Imprecise conjugate prior densities for the one-parameter exponential family of distributions
Coolen, F.P.A.

Published: 01/01/1991

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

Memorandum COSOR 91-36
Imprecise Conjugate Prior Densities for the
One-Parameter Exponential Family of Distributions
F.P.A. Coolen

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

Eindhoven, December 1991
The Netherlands
Imprecise Conjugate Prior Densities for the
One-Parameter Exponential Family of Distributions

F.P.A. Coolen
Eindhoven University of Technology
Department of Mathematics and Computing Science

Abstract
A generalization of the standard Bayesian theory of statistical inference is presented for members of the one-parameter exponential family of distributions, such that imprecise prior densities are allowed. This enables representation of lack of perfect prior information about the probability distribution for the parameter of interest. A model is suggested with imprecise conjugate prior densities to enable simple updating of the prior densities if new data come available.
1. Introduction

The Bayesian theory of statistical inference demands a prior distribution for some parameter, that represents available information before experimental data come available. Walley [9] presents a clear and extensive discussion about the major drawbacks of presenting a certain amount of information, or lack of information, by a single prior distribution. If, for example, the prior information consists of opinions of several experts, it is important for the decision maker to report the level of conflict among the experts. This is not possible by using a single distribution.

To this end the concept of imprecise probabilities is introduced (Walley [9]), where not just one value is assigned as the probability of a certain event, but an interval of values is chosen, with the bounds called the lower and upper probabilities for this event. This interval can be interpreted as a set of possible values for the unknown probability $P(A)$, between which you do not want to make further distinction, given the available information. Of course this implies that these bounds depend on you and your information, and therefore have a subjective nature.

This idea suggests small intervals in case of much relevant information, while lack of information should lead to large intervals. Total lack of information about the probability of a certain event can be represented by assigning the values zero and one to the lower and upper probabilities for this event, respectively.

In Coolen [1] another possible interpretation, that is used by Walley [9] and relates to the subjective nature of probabilities as advocated by De Finetti [2], is discussed. This alternative interpretation is necessary for Walley's more general concept of imprecise previsions, but can be avoided in this report. Note that in the standard Bayesian theory, as advocated by De Finetti [2], the lower and upper probabilities for $A$ are equal. In this situation we call the probability for event $A$ precise.

In this report the standard Bayesian theory is generalized by allowing imprecise prior densities for the parameter of a precise parametric distribution. We have not discovered earlier work on such models, and restrict to probability distributions that belong to the one-parameter exponential family.
This theory is different to sensitivity analysis, with regard to the prior, in the standard Bayesian framework. Our approach treats lack of perfect knowledge within the concept, while the use of sensitivity analysis leads to the contradictory fact that a concept is used that needs complete knowledge of probabilities, whereas the absence of this knowledge is the reason for the sensitivity analysis. Another important difference is that in sensitivity analysis only a finite number of distributions can be compared, while in our approach a set of an infinite number of prior distributions is used (except in case of precision).

To enable simple updating of the imprecise prior densities, after new data come available, we restrict to prior distributions that belong to a conjugate family of distributions. If, according to an assumed parametric distribution for the random variable of interest, with parameter $\theta$, the likelihood function is $L(\theta|x)$, then a class $\Pi$ of prior distributions is said to form a conjugate family if the posterior density, $p(\theta|x) \propto p(\theta)L(\theta|x)$, is in the class $\Pi$ for all $x$ if the prior density $p(\theta) \in \Pi$ (the symbol $\propto$ indicates 'equal but for a constant factor'). This means that the posterior distribution has the same form as the prior, and updating is simply done by changing some parameters of the prior distribution (called hyperparameters). This avoids the calculation of integrals that is generally necessary to update prior distributions. For many models the conjugate priors can be updated using only a few sufficient statistics of a set of data that comes available. If the class $\Pi$ is large, meaning that it contains many distributions with very different shapes, restriction to such a class is not an important objection for practical use of the Bayesian theory.

In section 2 of this report imprecise probabilities are introduced and some important results are summarized. Section 3 introduces the one-parameter exponential family of distributions in a general form, and in section 4 a general form for imprecise conjugate prior densities, for members of this family, is suggested, and some results are presented. These lower and upper prior densities are of the same form, but differ some factor that relates to the amount of imprecision, so in fact to the amount of available information or the level of confidence in the prior knowledge about the necessary probability distributions. In case of updating this factor should also change, as the imprecision will generally decrease if new information is
gathered, and in section 5 a simple form for this factor is suggested. Section 6 provides examples for some well-known members of this family of distributions: the Weibull, normal, Poisson, (negative) binomial and gamma distributions. Finally, in section 7 some concluding remarks are presented. Also some appendices are added to this report. The first is an overview of standard functions and distributions used in this report, while the other appendices provide calculations for section 6.
2. Imprecise Probabilities

In this section some relations are given that are used in this report. These relations are presented in Coolen [1] or Walley [9], where also a complete introduction and extensive discussion of lower and upper probabilities are given. We restrict to a short introduction.

The lower and upper probabilities of an event \( \mathcal{A} \in \Omega \) (the set of all possible events) are denoted by \( \underline{P}(\mathcal{A}) \) and \( \overline{P}(\mathcal{A}) \) respectively, and must satisfy the following basic axioms (see Coolen [1] or Wolfenson and Fine [10]):

(A1) For all \( \mathcal{A} \in \Omega \): \( \underline{P}(\mathcal{A}) \geq 0 \).

(A2) \( \overline{P}(\Omega) = 1 \).

(A3) For all \( \mathcal{A}, \mathcal{B} \in \Omega \), with \( \mathcal{A} \cap \mathcal{B} = \emptyset \): \( \underline{P}(\mathcal{A}) + \underline{P}(\mathcal{B}) \leq \underline{P}(\mathcal{A} \cup \mathcal{B}) \) and \( \overline{P}(\mathcal{A}) + \overline{P}(\mathcal{B}) \geq \overline{P}(\mathcal{A} \cup \mathcal{B}) \).

(A4) For all \( \mathcal{A} \in \Omega \): \( \underline{P}(\mathcal{A}) + \underline{P}(\mathcal{A}^c) = 1 \), where \( \mathcal{A}^c \) is the complement of \( \mathcal{A} \).

Elementary consequences of these axioms are (proven in Coolen [1]):

(C1) \( \underline{P}(\emptyset) = \underline{P}(\emptyset) = 0 \).

(C2) \( \overline{P}(\Omega) = 1 \).

(C3) \( \overline{P}(\mathcal{A}) \geq \underline{P}(\mathcal{A}) \).

(C4) \( \mathcal{A} \subseteq \mathcal{B} \) implies \( \underline{P}(\mathcal{B}) \geq \underline{P}(\mathcal{A}) \) and \( \overline{P}(\mathcal{B}) \geq \overline{P}(\mathcal{A}) \).

(C5) \( \mathcal{A} \cap \mathcal{B} = \emptyset \) implies \( \underline{P}(\mathcal{A} \cup \mathcal{B}) \leq \underline{P}(\mathcal{A}) + \overline{P}(\mathcal{B}) \).

Walley [9] presents the foundation of a more general theory, and also regards problems of coherence. The imprecise probabilities used in this report do not lead to incoherence.

Special cases of imprecise probabilities are lower and upper cumulative distribution functions (cdf) for a real variable \( X \). These are the lower and upper probabilities of the events \( X \leq x \) for \( x \in \mathbb{R} \), and are denoted by \( \underline{F}(x) = \underline{P}(X \leq x) \) and \( \overline{F}(x) = \overline{P}(X \leq x) \).

A suitable method to elicit a persons opinion about \( X \) is to ask him to assign two functions, say \( l(x) \) and \( u(x) \) with \( 0 \leq l(x) \leq u(x) \) for all \( x \), such that all functions \( h \) between \( l \) and \( u \) can, after normalization, be regarded as probability density functions (pdf) for \( X \). The functions \( l \) and \( u \) are called lower and upper density functions. We assume, in this report, that \( l \) and \( u \) have positive finite integrals (although theoretically this is not
necessary), and that \( l \) and \( u \) are continuous functions. If lower and upper density functions are given, then the lower and upper cdf's, defined by

\[
F(x) = \frac{\int_{-\infty}^{x} l(\omega) \, d\omega + \int_{x}^{\infty} u(\omega) \, d\omega}{\int_{-\infty}^{\infty} l(\omega) \, d\omega + \int_{-\infty}^{\infty} u(\omega) \, d\omega}
\]

and

\[
\bar{F}(x) = \frac{\int_{-\infty}^{x} u(\omega) \, d\omega}{\int_{-\infty}^{\infty} u(\omega) \, d\omega + \int_{-\infty}^{\infty} l(\omega) \, d\omega}
\]

are the bounds of all cdf's that can be constructed from densities (after normalization) that lie between \( l \) and \( u \).

The lower and upper cdf's, constructed from the lower and upper densities, have pdf's \( f \) and \( \bar{f} \), respectively, with

\[
f(x) = \frac{l(x) \int_{x}^{\infty} u(\omega) \, d\omega + u(x) \int_{-\infty}^{x} l(\omega) \, d\omega}{\left( \int_{-\infty}^{\infty} l(\omega) \, d\omega + \int_{-\infty}^{\infty} u(\omega) \, d\omega \right)^2}
\]

and

\[
\bar{f}(x) = \frac{u(x) \int_{x}^{\infty} l(\omega) \, d\omega + l(x) \int_{-\infty}^{x} u(\omega) \, d\omega}{\left( \int_{-\infty}^{\infty} u(\omega) \, d\omega + \int_{-\infty}^{\infty} l(\omega) \, d\omega \right)^2}.
\]

Within the Bayesian framework, let \( l(\theta) \) and \( u(\theta) \) be imprecise prior densities, with \( \theta \in \Theta \) (for some parameter space \( \Theta \)) the parameter of a sampling distribution with pdf \( f(x|\theta) \).

After observing data \( x \), the updated versions of \( l(\theta) \) and \( u(\theta) \) are

\[
l(\theta|x) = L(\theta|x)l(\theta) \quad \text{and} \quad u(\theta|x) = L(\theta|x)u(\theta),
\]

respectively, with \( L(\theta|x) \) the likelihood function defined by the chosen model. The corresponding updated lower and upper cdf's, \( F(\theta|x) \) and \( \bar{F}(\theta|x) \), and corresponding pdf's are derived as above.

Consequent definitions of lower and upper predictive densities, based on prior densities \( l(\theta) \) and \( u(\theta) \), are:

\[
\int_{\Theta} f(x|\theta) l(\theta) \, d\theta \quad \text{and} \quad \int_{\Theta} f(x|\theta) u(\theta) \, d\theta,
\]

respectively.
Again these are not pdf's. According lower and upper predictive cdf's for X are derived as above.
The predictive densities are updated by replacing \( l(\theta) \) and \( u(\theta) \) by \( l(\theta|x) \) and \( u(\theta|x) \).

For an event \( A \), the degree of imprecision is defined by \( \Delta(A) = \bar{P}(A) - \underline{P}(A) \), that is zero only when \( \underline{P}(A) = \bar{P}(A) \) (then the probability for \( A \) is called precise), and one if \( \underline{P}(A) = 0 \) and \( \bar{P}(A) = 1 \) (then the probabilities for \( A \) are called vacuous). For all other situations \( 0 < \Delta(A) < 1 \).

We also use a simple measure for the amount of information concerning \( A \) in \( \bar{P} \), that is introduced by Walley [9, section 5.3.7]: \( I(A) = \frac{1}{\Delta(A)} - 1 \). This is zero if the probabilities for \( A \) are vacuous, and infinite if the probability for \( A \) is precise.
3. The One-Parameter Exponential Family of Distributions

Many of the common statistical distributions have a similar form. This leads to the definition that a distribution belongs to the one-parameter exponential family (Lee [4]) if its pdf can be put into the form

\[ p(x|\theta) = g(x)h(\theta)\exp\{t(x)\psi(\theta)\}, \]

or equivalently if the likelihood of \( n \) independent observations \( x = (x_1, x_2, \ldots, x_n) \) from this distribution is

\[ L(\theta|x) \propto h(\theta)^n\exp\left\{ \sum_{i=1}^{n} t(x_i)\psi(\theta) \right\}. \]

Here \( X \) is the random variable of interest, and its distribution depends on a one-dimensional parameter \( \theta \), say \( \theta \in \Theta \). If \( X \) is continuous, we also write \( f(x|\theta) \) instead of \( p(x|\theta) \).

From the form of the likelihood it is clear that \( (n, \sum t(x_i)) \) are sufficient statistics of the observations.

If a distribution belongs to this family, there is an unambiguous definition of a conjugate family of precise prior distributions. It is defined to be the family \( \Pi \) of pdf's such that

\[ p(\theta) \propto h(\theta)^\nu\exp\{\tau\psi(\theta)\}. \]

Here \( (\nu, \tau) \) are hyperparameters, that can be interpreted as sufficient statistics of imaginary observations, where \( \nu \) and \( \tau \) correspond to \( n \) and \( \sum t(x_i) \) respectively. The prior pdf is uniquely determined by \( (\nu, \tau) \).

If data come available, the prior pdf can be updated by Bayes’ rule, leading to the posterior pdf

\[ p(\theta|x) \propto h(\theta)^{\nu+n}\exp\{\tau+\sum t(x_i)\psi(\theta)\}. \]

The posterior pdf has the same form as the prior pdf, with \( (\nu, \tau) \) replaced by \( (\nu+n, \tau+\sum t(x_i)) \).

Based on the prior \( p(\theta) \), the predictive pdf is

\[ p(x) \propto \int p(x|\theta)p(\theta)d\theta, \]
while based on the posterior pdf, the predictive pdf is

\[ p(x|x) \propto \int \frac{p(x|\theta)p(\theta|x)}{p(x)} \, d\theta. \]

If \( X \) is continuous, we also write \( f(x) \) and \( f(x|x) \) instead of \( p(x) \) and \( p(x|x) \).

This leads to

\[ p(x) \propto g(x) \int h(\theta)^{\nu+1} \exp\{ (\tau + t(x))\psi(\theta) \} \, d\theta, \quad \text{and} \]

\[ p(x|x) \propto g(x) \int h(\theta)^{\nu+n+1} \exp\{ (\tau + \sum x_i + t(x))\psi(\theta) \} \, d\theta. \]

In section 6 examples of distributions that are members of this family are given.

An interesting generalization of this family of distributions is the two-parameter exponential family. A distribution belongs to this family if its pdf can be written as

\[ p(x|\theta, \varphi) = g(x) h(\theta, \varphi) \exp\{ t(x)\psi(\theta, \varphi) + \nu(x)\chi(\theta, \varphi) \}, \]

or, equivalently, if the likelihood of \( n \) independent observations \( x = \{ x_1, x_2, \ldots, x_n \} \) takes the form

\[ L(\theta, \varphi|x) \propto h(\theta, \varphi)^n \exp\left\{ \sum_{i=1}^n t(x_i)\psi(\theta, \varphi) + \sum_{i=1}^n \nu(x_i)\chi(\theta, \varphi) \right\}. \]

The family of prior densities conjugate to this likelihood takes the form

\[ p(\theta, \varphi) \propto h(\theta, \varphi)^\nu \exp\{ \tau\psi(\theta, \varphi) + \zeta\chi(\theta, \varphi) \}. \]

We do not discuss this family here, but mention it because the normal distribution with both mean and variance unknown is a member of this family. Results for the one-parameter family can be generalized.

The idea of the exponential family can easily be extended to a \( k \)-parameter exponential family in an obvious way, and also for those distributions results as presented in this report can be derived.
4. Imprecise Conjugate Priors

The distribution of a random variable $X$ is represented by its pdf $f(x|\theta)$ with $\theta \in \mathbb{R}$, and we restrict ourselves to imprecise priors with the following relation:

$$u(\theta) = c_0 l(\theta),$$

where $c_0 \geq 1$ is independent of $\theta$. We define $F_1(\theta)$ to be the cdf corresponding to $l(\theta)$ after normalization,

$$F_1(\theta) = \frac{\int_{-\infty}^{\theta} l(\omega) d\omega}{\int_{-\infty}^{\infty} l(\omega) d\omega}.$$

Formulas (1) and (2) of section 2 lead to

$$F(\theta) = \left[1 + c_0 \left(\frac{1}{F_1(\theta)} - 1\right)\right]^{-1} \quad \text{and} \quad \bar{F}(\theta) = \left[1 + c_0^{-1} \left(\frac{1}{F_1(\theta)} - 1\right)\right]^{-1},$$

respectively.

(The according lower and upper pdf's are easily found by taking the first derivatives.)

The resulting prior imprecision is

$$\Delta(\theta) = \bar{F}(\theta) - F(\theta) = \frac{c_0^2 - 1}{(c_0^{-1})^2 + c_0 [F_1(\theta)(1-F_1(\theta))]}^{-1}.$$

It is known that $0 \leq F_1(\theta)(1-F_1(\theta)) \leq \frac{1}{4}$, with the maximum value for $\theta = \theta_m$, where $\theta_m$ is the median of the distribution with cdf $F_1$. Hence

$$0 \leq \Delta(\theta) \leq \Delta(\theta_m) = \frac{c_0^{-1}}{c_0^2 + 1}.$$

If $n$ independent observations come available, the likelihood $L(\theta|x)$, according to the chosen model $f(x|\theta)$, is used to update $l$ and $u$, and $c_0$ is replaced by $c_n$. This leads to

$$l(\theta|x) = L(\theta|x) l(\theta),$$

and

$$u(\theta|x) = \frac{c_n}{c_0} L(\theta|x) u(\theta) = c_n l(\theta|x), \quad \bar{F}(\theta|x) = \left[1 + c_n \left(\frac{1}{F_1(\theta|x)} - 1\right)\right]^{-1} \quad \text{and} \quad \bar{F}(\theta|x) = \left[1 + c_n^{-1} \left(\frac{1}{F_1(\theta|x)} - 1\right)\right]^{-1},$$

with

$$F_1(\theta|x) = \frac{\int_{-\infty}^{\theta} l(\omega|x) d\omega}{\int_{-\infty}^{\infty} l(\omega|x) d\omega}.$$
The resulting posterior imprecision is

\[ \Delta(\theta | x) = \bar{F}(\theta | x) - \hat{F}(\theta | x) = \frac{c_n^2 - 1}{(c_n - 1)^2 + c_n [\hat{F}_1(\theta | x)(1-\hat{F}_1(\theta | x))]^{-1}}. \]

Now again

\[ 0 \leq \Delta(\theta | x) \leq \Delta(\theta_{m,x}) = \frac{c_n - 1}{c_n + 1}, \text{ with } \theta_{m,x} \text{ the median of the distribution} \]

with cdf \( \hat{F}_1(\theta | x) \).

The maximum imprecision keeps the same form, and only depends on \( c_n \). Now it is clear that we want \( c_n \) to depend on \( n \). If we get more observations, so \( n \) increases, we get more information, so it seems reasonable that the amount of imprecision decreases. To achieve this, we need to define \( c_n \) such that the maximum imprecision decreases as \( n \) increases. This is discussed in section 5.

Before this, we give the predictive distributions and imprecision, and propose a form for \( \ell(\theta) \), for the models described in section 3.

Based on the prior densities \( \ell(\theta) \) and \( u(\theta) \), with \( u(\theta) = c_0 \ell(\theta) \), the lower and upper predictive densities are

\[ l_x(x) = \int_{-\infty}^{\infty} f(x|\theta) \ell(\theta) d\theta \]

and

\[ u_x(x) = \int_{-\infty}^{\infty} f(x|\theta) u(\theta) d\theta, \text{ respectively, so } u_x(x) = c_0 l_x(x). \]

The according lower and upper cdf's are

\[ F_{-X}(x) = \left[ 1 + c_0 \left( \frac{1}{F_{X,1}(x)} - 1 \right) \right]^{-1} \]

and

\[ \bar{F}_{-X}(x) = \left[ 1 + c_0 \left( \frac{1}{F_{X,1}(x)} - 1 \right) \right]^{-1}, \]

respectively, with \( F_{X,1} \) the cdf corresponding to \( l_x \) after normalization.

This leads to prior imprecision

\[ \Delta_X(x) = \bar{F}_X(x) - F_{-X}(x) = \frac{c_0^2 - 1}{(c_0 - 1)^2 + c_0 [F_{X,1}(x)(1-F_{X,1}(x))]^{-1}}. \]

Again \( 0 \leq \Delta_X(x) \leq \Delta_X(x_m) = \frac{c_0 - 1}{c_0 + 1}, \text{ with } x_m \text{ the median of the distribution} \)

with cdf \( F_{X,1} \).

After updating, again the same form for the imprecision is derived, with \( c_0 \) replaced by \( c_n \), while

\[ l_x(x|x) = \int_{-\infty}^{\infty} f(x|\theta) \ell(\theta|x) d\theta \]

and

\[ u_x(x|x) = \int_{-\infty}^{\infty} f(x|\theta) u(\theta|x) d\theta, \text{ so } u_x(x|x) = c_n l_x(x|x). \]
So the choice of $l$ and $u$, such that $u(\theta) = c_n l(\theta)$ with $c_n \geq 1$ independent of $\theta$, leads to equality of the maximum imprecision for $\theta$ and for $X$, $\Delta_{\theta} = \Delta_X(x)$. 

For a member of the one-parameter exponential family, $l$ can be defined proportional to a precise conjugate prior distribution, leading to simple calculations. Hence we define $l(\theta) = h(\theta) \nu \exp\{\tau \psi(\theta)\}$, with hyperparameters $(\nu, \tau)$, and $u(\theta) = c_0 l(\theta)$, with $c_0 \geq 1$.

This gives

$$F_1(\theta) = \frac{\int_{-\infty}^{\theta} l(\omega) d\omega}{\int_{-\infty}^{\infty} l(\omega) d\omega} = \frac{\int_{-\infty}^{\theta} h(\omega) \nu \exp\{\tau \psi(\omega)\} d\omega}{\int_{-\infty}^{\infty} h(\omega) \nu \exp\{\tau \psi(\omega)\} d\omega},$$

and the according $F(\theta)$ and $\bar{F}(\theta)$ follow by application of the above formulas.

Updating is done by replacing $(\nu, \tau)$ by $(\nu+n, \tau + n \sum_{i=1}^{n} t(x_i))$, and replacing $c_0$ by $c_n$. Also the predictive lower and upper cdf's are easily derived, using the results in this section and section 3. In section 6 we give examples for several members of this family of distributions.
5. Interpretation and choice of $c_n$

According to the model assumed in section 4, the imprecision is an increasing function of $c_n$. If $c_n = 1$ then the resulting maximal imprecision $\Delta(\theta_m|x) = 0$, where $x = (x_1, \ldots, x_n)$, while $\Delta(\theta_m|x) \to 1$ if $c_n \to \infty$. The degree of imprecision is determined by $c_n$, so it is logical to define $c_n$ as a strictly decreasing function of $n$, with $c_n \to 1$ for $n \to \infty$.

For the model discussed in this report, we propose

$$c_n = \frac{c_0 + n/\xi}{1 + n/\xi},$$

with $c_0 \geq 1$ and $\xi \in \mathbb{N}_+$. It follows that $c_n = c_0$ for $n = 0$, that $c_n$ is a strictly decreasing function of $n$, and that $c_n \to 1$ for $n \to \infty$.

Together with a prior distribution values of $c_0$ and $\xi$ must be chosen. Here $c_0$ relates to the prior imprecision, through the relation

$$\Delta(\theta_m) = \Delta_x(x_m) = \frac{c_0 - 1}{c_0 + 1},$$

that leads to

$$c_0 = \frac{1 + \Delta(\theta_m)}{1 - \Delta(\theta_m)} = \frac{1 + \Delta_x(x_m)}{1 - \Delta_x(x_m)}.$$

Especially this last relation can be useful in practice to choose $c_0$ (practical elicitation of expert opinions, and assessment of lower and upper probabilities, will be discussed in a succeeding report).

For the interpretation of $\xi$ we look at the relation between the amount of information, and $c_n$. We use Walley's measure for the amount of information, described in section 2, and only use the maximal imprecision caused by $F$ and $\bar{F}$. As shown in section 4, it does not matter whether we take lower and upper cdf's for the parameter $\theta$ or the according lower and upper predictive cdf's for $X$.

Let $I_n = \Delta(\theta_m|x)^{-1} - 1$ be the amount of information available after $n$ observations, where the imprecision is based on the cdf's that result after updating of the priors, with $\theta_m$ the median of the updated distribution with cdf $F_l(\theta|x)$, as presented in section 4. Further, let $I_0 = \Delta(\theta_m)^{-1} - 1$ be the prior amount of information, with $\theta_m$ the median of $F_l(\theta)$. 

13
The above definition of $c_n$ leads to 

$$\Delta(\theta_m) = \frac{c_0^{\alpha-1}}{c_0^{\alpha} + \lambda},$$

$$\Delta(\theta_{m,x} | x) = \frac{c_0^{\alpha-1}}{c_0^{\alpha} + 2n/\xi}, \quad I_0 = \frac{2}{c_0^{\alpha-1}} \quad \text{and} \quad I_n = \frac{2(1+n/\xi)}{c_0^{\alpha-1}}.$$  

Interpretation of $\xi$ (when restricted to $\xi \in N_1$) is possible by $I_0 = 2I_0$, so $\xi$ is the number of data $x_i$ that provides an equal amount of information as the prior information does.

This proposed form of $c_n$ is only one of an infinite amount of possible choices. However, this $c_n$ satisfies all intuitively logical conditions, and has a simple form. In practice, $c_n$ must be chosen by the decision maker, the person who must provide a solution to a certain decision problem, and hereto uses prior information (for example expert opinions), and possibly experimental data. To reach a decision, a model is used with some unknown parameter, and the prior information must lead to the choice of imprecise prior densities. To apply the model presented in this report only a few parameters need to be assessed, namely the hyperparameters of the conjugate lower prior density together with $c_0$ and $\xi$ to define $c_n$. The advantage of this model is that only simple calculations are required. The decision maker must use the prior information to choose the hyperparameters and $c_0$, while he must choose $\xi$ by comparing the value he assigns to the prior information with the value of a number of independent data that may come available, for example by experiments.
6. Examples

From section 4 it results that only $F_X(\theta)$ and $F_{X_1}(x)$ are needed to determine $F(\theta)$, $F(\theta)$, $F_X(x)$ and $F_{X_1}(x)$. Because updating is done by changing the hyperparameters and $c_0$ to $c_n$, the cdf's of the conjugate prior and the according predictive for some hyperparameters $(\nu, \tau)$ are all that is necessary for statistical inference, when a distribution from the one-parameter exponential family is chosen, together with $c_n$, and the imprecise priors $1(\theta)$ and $\nu(\theta)$ are as in section 4.

In this section the necessary distributions for some members of this family are provided, with notation of distributions and some standard functions as presented in appendix 1.

1. Weibull distribution

Let $X \sim W(\alpha, \beta)$, with $\beta$ a given constant. This distribution has pdf

$$f(x) = \alpha \beta x^{\beta-1} \exp(-\alpha x^\beta), \quad \text{for } x \geq 0, \quad \alpha > 0, \quad \beta > 0.$$  

This is a member of the one-parameter exponential family, with $g(x) = \beta x^{\beta-1}$, $h(\alpha) = \alpha$, $t(x) = x^\beta$ and $\psi(\alpha) = -\alpha$.

A conjugate prior distribution is $\alpha \sim G(\nu+1, \tau)$, with $\nu \in \mathbb{N}_+$ and $\tau > 0$ [see Martz and Waller [5]], while the predictive distribution for $X$, based on this prior, can be found by the fact that the predictive distribution for $X^{1/\beta}$ is a Pareto distribution, $X^{1/\beta} \sim Pa(\nu+1, \tau)$.

The pdf of this predictive distribution is (see appendix 2)

$$f(x) = (\nu+1)\tau^{\nu+1} \beta x^{\beta-1} (\tau + x^\beta)^{-(\nu+2)}.$$  

The corresponding cdf is

$$F(x) = 1 - \left(\frac{\tau}{\tau + x^\beta}\right)^{\nu+1}.$$  

Updating, after $n$ independent observations $x = \{x_1, \ldots, x_n\}$, is done by replacing the hyperparameters $(\nu, \tau)$ by $(\nu+n, \tau+n \sum x_i^\beta)$.

The distribution $W(\alpha, 1)$ is equal to the exponential distribution $\text{Exp}(\alpha)$, so results for this distribution follow immediately.
2. Normal distribution

Firstly, let \( X \sim N(\mu, \sigma^2) \), with \( \sigma^2 > 0 \) a known constant [see Lee [4]]. This distribution has pdf \( f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right] \), for \( x \in \mathbb{R} \), with \( \mu \in \mathbb{R} \).

This is a member of the one-parameter exponential family, with
\[
g(x) = (2\pi\sigma^2)^{-1/2} \exp \left[ -\frac{x^2}{2\sigma^2} \right], \quad h(\sigma^2) = \exp \left[ -\frac{\mu^2}{2\sigma^2} \right], \quad t(\sigma^2) = \frac{\mu^2}{\sigma^2} \text{ and } \psi(\sigma^2) = \frac{\mu^2}{\sigma^2}.\]

A conjugate prior is \( \mu \sim N \left( \frac{\tau}{\nu}, \frac{\sigma^2}{\nu} \right) \), with \( \nu \in \mathbb{N}_+ \) and \( \tau \in \mathbb{R} \), while the predictive, based on this prior, is \( X \sim N \left( \frac{\tau}{\nu}, \frac{(\nu+1)\sigma^2}{\nu} \right) \).

When \( x \) come available, updating is done by replacing \( (\nu, \tau) \) by \( (\nu+n, \tau+\sum x_i) \).

Secondly, let \( X \sim N(\mu, \sigma^2) \), with \( \mu \in \mathbb{R} \) a known constant. The pdf has the form of the one-parameter exponential family with \( g(x) = (2\pi)^{-1/2} \), \( h(\sigma^2) = (\sigma^2)^{-1/2} \), \( t(\sigma^2) = (x-\mu)^2/2 \) and \( \psi(\sigma^2) = -1/\sigma^2 \). A conjugate prior is \( \sigma^2 \sim IG \left( \frac{\nu-2}{2}, \tau \right) \), with \( \nu \in \mathbb{N}_+ \setminus \{1, 2\} \) and \( \tau > 0 \).

The predictive distribution has (see appendix 3) pdf
\[
f(x) = \left[ \sqrt{2} \frac{(1-\nu/2)^{1-\nu/2}}{B \left( \frac{\nu}{2}, \frac{\nu}{2} \right)} \right] \left( \frac{(\nu+1)(x-\mu)^2}{2} \right)^{-1/2} \exp \left[ -\frac{(x-\mu)^2}{2\sigma^2} \right], \text{ for } x \in \mathbb{R}.
\]

In literature, this distribution is known as a generalized Cauchy distribution [see Rider [8]].

The according cdf is (see appendix 3):

- for \( x \leq \mu \) \( F(x) = \frac{1}{2} - \frac{1}{2} \frac{B \left( \frac{\nu}{2}, \frac{\nu}{2}, \frac{\nu-1}{2} \right)}{B \left( \frac{\nu}{2}, \frac{\nu}{2} \right)}, \text{ with } \kappa(x) = \frac{(x-\mu)^2/2}{\tau+(x-\mu)^2/2} \),

- for \( x > \mu \) \( F(x) = \frac{1}{2} + \frac{1}{2} \frac{B \left( \frac{\nu}{2}, \frac{\nu}{2}, \frac{\nu-1}{2} \right)}{B \left( \frac{\nu}{2}, \frac{\nu}{2} \right)}, \text{ with the same } \kappa(x) \).

When \( x \) come available, updating is done by replacing \( (\nu, \tau) \) by \( (\nu+n, \tau+\sum (x_i-\mu)^2/2) \).
3. Poisson distribution

Let $X \sim P(\lambda)$, for $x \in \mathbb{N}$, with $\lambda > 0$ (see Lee [4]). This distribution has pdf $p(x) = e^{-\lambda} \frac{\lambda^x}{x!}$, which is a member of the one-parameter exponential family with $g(x) = (x!)^{-1}$, $h(\lambda) = e^{-\lambda}$, $t(x) = x$ and $\psi(\lambda) = \ln \lambda$. A conjugate prior is $\lambda \sim G(\tau + 1, \nu)$, with $\nu \in \mathbb{N}_+$, $\tau \in \mathbb{N}$. The predictive distribution for $X$, based on this prior, is $X \sim NB\left(\tau + 1, \frac{\nu}{\nu + 1}\right)$. When $x$ come available, updating is done by replacing $(\nu, \tau)$ by $(\nu + n, \tau + \sum x_i)$.

4. Binomial distribution

Let $X \sim Bin(k, p)$, with $k \in \mathbb{N}_+$ constant and known (see Press [6]). This distribution has pdf $p(x) = \binom{k}{x} p^x (1-p)^{k-x}$, for $x \in \{0, 1, \ldots, k\}$, with $0 < p < 1$ (we assume $p \in (0, 1)$). This can be written in the general form by taking $g(x) = \binom{k}{x}$, $h(p) = (1-p)^k$, $t(x) = x$ and $\psi(p) = \ln\left(\frac{p}{1-p}\right)$. A conjugate prior is $p \sim Be(\tau + 1, \nu k - \tau + 1)$, with $\nu \in \mathbb{N}_+$ and $\tau \in \mathbb{N}$. The predictive distribution, based on this prior, is a Polya distribution, $X \sim Pol(k, \tau + 1, \nu k - \tau + 1)$, that is also known in literature as a beta-binomial distribution. When $x$ come available, updating is done by replacing $(\nu, \tau)$ by $(\nu + n, \tau + \sum x_i)$.

5. Negative Binomial distribution

Let $X \sim NB(k, p)$, with $k \in \mathbb{N}_+$ constant and known. This distribution has pdf $p(x) = \binom{k+x-1}{x} p^x (1-p)^{k-x}$, for $x \in \mathbb{N}$, with $0 < p < 1$ (again we assume $p \in (0, 1)$). This is a member of the one-parameter exponential family, with $g(x) = \binom{k+x-1}{x}$, $h(p) = p^k$, $t(x) = x$ and $\psi(p) = \ln(1-p)$. A conjugate prior is $p \sim Be(\nu k + 1, \tau + 1)$, with $\nu \in \mathbb{N}_+$ and $\tau \in \mathbb{N}$. The predictive pdf, based on this prior, is found by using the general form: $p(x) \propto \binom{k+x-1}{x} B((\nu+1)k+1, \tau + x + 1)$. 

17
In appendix 4 this distribution, known in literature as a Beta-Pascal distribution (see Raiffa and Schlaifer [7]), is discussed, and it is shown that \( p(x) = \frac{k+x-1}{x} \frac{\mathbf{B}(\nu+1)k+1,\tau+x+1)}{\mathbf{B}(\nu k+1,\tau+1)}. \)

For our purposes, however, the according cdf is needed. This is not available in a simple form, so term-by-term calculation of all pdf values is needed. In appendix 4 some recursion relations are given that simplify these calculations.

When \( x \) come available, updating is done by replacing \((\nu,\tau)\) by \((\nu+n,\tau+\sum_{i=1}^{n}x_i)\).

6. Gamma distribution

Let \( X \sim \Gamma(a,\beta) \), with \( a>0 \) a given constant. This distribution has pdf
\[
f(x) = \beta(\beta x)^{a-1}e^{-\beta x} / \Gamma(a), \quad \text{for } x \geq 0, \text{ with } \beta > 0.
\]
This is a member of the one-parameter exponential family, with \( g(x)=x^{a-1}/\Gamma(a) \), \( h(\beta)=\beta^a \), \( t(x)=x \) and \( \psi(\beta)=-\beta \).

A conjugate prior distribution is \( \beta \sim \Gamma(\alpha v+1,\tau) \), with \( \nu \in \mathbb{N}_+ \) and \( \tau \geq 0 \). The corresponding predictive distribution (see appendix 5) has pdf
\[
f(x) = \frac{\tau^{\alpha v+1}x^{\alpha-1}}{\mathbf{B}((\alpha,\alpha v+1)+x+1)} \quad \text{for } x \geq 0, \text{ and corresponding cdf}
\]
\[F(x) = \frac{\mathbf{B}(x,\alpha v+1)}{\mathbf{B}(\alpha,\alpha v+1)}.
\]
When \( x \) come available, updating is done by replacing \((\nu,\tau)\) by \((\nu+n,\tau+\sum_{i=1}^{n}x_i)\).

We remark that also the gamma distribution with known \( \beta \) belongs to this family of distributions, with \( g(x)=\beta(\beta x)^{a-1}e^{-\beta x} \), \( h(\alpha)=1/\Gamma(\alpha) \), \( t(x)=\ln(\beta x) \) and \( \psi(\alpha)=a-1 \). For this situation the results are less attractive, and in practice this model is rarely of interest.
7. Concluding Remarks

In this report a simple general form for imprecise prior densities is proposed for members of the one-parameter exponential family of distributions. These densities have the form of the pdf of some member of a conjugate family, which enables updating by simple means. These models can be regarded as first suggestions for statistical models with imprecise prior densities, a generalization of the standard Bayesian framework of statistical inference.

The amount of imprecision should reflect the lack of knowledge about the prior distribution, and a simple function $c_n$ is proposed that describes the degree of imprecision after new data have come available. This function depends on the prior degree of imprecision, and on the number of data that is supposed to provide an amount of information equal to the prior information. The decision maker must choose values for two interpretable parameters to define this function.

It is clear that the use of these models, and also of the entire concept of imprecise probabilities, can only be evaluated by practical application. The most important task for future research in this area is to work out all necessary steps of a suitable procedure to apply this method to practical decision problems.

To this end methods to elicit expert opinions are very important. In a following report this will be discussed for the situation that the random variable of interest represents the lifetime of a component.

It is also necessary to develop more models, but for practical use it is useful, although not necessary, to keep the necessary amount of calculations (e.g. in case of updating) small. To this end imprecise conjugate priors are attractive.

The role of the decision maker, when models as proposed here are used, should also be studied in practice. Especially a comparative (case-) study of possible alternatives to our suggested form of $c_n$ will be useful.

Finally, the use of this concept to solve decision problems needs to be analyzed. Here decisions might have a statistical nature (estimators, tests of hypotheses) or immediate practical importance (e.g. when a good solution to a certain decision problem is needed).
References


Appendix 1  Standard functions and distributions

We give an overview of standard functions and distributions, used in this report.

1. Gamma function
\[ \Gamma(z) = \int_0^\infty u^{z-1} e^{-u} du, \quad z > 0. \]

2. Incomplete Gamma function
\[ \Gamma_x(z) = \int_0^x u^{z-1} e^{-u} du, \quad z > 0, \quad x > 0. \]

3. Beta function
\[ B(z, w) = \int_0^1 u^{z-1} (1-u)^{w-1} du, \quad z > 0, \quad w > 0. \quad B(z, w) = \frac{\Gamma(z) \Gamma(w)}{\Gamma(z+w)}. \]

4. Incomplete Beta function
\[ B_x(z, w) = \int_0^x u^{z-1} (1-u)^{w-1} du, \quad z > 0, \quad w > 0, \quad 0 < x \\leq 1. \]

5. Normal distribution \( X \sim N(\mu, \sigma^2) \)
\[ f(x) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad x \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0. \]
\[ F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \text{ with } \Phi \text{ the cdf of the standard normal distribution}, \]
\[ \Phi(z) = \int_{-\infty}^{z} (2\pi)^{-1/2} \exp\left[-\frac{x^2}{2}\right] dx, \quad z \in \mathbb{R}. \]

6. Exponential distribution \( X \sim \text{Exp}(\lambda) \)
\[ f(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \lambda > 0. \]
\[ F(x) = 1 - e^{-\lambda x}. \]

7. Weibull distribution \( X \sim W(\alpha, \beta) \)
\[ f(x) = \alpha \beta x^{\beta-1} \exp(-\alpha x^\beta), \quad x \geq 0, \alpha > 0, \beta > 0. \]
\[ F(x) = 1 - \exp(-\alpha x^\beta). \]
8. Gamma distribution \( X - G(\alpha, \beta) \)
\[ f(x) = \beta (\beta x)^{\alpha-1} e^{-\beta x} / \Gamma(\alpha), \quad x \geq 0, \quad \alpha > 0, \quad \beta > 0. \]
\[ F(x) = \frac{\Gamma(\beta x (\alpha) / \Gamma(\alpha)). \]

9. Inverted Gamma distribution \( X - IG(\alpha, \beta) \)
\[ f(x) = \beta x^{-(a+1)} \exp(-\beta / x) / \Gamma(\alpha), \quad x > 0, \quad \alpha > 0, \quad \beta > 0. \]
\[ F(x) = 1 - \frac{\Gamma(\beta / x (\alpha) / \Gamma(\alpha)). \]

By comparing this distribution (see Martz and Waller [5, p.101]) to the inverse Chi-squared distribution (see Lee [4, p.236-237]) it can be seen that these two are identical, only with different parameters. Some authors prefer to work with the inverse Chi-squared distribution.

10. Pareto distribution \( X - Pa(\alpha, \gamma) \)
\[ f(x) = \alpha \gamma \alpha^{\alpha+1} (\gamma + x)^{-\alpha+1}, \quad x \geq 0, \quad \alpha > 0, \quad \gamma > 0. \]
\[ F(x) = 1 - \left( \frac{\gamma}{\gamma + x} \right)^\alpha. \]

We have chosen this definition of the Pareto distribution for convenience. Usually \( Y = X + \gamma \) is said to have a pareto distribution, in which case
\[ f_y(y) = \alpha \gamma \alpha^{\alpha+1}, \quad \text{for } y \geq \gamma. \]

11. Beta distribution \( X - Be(\alpha, \beta) \)
\[ f(x) = x^{\alpha-1} (1-x)^{\beta-1} / B(\alpha, \beta), \quad 0 < x < 1, \quad \alpha > 0, \quad \beta > 0. \]
\[ F(x) = B_x (\alpha, \beta) / B(\alpha, \beta). \]

12. Polya distribution \( X - Pol(n, r, s) \)
\[ p(x) = \left( \frac{n}{x} \right) B(r+x, s+n-x) / B(r, s), \quad x \in \{0, 1, \ldots, n\}, \quad n \in \mathbb{N}^+, \quad r > 0, \quad s > 0. \]
(We use \( \mathbb{N} = \{0, 1, \ldots\} \), and \( \mathbb{N}^+ = \mathbb{N} \setminus \{0\} \).
This distribution is also known as beta-binomial distribution.

13. Poisson distribution \( X - P(\lambda) \)
\[ p(x) = e^{-\lambda} \lambda^x / x!, \quad x \in \mathbb{N}, \quad \lambda > 0. \]

22
14. **Binomial distribution** \(X \sim \text{Bin}(k, p)\)

\[ p(x) = \binom{k}{x} p^x (1-p)^{k-x}, \quad x \in \{0, 1, \ldots, k\}, \; k \in \mathbb{N}_+, \; 0 \leq p \leq 1. \]

15. **Negative Binomial distribution** \(X \sim \text{NB}(k, p)\)

(In statistics, the name 'negative binomial distribution' is used for random variables with several different definitions. We define \(X\) to be the number of failures until the \(k\)th success.)

\[ p(x) = \binom{k+x-1}{x} p^x (1-p)^{k-x}, \quad x \in \mathbb{N}, \; k \in \mathbb{N}_+, \; 0 \leq p \leq 1. \]

**Appendix 2** 

**Predictive distribution for the Weibull distribution with known \(\beta\), and gamma prior for \(\alpha\).**

The predictive pdf is,

\[ f(x) \propto \beta x^{-1} \int_0^\infty \alpha^\nu x \exp(-\alpha(x^\beta)) dx = \Gamma(\nu+2) \beta x^{\beta-1} (x^\beta)^{-(\nu+2)}, \quad \text{for } x \geq 0, \text{ where} \]

the equality can be proven by defining \(u = \alpha(x^\beta)\) in the integral. We know that \(f(x) = c_N x^{\beta-1} (x^\beta)^{-(\nu+2)}\), with \(c_N\) the normalizing constant,

\[ c_N = \left[ \int_0^\infty x^{\beta-1} (x^\beta)^{-(\nu+2)} dx \right]^{-1} = \beta (\nu+1) (\nu+1). \]

For this last equality, we must use the following result:

Let \(m > 0\), \(a > -1\), \(b > 0\) and \(c > \frac{a+1}{b}\), then

\[ \int_0^\infty x^{a(m+x^b)} y^{-c} dy = \int_0^\infty \frac{a}{b} \left( \frac{m+y}{1-y} \right)^{-c} \left( \frac{m}{1-z} \right)^{-2} dz = \frac{a+1}{b} \left( \frac{a+1}{b} \right)^{-c} \left( \frac{a+1}{b} \right)^{-c} \left( \frac{a+1}{b} \right)^{-c} \left( \frac{a+1}{b} \right)^{-c}. \]

In calculating the above integral, we first define \(y = x^\beta\), followed by \(z = \frac{y}{m+y}\).

We use this result to calculate \(c_N\), by taking \(a = \beta - 1\), \(m = \tau\), \(b = \beta\) and \(c = \nu+2\), for which values all conditions are fulfilled.

The corresponding cdf is

\[ F(x) = \int_0^x (\nu+1) \tau^{\nu+1} \omega^{\beta-1} (\tau+\omega^\beta)^{-(\nu+2)} d\omega = 1 - \left( \frac{\tau+x^\beta}{\tau+\beta} \right)^{\nu+1}. \]
Appendix 3  
Predictive distribution for the normal distribution with known \( \mu \), and inverted gamma prior for \( \sigma^2 \).

Using the general results for members of the one-parameter exponential family of distribution, we derive the predictive pdf:

\[
f(x) \propto \int_0^\infty (\sigma^2)^{-\frac{(v+1)}{2}} \exp\left[-\frac{(\tau+(x-\mu)^2)}{2\sigma^2}\right] \, d\sigma^2 = \Gamma\left(\frac{v-1}{2}\right) \left(\tau+(x-\mu)^2/2\right)^{-\frac{(v-1)}{2}},
\]

where the equality can be proven by using the pdf of an inverted gamma distribution. Hence \( f(x) = c \left(\tau+(x-\mu)^2/2\right)^{-\frac{(v-1)}{2}} \), with normalizing constant \( c \) such that

\[
c^{-1} = \int_{-\infty}^{\infty} \left(\tau+(x-\mu)^2/2\right)^{-\frac{(v-1)}{2}} dx = \sqrt{2\tau} \left(\frac{1}{2}, \frac{v-1}{2}\right) B\left[\frac{1}{2}, \frac{v}{2-1}\right].
\]

This result is derived by calculating the integral over the interval \([\mu, \infty)\), and multiplying this by two (the function is symmetric around \( \mu \)). The integral \( \int_{-\infty}^{\infty} \left(\tau+(x-\mu)^2/2\right)^{-\frac{(v-1)}{2}} dx \) is solved by first defining \( y=(x-\mu)^2/2 \),

and then using the standard integral presented in appendix 2, with \( a=-1/2 \), \( \alpha=\tau \), \( \beta=1 \) and \( c=(v-1)/2 \).

This leads to the predictive pdf

\[
f(x) = \sqrt{2\tau} \left(\frac{1}{2}, \frac{v}{2-1}\right) B\left[\frac{1}{2}, \frac{v-1}{2}\right] \left(\tau+(\frac{x}{2})^2\right)^{-\frac{(v-1)}{2}}, \text{ for } x \in \mathbb{R}.
\]

To calculate the according cdf, another standard integral is needed, that is analogous to the integral above and in appendix 2, but now on a finite interval. This standard integral is:

\[
\int_0^x \frac{a}{m+x} \, dx = b \left(\frac{a+1-c}{m}\right) B\left(\frac{a+1-c}{m'+b}\right), \text{ for } k \geq 0, \text{ and the same conditions for } a, m, b \text{ and } c \text{ as in appendix 2. This integral is proven in the same way as the integral in appendix 2, and the only difference is that the Beta function is replaced by the incomplete Beta function. The cdf is easily found by using the same method used to calculate the normalizing constant for the pdf:}
\]
for $x \leq \mu$ \[ F(x) = \frac{1}{2} - \frac{1}{2} \left[ \frac{B(\kappa(x) \left[ \frac{1}{2}, \frac{\nu-1}{2} \right])}{B\left[ \frac{1}{2}, \frac{\nu-1}{2} \right]} \right], \quad \text{with} \quad \kappa(x) = \frac{(x-\mu)^2/2}{\tau+(x-\mu)^2/2}, \]

for $x > \mu$ \[ F(x) = \frac{1}{2} + \frac{1}{2} \left[ \frac{B(\kappa(x) \left[ \frac{1}{2}, \frac{\nu-1}{2} \right])}{B\left[ \frac{1}{2}, \frac{\nu-1}{2} \right]} \right], \quad \text{with the same} \quad \kappa(x). \]

In Johnson and Kotz [3] a generalization of the Cauchy distribution is shortly discussed, based on a paper by Rider [8]. The generalized Cauchy distribution is defined by its pdf:

\[ f(x) = \frac{k \Gamma(h)}{2 \lambda \Gamma(k^{-1}) \Gamma(h-k^{-1})} \left[ 1 + \left( \frac{x-\xi}{\lambda} \right)^{k-1} \right]^{-h}, \quad \text{for} \quad x \in \mathbb{R}, \quad \text{with} \quad \lambda, \; k \; \text{and} \; h \; \text{all positive.} \]

The predictive pdf derived above has this form, with $\xi=\mu$, $\lambda=\sqrt{2\tau}$, $k=2$ and $h=(\nu-1)/2$.

**Appendix 4**

*Predictive distribution for the negative binomial distribution with beta prior for $p$.*

Using the general form of the predictive pdf for a member of the one-parameter exponential family of distributions, based on a conjugate prior, we get

\[ p(x) \propto \left( \frac{x+1}{x} \right)^{(k+x-1)} B((\nu+1)k+1, \tau+x+1). \]

Raiffa and Schlaifer [7] discuss the beta-Pascal distribution, which has pdf

\[ p(n|r',n',r) = \frac{(r+r'-1)! (n+n'-r-r'-1)! (n-1)! (n'-1)!}{(r-1)! (r'-1)! (n-r)! (n'-r'-1)! (n+n'-1)!}, \quad \text{for} \quad n \in \mathbb{N}_+, \; r \in \mathbb{N}_+, \; n \geq r, \; n' > r' > 0. \]

The predictive pdf can be written in the same form, by defining $n=x+\tau+1$, $r=\tau+1$, $n'=\nu+1)k+1$ and $r'=\nu k+1$. Then the beta-Pascal pdf is a normalized mass function for $X \in \mathbb{N}$, by defining $p_X(x)=p(x+\tau+1|\tau, \nu, k)$ for all $x \in \mathbb{N}$. Hence we derive predictive pdf

\[ p(x) = \left( \frac{x+1}{x} \right)^{(k+x-1)} B((\nu+1)k+1, \tau+x+1)/B(\nu k+1, \tau+1). \]
To calculate the according cdf one has to compute all necessary pdf values term-by-term. Raiffa and Schlaifer [7] provide the following recursion relations that simplify these computations:

\[
p(n | r', n', r) = \frac{(n-1)(n+n'-r-x'-1)}{(n+n'-1)(n-r)} p(n-1 | r', n', r), \text{ and also } \]

\[
p(n | r', n', r) = \frac{(n+n')(n+1-r)}{n(n+n'-r-x')} p(n+1 | r', n', r).
\]

As a base of the recursion, one term must be evaluated by use of the complete formula for \( p(n | r', n', r) \).

**Appendix 5**  
*Predictive distribution for the gamma distribution with known \( \alpha \), and gamma prior for \( \beta \).*

Using the general results, we derive for the predictive pdf:

\[
f(x) \propto \frac{x^{\alpha-1}}{(\tau+x)^{\alpha \nu + \alpha + 1}} \text{ for } x \geq 0.
\]

To calculate the normalizing constant, the standard integral of appendix 2 is used, with \( a=\alpha-1 \), \( b=1 \), \( c=\alpha \nu + \alpha + 1 \) and \( m=\tau \).

This leads to

\[
f(x) = \frac{\tau^{\alpha \nu + \alpha + 1} \tau^{\alpha-1}}{B(\alpha, \alpha \nu + 1)(\tau+x)^{\alpha \nu + \alpha + 1}} \text{ for } x \geq 0.
\]

By rewriting this pdf to

\[
f(x) = \frac{\tau}{B(\alpha, \alpha \nu + 1)(\tau+x)^2} \left( \frac{x}{\tau+x} \right)^{\alpha-1} \left( \frac{\tau}{\tau+x} \right)^{\alpha \nu},
\]

it shows that it relates to a Beta distribution.

To derive the cdf, the standard integral of appendix 3 has to be used, with the same parameters as above, leading to

\[
F(x) = \frac{B_x(\alpha, \alpha \nu + 1)}{B(\alpha, \alpha \nu + 1)}.
\]
List of COSOR-memoranda - 1991

<table>
<thead>
<tr>
<th>Number</th>
<th>Month</th>
<th>Author</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td>91-03</td>
<td>January</td>
<td>M.W.P. Savelsbergh</td>
<td>The vehicle routing problem with time windows: minimizing route duration.</td>
</tr>
<tr>
<td>91-04</td>
<td>January</td>
<td>M.W.I. van Kraaij</td>
<td>Some considerations concerning the problem interpreter of the new manpower planning system formasy.</td>
</tr>
<tr>
<td>91-06</td>
<td>March</td>
<td>R.J.G. Wilms</td>
<td>Properties of Fourier-Stieltjes sequences of distribution with support in [0,1].</td>
</tr>
<tr>
<td>91-07</td>
<td>March</td>
<td>F. Coolen, R. Dekker, A. Smit</td>
<td>Analysis of a two-phase inspection model with competing risks.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td>91-10</td>
<td>May</td>
<td>P.J. Zwietering, E.H.L. Aarts,</td>
<td>Sorting With A Neural Net.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>J. Wessels</td>
<td></td>
</tr>
<tr>
<td>91-11</td>
<td>May</td>
<td>F. Coolen</td>
<td>On some misconceptions about subjective probability and Bayesian inference.</td>
</tr>
<tr>
<td>Date</td>
<td>Month</td>
<td>Authors</td>
<td>Topic</td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>91-12</td>
<td>May</td>
<td>P. van der Laan</td>
<td>Two-stage selection procedures with attention to screening.</td>
</tr>
<tr>
<td>91-14</td>
<td>June</td>
<td>J. Korst, E. Aarts, J.K. Lenstra, J. Wessels</td>
<td>Periodic assignment and graph colouring.</td>
</tr>
<tr>
<td>91-16</td>
<td>July</td>
<td>P. Deheuvels, J.H.J. Einmahl</td>
<td>Approximations and Two-Sample Tests Based on P - P and Q - Q Plots of the Kaplan-Meier Estimators of Lifetime Distributions.</td>
</tr>
<tr>
<td>91-19</td>
<td>August</td>
<td>P. van der Laan</td>
<td>The efficiency of subset selection of an almost best treatment.</td>
</tr>
<tr>
<td>91-20</td>
<td>September</td>
<td>P. van der Laan</td>
<td>Subset selection for an ( \alpha )-best population: efficiency results.</td>
</tr>
<tr>
<td>91-22</td>
<td>September</td>
<td>R.J.M. Vaessens, E.H.L. Aarts, J.H. van Lint</td>
<td>Genetic Algorithms in Coding Theory - A Table for ( A_1(n,d) ).</td>
</tr>
<tr>
<td>91-23</td>
<td>September</td>
<td>P. van der Laan</td>
<td>Distribution theory for selection from logistic populations.</td>
</tr>
</tbody>
</table>
91-24 October I.J.B.F. Adan
J. Wessels
W.H.M. Zijm
Matrix-geometric analysis
of the shortest queue
problem with threshold
jockeying.

91-25 October I.J.B.F. Adan
J. Wessels
W.H.M. Zijm
Analysing Multiprogramming
Queues by Generating
Functions.

91-26 October E.E.M. van Berkum
P.M. Upperman
D-optimal designs for an
incomplete quadratic model.

91-27 October R.P. Gilles
P.H.M. Ruys
S. Jilin
Quasi-Networks in Social
Relational Systems.

91-28 October I.J.B.F. Adan
J. Wessels
W.H.M. Zijm
A Compensation Approach for
Two-dimensional Markov
Processes.

91-29 November J. Wessels
Tools for the Interfacing
Between Dynamical Problems
and Models within Decision
Support Systems.

91-30 November G.L. Nemhauser
M.W.P. Savelsbergh
G.C. Sigismondi
Constraint Classification
for Mixed Integer Program­
ing Formulations.

91-31 November J.Th.M. Wijnen
Taguchi Methods.

91-32 November F.P.A. Coolen
The Theory of Imprecise
Probabilities: Some Results
for Distribution Functions,
Densities, Hazard Rates and
Hazard Functions.

91-33 December S. van Hoesel
A. Wagelmans
On the P-coverage problem
on the real line.

91-34 December S. van Hoesel
A. Wagelmans
On setup cost reduction in
the economic lot-sizing
model without speculative
motives.

91-35 December S. van Hoesel
A. Wagelmans
On the complexity of post-
optimality analysis of
0/1 programs.

91-36 December F.P.A. Coolen
Imprecise Conjugate Prior
Densities for the One-
Parameter Exponential
Family of Distributions.