On non-informativeness in a classical Bayesian inference problem
Coolen, F.P.A.

Published: 01/01/1992

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):
Memorandum COSOR 92-35

On non-informativeness in a classical Bayesian inference problem

F.P.A. Coolen

Eindhoven, September 1992
The Netherlands
ON NON-INFORMATIVENESS IN A CLASSICAL BAYESIAN INFERENCE PROBLEM

F.P.A. Coolen
Eindhoven University of Technology
Department of Mathematics and Computing Science
P.O. Box 513, 5600 MB Eindhoven, The Netherlands

Abstract
Not only between frequentists and Bayesians, but also among Bayesians, there is discrepancy on the answer of the essential question: 'Given \( \alpha \) successes in \( n \) previous trials, what is the probability of success at the next trial (all trials identical and independent)?'.

In this paper the Bayesian answer to this problem is discussed considering non-informative prior subjective data. Also some paradoxes related to this problem are discussed. The actual paradox seems to be that our feeling about using subjective information additional to real data that have become available depends on these data, which is not allowed in the Bayesian framework.

Key Words
Bernoulli trials, Bayesian statistics, non-informative prior, predictive probability, Hardy's paradox, Bing's paradox, rule of succession.
1. Introduction

A classical problem studied by statisticians is (Dale, 1991):
'Given $\alpha$ successes in $n$ previous trials, what is the probability of success at the next trial (all trials identical and independent)?'.

This is a standard Bernoulli situation. Independent random variables $X_i$ ($i=1,2,\ldots$) with identical binomial probability mass functions,

\[
P_X(X_i=x|\pi) = \pi^x(1-\pi)^{1-x},
\]

are sampled ($x=0$ or $1$ and $0\leq\pi\leq1$, $X_i$ are exchangeable according to De Finetti, 1974). An observation (a trial) of $X_i$ with value $1$ is called a success and an observation with value $0$ a failure.

If $n$ previous trials have resulted in $\alpha$ successes and $n-\alpha$ failures, this is denoted as data $(n,\alpha)$. The above problem is to find $P_X(X_{n+1}=1|(n,\alpha))$.

In the frequency theory of statistics both the maximum likelihood and the \textit{\alpha}-moment estimator for the parameter $\pi$ are equal to $\hat{\pi}=\alpha/n$, and an estimation of $P_X(X_{n+1}=1|(n,\alpha))$ is $\hat{P}_X(X_{n+1}=1|\hat{\pi},(n,\alpha)) = \hat{\pi} = \alpha/n$ (that is, if $n$ was fixed before the experiment, see Lindley and Phillips (1976)). In the Bayesian context the answer depends on the prior that is chosen before the data $(n,\alpha)$ become available, based on the knowledge or ideas about the outcome of such a trial. However, there are many situations in which one would prefer not to use any prior subjective information additional to $(n,\alpha)$.

Walley (1991) shows that the standard Bayesian framework is not suitable for indicating the amount of information used in defining a probability, and the best solution is derived by generalization of the probability concept to imprecise probabilities. Coolen (1992) proposes a model for the above problem, and discusses the behaviour of imprecision. Nonetheless, it is interesting to discuss this problem restricted to the precise Bayesian framework, as there still is disagreement among Bayesians about the solution.

In section 2 of this paper the Bayesian analysis is discussed, and it is concluded that the main problem is whether or not one wants to use any prior information. The ideas about this seem to depend on the data, which is in fact not allowed within the Bayesian theory, leading to some kind of paradox.

In section 3 some paradoxes in literature are discussed briefly, together with a simple example of the rule of succession.
2. Non-informativeness and the Bayesian solution

In the Bayesian context the parameter $\pi$ is regarded as a random variable, and the density of a conjugate prior distribution is a beta,

$$g_{\pi}(p) \propto p^{\alpha-1}(1-p)^{\beta-1} \text{ for } 0 \leq p \leq 1 \quad (\alpha, \beta > 0),$$

referred to by $\pi$-$\text{Be}(\alpha, \beta)$. Colombo and Constantini (1980) show that the beta family satisfies reasonable hypotheses for choosing prior distributions for the parameter $\pi$. The fact that prior imaginary data play an analogous role to real data is perhaps the most satisfactory argument for restriction to the beta family.

If $n$ trials lead to $\mathit{a}$ successes this prior can be updated giving a posterior $\pi$-$\text{Be}(\alpha+\mathit{a}, \beta+n-\mathit{a})$. The predictive distribution for $X_1$ (Raiffa and Schlaifer, 1961) related to the beta prior has probability mass function

$$P(X_1 = x) = \frac{B(\alpha+\mathit{x}, \beta+1-\mathit{x})}{B(\alpha, \beta)} \text{ for } x \in \{0, 1\},$$

with $B$ the beta function, so

$$P(X_{1+1} = 1) = \frac{\alpha}{\alpha + \beta} \quad \text{ and } \quad P(X_{1+1} = 0) = \frac{\beta}{\alpha + \beta} \text{ for } i \geq 1.$$ 

On updating $(\alpha, \beta)$ is replaced by $(\alpha+\mathit{n}, \beta+\mathit{n}-\mathit{a})$, leading to posterior predictive mass function

$$P(X_{n+1} = 1 \mid (n, \mathit{a})) = \frac{\alpha + \mathit{a}}{\alpha + \beta + n} \quad \text{ and } \quad P(X_{n+1} = 0 \mid (n, \mathit{a})) = \frac{\beta + n - \mathit{a}}{\alpha + \beta + n} \text{ for } i \geq 1.$$ 

The prior parameters $\alpha$ and $\beta$ can be interpreted as imaginary data $(\alpha+\beta, \alpha)$, and the effect of data on the posterior does not depend on whether these data are imaginary or real.

Many priors have been proposed to represent total lack of knowledge (Walley, 1991; section 5.5.2). Restricted to the above model with prior $\pi$-$\text{Be}(\alpha, \beta)$, the posterior predictive mass function is entirely determined by the data $(n, \mathit{a})$ only if $\alpha = \beta = 0$, which is formally not possible as this prior is improper, but this does not prevent the determination of the posterior distribution, $\pi$-$\text{Be}(\mathit{a}, n-\mathit{a})$, which is proper if $0 < \mathit{a} < n$.

This improper prior $\pi$-$\text{Be}(0, 0)$ has been proposed by, amongst others, Haldane (1945), Novick and Hall (1965), Jaynes (1968) and Villegas (1977). The resulting posterior predictive probabilities are

$$P(X_{n+1} = 1 \mid (n, \mathit{a})) = \frac{\mathit{a}}{n}$$

and

$$P(X_{n+1} = 0 \mid (n, \mathit{a})) = \frac{n - \mathit{a}}{n} \text{ for } i \geq 1,$$

which is identical to the frequentists' solution when $n$ was fixed before the experiment, so it seems that this discrepancy is resolved when Bayesians represent lack of knowledge with this particular prior.

Of course, this prior can only be used if data actually become available, and it is impossible to derive at statistical inferences based on this prior only. This is not unreasonable, as one can only reach inferences if some information is used, and total lack of historical data explicitly asks for
the use of subjective information. But to solve the stated problem, one
could regard \((n, \alpha)\) as information prior to the question what the probability
of the event \(X_{n+1} = 1\) is, and hence at this moment in time \(\pi - \text{Be}(\alpha, n - \alpha)\) could
be regarded as a prior based on actual historical data, which is intuitively
attractive.

A second problem, using \(\alpha = \beta = 0\), occurs if \(\alpha = 0\) or \(\alpha = n\), in which cases the
posterior is improper, and the posterior predictive probabilities are
\[
P_X(X_{n+1} = 1 | (n, 0)) = 0 \quad \text{and} \quad P_X(X_{n+1} = 1 | (n, n)) = 1
\]
respectively, for \(i \neq 1\). To many Bayesians this seems to be unreasonable (especially for \(n = 1\)), but not
accepting this means that one actually wants to take subjective opinion into
account, in a way other than through a prior.

From this discussion one could conclude that in many situations one would
feel that a really non-informative prior is desirable if \(0 < \alpha < n\), but not in
the other situations. As the prior must be chosen before the data become
available, this cannot formally be embedded in the Bayesian framework. This
is the reason that Bayesians generally use informative priors in problems
like the one above, so choosing positive \(\alpha\) and \(\beta\), even if the resulting
answer is not the answer to the question asked, but to that question with an
extra assumption about prior knowledge. The most often proposed solution is
based on the uniform prior \(\pi - \text{Be}(1, 1)\), that was also used by Bayes (1763)
and Laplace (1812). The corresponding answer to the above question then is
\[
P_X(X_{n+1} = 1 | (n, \alpha)) = \frac{\alpha + 1}{n + 2}
\]
by many authors seen as the correct solution to the question (this is called Laplace's rule of succession (Dale, 1991;
chapter 6)).

Lotze (Dale, 1991; section 8.4) gives a 'proof' for the solution
\[
P_X(X_{n+1} = 1 | (n, n)) = \frac{n + 1}{n + 2}
\]
based on the uniform prior, by stating that "the denominator equals the sum of possible cases, because after \(n\) real trials 2
possible cases can occur, either a success or a failure, at the next trial,
whereas the numerator gives the number of successes for these situations".
Of course this 'predictive' argument is very doubtful, as it seems to be
that the two possible results of the following trial are assumed to be
equally likely, so that one did not learn anything from the past results.
Based on the same idea of dividing the future expected number of successes
by the total number of trials (after data \((n, \alpha)\) are derived), a more logical
result for general \((n, \alpha)\) is given by (for \(m=1,2,\ldots\)):
\[ P_X(X_{n+m} = 1 | (n, \omega)) = \frac{\alpha + \sum_{i=1}^{m} P_X(X_{n+i} = 1 | (n, \omega))}{n + m}, \]
and, combining this with the assumption that, for prior \( \pi-Be(\alpha, \beta) \), \( P_X(X_i = 1 | (n, \omega)) = \frac{\alpha + \omega}{\alpha + \beta + n} \) for all \( i \leq n+1 \), we get \( \frac{\alpha + \omega}{\alpha + \beta + n} = \frac{\alpha + \omega}{n + m} \), which holds if \( \frac{\alpha}{\alpha + \beta} = \frac{\omega}{n} \) or \( \alpha = \beta = 0 \). As \( \alpha \) and \( \beta \) have to be chosen before the data become available, the only possible choice satisfying this condition is \( \alpha = \beta = 0 \). Of course, for \( \omega = 0 \) or \( \omega = n \) again one would somehow feel the urge to use subjective information. This is further discussed in section 3.

Lee (1989, section 3.2) mentions the fact that the mode of the posterior distribution \( \pi-Be(\alpha+\omega, \beta+n-\omega) \) is equal to \( \frac{\omega}{n} \) if and only if \( \alpha = \beta = 1 \), and regards this to be an estimator for \( \pi \) that relates to the maximum likelihood estimator in the frequentist approach, and that is therefore attractive. This argument is rather non-Bayesian, and it must be remarked that we are not primarily interested in estimating \( \pi \) but rather in calculating \( P_X(X_{n+1} = 1 | (n, \omega)) \), which equals \( \frac{\omega}{n} \) if and only if \( \alpha = \beta = 0 \).

Within the subjective probability theory related to personal betting behaviour (De Finetti, 1974), acceptance of a prior \( \pi-Be(\alpha, \beta) \) with positive \( \alpha \) or \( \beta \) also leads to remarkable betting behaviour. Within this theory, a person is for one and only one price \( P \) indifferent between buying or selling a bet on a certain event of interest that pays 1 'unit' if this event occurs, and nothing if it does not. Here the event of interest is \( X_{n+1} = 1 \) given data \( (n, \omega) \), and \( P = P_X(X_{n+1} = 1 | (n, \omega)) = \frac{\alpha + \omega}{\alpha + \beta + n} \). Within this theory, \( P \) would change if data \( (kn, k\omega) \) had been observed instead of \( (n, \omega) \), for \( k = 2, 3, \ldots \), except if \( \frac{\alpha}{\alpha + \beta} = \frac{\omega}{n} \) or \( \alpha = \beta = 0 \). We strongly feel that not using any additional subjective information should lead to betting behaviour depending only on the ratio \( \frac{\omega}{n} \), so \( P \) should be equal for all \( k \), where again data \( (n, 0) \) or \( (n, \omega) \) lead to intuitive problems. Also from this point of view the only logical choice is \( \alpha = \beta = 0 \).

Other arguments against positive \( \alpha \) or \( \beta \), are given in section 3 while discussing some paradoxes that are known from literature, together with Laplace’s rule of succession.
3. Some paradoxes

We consider two paradoxes discussed by Dale (1991), originating from the problem that choosing a uniform prior \( \pi \text{-Be}(1,1) \) does not represent total lack of prior information.

Hardy’s paradox (Dale, 1991; section 8.10) is:
"Suppose that, of 1000 lives exposed to risk at age 70, 900 survive to age 71 and 800 to age 72. Using Laplace’s formula one finds the predictive probabilities (let \( Y \) denote the age to which one survives, assuming exchangeability)
\[
P_Y(Y=71|Y=70) = \frac{901}{1002} \quad \text{and} \quad P_Y(Y=72|Y=71) = \frac{801}{902}, \quad \text{so}
\]
\[
P_Y(Y=72|Y=71) \times P(Y=71|Y=70) = \frac{901 \times 801}{1002 \times 902}, \quad \text{which is not equal to}
\]
\[
P_Y(Y=72|Y=70) = \frac{801}{1002} \quad \text{obtained from the original data.}"

Hardy concludes that, while the usual formula \( \frac{\alpha}{n} \) may be open to a theoretical objection, no better formula can be found. Indeed, when we have data as presented here, for predictive probabilities the posterior
\( \pi \text{-Be}(\alpha, n - \alpha), \) used once or twice depending on whether one conditions on \( Y=71 \) or not, leads to the most logical answer,
\[
P_Y(Y=72|Y=70) = P_Y(Y=72|Y=71) \times P_Y(Y=71|Y=70) = \frac{800}{1000}.
\]

Bing’s paradox (Dale, 1991; section 8.5) relates to a multinominal problem strongly related to the binomial problem discussed in this paper. We do not explicitly discuss the multinominal rule of succession (Dale, 1991), but the idea is the same as before:
"Suppose that 100 trials have yielded \( A, B \) and \( C \) respectively 49, 37 and 14 times, then the probability that the 101st trial will yield none of these three letters (but ‘something else’) is, according to the rule of succession, \( 1/104 \). On the other hand, if we merely consider that a letter has been drawn in all 100 trials, then the probability against drawing a letter on the next trial is \( 1/102 \), again according to the rule of succession."

For this situation the problem is also resolved if the multinominal prior were really non-informative, instead of assuming an imaginary set of data prior to the observations consisting of \('1 A, 1 B, 1 C \) and \('something else‘) leading to \( 1/104 \), or \('1 letter and 1 non-letter‘) leading to \( 1/102 \).
These paradoxes clearly show the problems that the uniform prior may lead
to. We end this discussion by looking at another example from Dale (1991;
section 8.9, originally by Poisson):
"Two playing-cards are lying face-down on a table. One is turned over and
found to be black. What is the probability that the second card is black?". 
It is important to remark that no more information is available about the
cards. Now there are several possible ways to try to answer this problem
(let $C_i$ denote the color of the i-th card, $i=1,2$, and a color is denoted by
its first letter).

Firstly, one could use Laplace's rule of succession, leading to
$P(C_2=B|C_1=B) = 2/3$. This is based on the uniform prior, which could be
defended by the idea that playing-cards are either black or red. This leads
to the remarkable result that the prior probability of both cards having the
same color is equal to $2/3$, when calculated by conditioning on the color of
the first card that is turned over, while the prior for this situation can
also be calculated by $\{P(C_1=B)\}^2 + \{P(C_1=R)\}^2 = 1/2$. We do not know any
reasonable argument why this difference should exist (compare Hardy's
paradox). Further, if more than two colors were possible, a uniform prior
for the multinomial case would lead to the same problem as Bing's paradox.

Secondly, one could assume, that both cards are taken randomly from an
infinite population of 50% black and 50% red cards, leading to
$P(C_2|C_1) = 1/2$ by independence. Here the extra assumption is obvious. If
the population is assumed to be finite, then dependence will cause
$P(C_2|C_1) < 1/2$.

Thirdly, as a frequentists' approach, an assumption like the former can be
made with $100*p\%$ black cards and $100*(1-p)\%$ red cards. Now the question
arises what stopping rule is used (one needs to know the sample space), but
the amount of data is too small to give a reasonable estimate of $p$.

Fourthly, using the non-informative improper prior $\pi \sim Be(0,0)$ would lead to
$P(C_2|C_1) = 1$, which all of us would regard to be unreasonable. This is just
the point stated earlier, that we actually want to use additional subjective
information if we get only very little data. This prior, or better to say
the posterior resulting from this prior, only makes sense if $0<\omega<n$, in which
case the posterior is, for inferences on future observations, the only
real-data-based distribution of $\pi$.

Lastly, one could enhance a complete subjective method and, after $C_1=B$ is
derived, try to find one single rate at which you would be indifferent
between selling or buying a bet that pays 1 'unit' if \( C_2 = B \), and nothing otherwise (De Finetti (1974), according to whom one and only one such rate must exist, see Walley (1991) for comments on this assumption and generalization).

Perhaps the most logical thing to do, in such situation, is to admit that there is too little information, and therefore inferences must, for the largest part, be based on subjective information. From this point of view the last mentioned solution seems to be best. Solutions based on the rule of succession rely too strongly on the prior distribution and therefore on information that is not available or on implicit assumptions; these are not the right solutions to problems of this kind.

If there is enough data, with \( 0 < \alpha < n \), then the posterior \( \pi - \text{Be}(\alpha, n - \alpha) \) is by far the most logical to be used for inferences on future observations. One important difference between this and the frequentist approach is that this posterior is actually an entirely data-based distribution of \( \pi \), and therefore Highest Density Regions (Lee, 1989) can be constructed, that are much easier and more logical to interpret than frequentists' confidence intervals. Another difference is that not only the maximum of the likelihood function is used to estimate \( P(X_{n+1} = 1 | (n, \alpha)) \), as in the frequentist approach, but this probability is calculated using the entire posterior distribution for \( \pi \).

As Lindley (1990) clearly states, both the Bayesians and frequentists need to add some assumption to the model, namely a prior or a sample space (information about the stopping rule) respectively. However, if a Bayesian were to regard the data \((n, \alpha)\) as actual prior information for answering the question as to what \( P(X_{n+1} = 1 | (n, \alpha)) \) is, then \( \pi - \text{Be}(\alpha, n - \alpha) \) would be the very best choice and inferences can be made using the predictive probabilities based on this distribution of \( \pi \). In some extreme cases, when \( \alpha = 0 \) or \( \alpha = n \), one would still feel the need to add subjective information to the real data to answer this question. Mostly these extreme cases are not of very great practical importance, and it is necessary to ask ourselves what we actually learn from data of the form \( \alpha = 0 \) or \( \alpha = n \). For many practical problems in which data \((n, \alpha)\) are available, statistical inferences based on \( \pi - \text{Be}(\alpha, n - \alpha) \) are most satisfactory, and give the right answer to the basic problem without adding extra information. It is really more logical, if data \((n, \alpha)\) are available, to use these data as prior information to the problem to be
solved than to use \((a+\beta+n, a+\alpha)\) for some nonnegative \(\alpha\) or \(\beta\), unless one wants to add subjective information. If the above problem is stated by a person presenting real data \((n, \alpha)\), then it is remarkable to add imaginary data \((\alpha+\beta, \alpha)\) to the real data before answering the question. For a statistician, the data \((n, \alpha)\) are real prior data (assuming that these are available when first confronted with the problem), and there is no need to update. The desire to find a prior that is both non-informative and that leads to satisfactory answers to this problem for all situations is paradoxical, as only after data have become available can we decide whether or not we want to add subjective information, which is not possible in a Bayesian analysis. The paradox is that real data have influence on our opinion about using no additional subjective information at all.

References


Laplace, P.S. de (1812) Théorie analytique des probabilités (Courcier, Paris).


<table>
<thead>
<tr>
<th>Number</th>
<th>Month</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>92-01</td>
<td>January</td>
<td>F.W. Steutel</td>
<td>On the addition of log-convex functions and sequences</td>
</tr>
<tr>
<td>92-02</td>
<td>January</td>
<td>P. v.d. Laan</td>
<td>Selection constants for Uniform populations</td>
</tr>
<tr>
<td>92-03</td>
<td>February</td>
<td>E.E.M. v. Berkum</td>
<td>Data reduction in statistical inference</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H.N. Linssen</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>D.A. Overdijk</td>
<td></td>
</tr>
<tr>
<td>92-04</td>
<td>February</td>
<td>H.J.C. Huijberts</td>
<td>Strong dynamic input-output decoupling: from linearity to nonlinearity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>H. Nijmeijer</td>
<td></td>
</tr>
<tr>
<td>92-05</td>
<td>March</td>
<td>S.J.L. v. Eijndhoven</td>
<td>Introduction to a behavioral approach of continuous-time systems</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J.M. Soethoudt</td>
<td></td>
</tr>
<tr>
<td>92-06</td>
<td>April</td>
<td>P.J. Zwietering</td>
<td>The minimal number of layers of a perceptron that sorts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E.H.L. Aarts</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td>92-07</td>
<td>April</td>
<td>F.P.A. Coolen</td>
<td>Maximum Imprecision Related to Intervals of Measures and Bayesian Inference with Conjugate Imprecise Prior Densities</td>
</tr>
<tr>
<td>92-08</td>
<td>May</td>
<td>I.J.B.F. Adan</td>
<td>A Note on “The effect of varying routing probability in two parallel queues with dynamic routing under a threshold-type scheduling”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>W.H.M. Zijm</td>
<td></td>
</tr>
<tr>
<td>92-09</td>
<td>May</td>
<td>I.J.B.F. Adan</td>
<td>Upper and lower bounds for the waiting time in the symmetric shortest queue system</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G.J.J.A.N. v. Houtum</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. v.d. Wal</td>
<td></td>
</tr>
<tr>
<td>92-10</td>
<td>May</td>
<td>P. v.d. Laan</td>
<td>Subset Selection: Robustness and Imprecise Selection</td>
</tr>
<tr>
<td>92-11</td>
<td>May</td>
<td>R.J.M. Vaessens</td>
<td>A Local Search Template</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E.H.L. Aarts</td>
<td>(Extended Abstract)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J.K. Lenstra</td>
<td></td>
</tr>
<tr>
<td>92-12</td>
<td>May</td>
<td>F.P.A. Coolen</td>
<td>Elicitation of Expert Knowledge and Assessment of Imprecise Prior Densities for Lifetime Distributions</td>
</tr>
<tr>
<td>92-13</td>
<td>May</td>
<td>M.A. Peters</td>
<td>Mixed $H_2/H_\infty$ Control in a Stochastic Framework</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A.A. Stoorvogel</td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>Month</td>
<td>Author</td>
<td>Title</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>--------</td>
<td>-------</td>
</tr>
<tr>
<td>92-14</td>
<td>June</td>
<td>P.J. Zwietering, E.H.L. Aarts, J. Wessels</td>
<td>The construction of minimal multi-layered perceptrons: a case study for sorting</td>
</tr>
<tr>
<td>92-15</td>
<td>June</td>
<td>P. van der Laan</td>
<td>Experiments: Design, Parametric and Nonparametric Analysis, and Selection</td>
</tr>
<tr>
<td>92-16</td>
<td>June</td>
<td>J.J.A.M. Brands, F.W. Steutel, R.J.G. Wilms</td>
<td>On the number of maxima in a discrete sample</td>
</tr>
<tr>
<td>92-17</td>
<td>June</td>
<td>S.J.L. v. Eijndhoven, J.M. Soethoudt</td>
<td>Introduction to a behavioral approach of continuous-time systems part II</td>
</tr>
<tr>
<td>92-18</td>
<td>June</td>
<td>J.A. Hoogeveen, H. Oosterhout, S.L. van der Velde</td>
<td>New lower and upper bounds for scheduling around a small common due date</td>
</tr>
<tr>
<td>92-19</td>
<td>June</td>
<td>F.P.A. Coolen</td>
<td>On Bernoulli Experiments with Imprecise Prior Probabilities</td>
</tr>
<tr>
<td>92-20</td>
<td>June</td>
<td>J.A. Hoogeveen, S.L. van de Velde</td>
<td>Minimizing Total Inventory Cost on a Single Machine in Just-in-Time Manufacturing</td>
</tr>
<tr>
<td>92-21</td>
<td>June</td>
<td>J.A. Hoogeveen, S.L. van de Velde</td>
<td>Polynomial-time algorithms for single-machine bicriteria scheduling</td>
</tr>
<tr>
<td>92-22</td>
<td>June</td>
<td>P. van der Laan</td>
<td>The best variety or an almost best one? A comparison of subset selection procedures</td>
</tr>
<tr>
<td>92-23</td>
<td>June</td>
<td>T.J.A. Storcken, P.H.M. Ruys</td>
<td>Extensions of choice behaviour</td>
</tr>
<tr>
<td>92-25</td>
<td>July</td>
<td>P.J. Zwietering, E.H.L. Aarts, J. Wessels</td>
<td>Exact Classification With Two-Layered Perceptrons</td>
</tr>
<tr>
<td>92-26</td>
<td>July</td>
<td>M.W.P. Savelsbergh</td>
<td>Preprocessing and Probing Techniques for Mixed Integer Programming Problems</td>
</tr>
<tr>
<td>Number</td>
<td>Month</td>
<td>Author</td>
<td>Title</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>---------------------------------</td>
<td>-----------------------------------------------------</td>
</tr>
<tr>
<td>92-27</td>
<td>July</td>
<td>I.J.B.F. Adan</td>
<td>Analysing $E_k</td>
</tr>
<tr>
<td></td>
<td></td>
<td>W.A. van de Waarsenburg</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td>92-28</td>
<td>July</td>
<td>O.J. Boxma</td>
<td>The compensation approach applied to a $2 \times 2$ switch</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G.J. van Houtum</td>
<td></td>
</tr>
<tr>
<td>92-29</td>
<td>July</td>
<td>E.H.L. Aarts</td>
<td>Job Shop Scheduling by Local Search</td>
</tr>
<tr>
<td></td>
<td></td>
<td>P.J.M. van Laarhoven</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J.K. Lenstra</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>N.L.J. Ulder</td>
<td></td>
</tr>
<tr>
<td>92-30</td>
<td>August</td>
<td>G.A.P. Kindervater</td>
<td>Local Search in Physical Distribution Management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M.W.P. Savelsbergh</td>
<td></td>
</tr>
<tr>
<td>92-31</td>
<td>August</td>
<td>M. Makowski</td>
<td>MP-DIT Mathematical Program data Interchange Tool</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M.W.P. Savelsbergh</td>
<td></td>
</tr>
<tr>
<td>92-32</td>
<td>August</td>
<td>J.A. Hoogeveen</td>
<td>Complexity of scheduling multiprocessor tasks with</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S.L. van de Velde</td>
<td>prespecified processor allocations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B. Veltman</td>
<td></td>
</tr>
<tr>
<td>92-33</td>
<td>August</td>
<td>O.J. Boxma</td>
<td>Tandem queues with deterministic service times</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J.A.C. Resing</td>
<td></td>
</tr>
<tr>
<td>92-34</td>
<td>September</td>
<td>J.H.J. Einmahl</td>
<td>A Bahadur-Kiefer theorem beyond the largest</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>observation</td>
</tr>
<tr>
<td>92-35</td>
<td>September</td>
<td>F.P.A. Coolen</td>
<td>On non-informativeness in a classical Bayesian</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>inference problem</td>
</tr>
</tbody>
</table>