How to Make Plausibility-Based Forecasting More Accurate

Kongliang Zhu, Nantiworn Thianpaen, and Vladik Kreinovich

Abstract In recent papers, a new plausibility-based forecasting method was proposed. While this method has been empirically successful, one of its steps – selecting a uniform probability distribution for the plausibility level – is heuristic. It is therefore desirable to check whether this selection is optimal or whether a modified selection would like to a more accurate forecast. In this paper, we show that the uniform distribution does not always lead to (asymptotically) optimal estimates, and we show how to modify the uniform-distribution step so that the resulting estimates become asymptotically optimal.

1 Plausbility-Based Forecasting: Description, Successes, and Formulation of the Problem

Need for prediction. One of the main objectives of science is, given the available data $x_1, \ldots, x_n$, to predict future values of different quantities y.

The usual approach to solving this problem consists of two stages:

- first, we find a *model* that describes the observed data; and
- then, we use this model to predict the future value of each of the quantities y.

In some cases, it is sufficient to have a *deterministic model*, that describes the dependence of each observed value on the known values describing the i-th observation

Kongliang Zhu
Faculty of Economics, Chiang Mai University, Thailand, e-mail: 258zkl@gmail.com

Nantiworn Thianpaen
Faculty of Economics, Chiang Mai University, Thailand, and Faculty of Management Sciences, Suratthani Rajabhat University, Thailand, e-mail: nantiworn@outlook.com

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University, El Paso, Texas 79968, USA, e-mail: vladik@utep.edu

and on the (unknown) parameters p of the model: $x_i = f_i(p)$. In this case, we can predict the value y as $y = f(p)$ for an appropriate function $f(p)$.

For example, in Newtonian's model of the Solar system, once we know the initial locations, initial velocities, and masses of all the celestial bodies (which, in this case, are the parameters p), we can predict the position and velocity of each body at any future moment of time.

In this deterministic case, we can use the known observed values to estimate the parameters p of the corresponding probabilistic model, and then we can use these parameters to predict the desired future values. This is how, e.g., solar eclipses can be predicted for centuries ahead.

Need for statistical prediction. In most practical problems, however, a fully deterministic prediction is not possible, since, in addition to the parameters p, both the observed values x_i and the future value y are affected by other parameters beyond our control, parameters that can be viewed as *random*. Thus, instead of a deterministic model, we have a general *probabilistic model* $x_i = f(p, z_1, \ldots, z_m)$ and $y = f(p, z_1, \ldots, z_m)$, where z_j are random variables.

Usually, we do not know the exact probability distribution for the variables z_i, but we know a finite-parametric family of distributions that contains the actual (unknown) distribution. For example, we may know that the distribution is Gaussian, or that it is uniform. Let q denote the parameter(s) that describe this distribution.

In this case, both x_i and y are random variables whose distribution depends on all the parameters $\theta = (p, q)$: $x_i \sim f_{i,\theta}$ and $y \sim f_\theta$.

In this case, to identify the model:

- we first estimate the parameters θ based on the observations $x_1, \ldots, x_n$, and then
- we use the distribution f_θ corresponding to these parameter values to predict the values y – or, to be more precise, to predict the probability of different values of y.

Need for a confidence interval. Since in the statistical case, we cannot predict the *exact* value of y, it is desirable to predict the *range* of possible values of y.

For many distributions – e.g., for a (ubiquitous) normal distribution – it is, in principle, possible to have arbitrarily small and arbitrarily large values, just the probability of these values is very small. In such situations, there is no *guaranteed* range of values of y.

However, we can still try to estimate a *confidence interval*, i.e., for a given small value $\alpha > 0$, an interval $[\underline{y}_\alpha, \overline{y}_\alpha]$ that contains the actual value y with confidence $1 - \alpha$. In other words, we would like to find an interval for which $\text{Prob}(y \in [\underline{y}_\alpha, \overline{y}_\alpha]) \geq \alpha$.

In the idealized situation, when we know the probabilities of different values of y – i.e., in precise terms, when we know the corresponding cumulative distribution function (cdf) $F(y) \stackrel{\text{def}}{=} \text{Prob}(Y \leq y)$ – then we know that $Y \leq F^{-1}\left(\frac{\alpha}{2}\right)$ with probability $\alpha/2$ and that $Y > F^{-1}\left(1 - \frac{\alpha}{2}\right)$ with probability $\alpha/2$. Thus, with probability $1 - \alpha$, we have $y \in [\underline{y}_\alpha, \overline{y}_\alpha]$, where $\underline{y}_\alpha = F^{-1}\left(\frac{\alpha}{2}\right)$ and

$$\overline{y}_\alpha = F^{-1}\left(1-\frac{\alpha}{2}\right).$$

In general, a statistical estimate based on a finite sample is only approximate. Thus, based on a finite sample, we can predict the value of the parameters θ only approximately – and therefore, we only have an approximate estimate of the probabilities of different values of y. So, instead of the actual cdf $F(y)$, we only know the bounds on the cdf: $\underline{F}(y) \le F(y) \le \overline{F}(y)$. We want to select the interval $[\underline{y}_\alpha, \overline{y}_\alpha]$ in such a way that the probability of being outside this interval is guaranteed not to exceed α.

For the lower bound $\underline{y}_\alpha$, all we know about the probability $F(\underline{y}_\alpha)$ of being smaller than this bound is that this probability is bounded, from above, by the known value $\overline{F}(\underline{y}_\alpha)$: $F(\underline{y}_\alpha) \le \overline{F}(\underline{y}_\alpha)$. Thus, to guarantee that this probability does not exceed $\dfrac{\alpha}{2}$, we must select a bound $\underline{y}_\alpha$ for which $\overline{F}(\underline{y}_\alpha) = \dfrac{\alpha}{2}$. In other words, we should take

$$\underline{y}_\alpha = (\overline{F})^{-1}\left(\frac{\alpha}{2}\right),$$

Similarly, the probability $1 - F(\overline{y}_\alpha)$ of being larger than the upper bound $\overline{y}_\alpha$ is bounded, from above, by the known value $1 - \underline{F}(\overline{y}_\alpha)$: $1 - F(\overline{y}_\alpha) \le \underline{F}(\overline{y}_\alpha)$. Thus, to guarantee that this probability does not exceed $\dfrac{\alpha}{2}$, we must select a bound $\overline{y}_\alpha$ for which $1 - \underline{F}(\overline{y}_\alpha) = \dfrac{\alpha}{2}$. In other words, we should take

$$\overline{y}_\alpha = (\underline{F})^{-1}\left(1-\frac{\alpha}{2}\right).$$

Plausibility-based forecasting: a brief reminder. In [1, 3, 4], a new forecasting method was proposed. In this method, we start by forming a *likelihood function*, i.e., a function that describes, for each possible value θ, the probability (density) of observing the values $x = (x_1, \ldots, x_n)$. If we assume that the probability density function corresponding to each observation x_i has the form $f_{i,\theta}(x_i)$, then, under the natural assumption that the observations $x_1, \ldots, x_n$ are independent, we conclude that:

$$L_x(\theta) = \prod_{i=1}^{n} f_{\theta_i}(x_i).$$

The likelihood function is normally used to find the *maximum likelihood* estimate for the parameters θ, i.e., the estimate $\hat{\theta}$ for which $L_x\left(\hat{\theta}\right) = \max\limits_{\theta} L_x(\theta)$.

In the plausibility-based approach to forecasting, instead of simply computing this value $\hat{\theta}$, we use the likelihood function to define the *plausibility function* as

$$\mathrm{pl}_x(\theta) = \frac{L_x(\theta)}{\sup\limits_{\theta'} L_x(\theta')} = \frac{L_x(\theta)}{L_x\left(\hat{\theta}\right)}.$$

Based on this plausibility function, we define, for each real number $\omega \in [0,1]$, a *plausibility region*

$$\Gamma_x(\omega) = \{\theta : \mathrm{pl}_x(\theta) \geq \omega\}.$$

We then represent a probability distribution for y as $y = g(\theta, z)$ for an auxiliary variable z whose distribution does not depend on θ. Usually, as z, we select a random variable which is uniformly distributed on the interval $[0,1]$. Such a representation is possible for each random variable with a probability density function $f_\theta(y)$ and corresponding cumulative distribution function $F_\theta(y)$: namely, we can simply take $g(\theta,z) = F_\theta^{-1}(z)$, where F^{-1} denotes an inverse function, i.e., a function for which $F_\theta^{-1}(F_\theta(x)) = x$ for all x.

Based on the plausibility regions, we then compute the belief and plausibility of each set A of possible values of θ as follows:

$$\mathrm{Bel}(A) = \mathrm{Prob}(g(\Gamma_x(\omega), z) \subseteq A)$$

and

$$\mathrm{Pl}(A) = \mathrm{Prob}(g(\Gamma_x(\omega), z) \cap A \neq \emptyset),$$

where both ω and z are uniformly distributed on the interval $[0,1]$. After that, we compute the lower and upper bounds on the cdf $F(y)$ for y as

$$\underline{F}(y) = \mathrm{Bel}((-\infty, y])$$

and

$$\overline{F}(y) = \mathrm{Pl}((-\infty, y]).$$

Then, for any given small value $\alpha > 0$, we predict that y is, with confidence $1 - \alpha > 0$, contained in the interval $[\underline{y}_\alpha, \overline{y}_\alpha]$, where $\underline{y}_\alpha = (\overline{F})^{-1}\left(\dfrac{\alpha}{2}\right)$ and $\overline{y}_\alpha = (\underline{F})^{-1}\left(1 - \dfrac{\alpha}{2}\right)$.

Remaining problem. While the new approach has led to interesting applications, the motivations for this approach are not very clear. To be more precise:

- it is clear why, to simulate z, we use a uniform distribution on the interval $[0,1]$ – because we represent the corresponding probabilistic model for y as $y = g(\theta, z)$ for exactly this distribution for z;
- what is less clear is why we select a uniform distribution for ω.

Yes, this ω-distribution sounds like a reasonable idea: we know that ω is located on the interval $[0,1]$, we do not know which values ω are more probable and which are less probable, so we select a uniform distribution. However, since we are not just making reasonable estimates, we are making predictions with confidence, it is desirable to come up with a more convincing justification for selecting the probability distribution for ω: namely, a justification that would explain why we believe that the predicted value y belongs to the above-constructed confidence interval.

Maybe we can get a justification, or maybe we can conclude that the above interval is only an approximation – and by selecting a different probability distribution for ω, we can make the resulting forecasting more accurate.

This is the problem that we will be analyzing in this paper.

2 On a Simple Example, Let Us Compare Plausibility-Based Forecasting With Known Methods

Let us consider the simplest possible situation. To analyze this problem, let us consider the simplest case:

- when we have only one parameter $\theta = \theta_1$,
- when the predicted value y simply coincides with the value of this parameter, i.e., when the probabilistic model $y = g(\theta, z)$ has the form $g(\theta, z) = \theta$, and
- when the likelihood $L_x(\theta)$ is continuous and strictly decreasing as we move away from the maximum likelihood estimate $\hat{\theta}$; in other words, we assume that:
  - for $\theta \le \hat{\theta}$ the likelihood function strictly increases, while
  - for $\theta \ge \hat{\theta}$ the likelihood function strictly decreases.

Comment. While the first two conditions are really restrictive, the third condition – monotonicity – is not very restrictive, it is true in the overwhelming majority of practical situations.

Let us analyze this simplest possible situation. Since in our case, $y = \theta$, the desired bounds on the predicted value y are simply bounds on the value θ of the corresponding parameter, bounds that contain θ with a given confidence α. In other words, what we want is a traditional confidence interval for θ.

In the above simplest possible situation, we can explicitly express the resulting confidence interval in terms of the likelihood function. According to the plausibility-based forecasting method, we select

$$\underline{F}(y) = \text{Bel}((-\infty, y]) = \text{Prob}(\{\theta : \text{pl}_x(\theta) \ge \omega\} \subseteq (-\infty, y]).$$

Since we assumed that the likelihood function $L_x(\theta)$ is increasing for $\theta \le \hat{\theta}$ and decreasing for $\theta \ge \hat{\theta}$, the plausibility function $\text{pl}_x(\theta)$ – which is obtained by $L_x(\theta)$ by dividing by a constant – also has the same property:

- the function $\text{pl}_x(\theta)$ is increasing for $\theta \le \hat{\theta}$, and
- the function $\text{pl}_x(\theta)$ is decreasing for $\theta \ge \hat{\theta}$.

In this case, the set $\{\theta : \text{pl}_x(\theta) \ge \omega\}$ is simply an interval $[\theta^-, \theta^+]$, whose endpoints can be described as follows:

- the lower endpoint θ^- is the value to the left of $\hat{\theta}$ for which $\text{pl}_x(\theta) = \omega$, and
- the upper endpoint θ^+ is the value to the right of $\hat{\theta}$ for which $\text{pl}_x(\theta) = \omega$.

In these terms, the condition that the set $\{\theta : \mathrm{pl}_x(\theta) \geq \omega\} = [\theta^-, \theta^+]$ is contained in $(-\infty, y]$ simply means that $\theta^+ \leq y$.

Since $\theta^+ \geq \hat{\theta}$, we thus have $y \geq \hat{\theta}$ as well. The plausibility function is strictly decreasing for $\theta \geq \hat{\theta}$, the inequality $\theta^+ \leq y$ is equivalent to $\mathrm{pl}(\theta^+) \geq \mathrm{pl}(y)$. By the construction of the value θ^+, we know that $\mathrm{pl}(\theta^+) = \omega$. Thus, the condition $\{\theta : \mathrm{pl}_x(\theta) \geq \omega\} \subseteq (-\infty, y]$ is simply equivalent to $\omega \geq \mathrm{pl}_x(y)$. Hence,

$$\underline{F}(y) = \mathrm{Prob}(\omega \geq \mathrm{pl}_x(y)).$$

When ω is uniformly distributed on the interval $[0,1]$, then, for all z, the probability $\mathrm{Prob}(\omega \geq z)$ that ω is in the interval $[z,1]$, is simply equal to the width of bthis interval, i.e., to $1 - z$. In particular, for $z = \mathrm{pl}_x(y)$, we have $\underline{F}(y) = 1 - \mathrm{pl}_x(y)$. In these terms, in the plausibility-based forecasting method, as the upper bound $\overline{\theta}_\alpha$ of the confidence interval, we select the value $\overline{\theta}_\alpha$ for which $1 - \mathrm{pl}_x(\overline{\theta}_\alpha) = 1 - \frac{\alpha}{2}$, i.e., for which

$$\mathrm{pl}_x(\overline{\theta}_\alpha) = \frac{\alpha}{2}.$$

Similarly, the condition that the set $\{\theta : \mathrm{pl}_x(\theta) \geq \omega\} = [\theta^-, \theta^+]$ has a non-empty intersection with $(-\infty, y]$ simply means that $\theta^- \leq y$.

Since $\theta^- \leq \hat{\theta}$, this inequality is always true for $y \geq \hat{\theta}$. So, for $y \geq \hat{\theta}$, we have $\overline{F}(y) = 1$. For $y \leq \hat{\theta}$, the inequality $\theta^- \leq y$ is equivalent to $\mathrm{pl}(\theta^-) \leq \mathrm{pl}(y)$. By the construction of the value θ^-, we know that $\mathrm{pl}(\theta^-) = \omega$. Thus, the condition $\{\theta : \mathrm{pl}_x(\theta) \geq \omega\} \cap (-\infty, y] \neq \emptyset$ is simply equivalent to $\omega \leq \mathrm{pl}_x(y)$. Hence,

$$\overline{F}(y) = \mathrm{Prob}(\omega \leq \mathrm{pl}_x(y)).$$

When ω is uniformly distributed on the interval $[0,1]$, then, for all z, the probability $\mathrm{Prob}(\omega \leq z)$ is simply equal to z. In particular, for $z = \mathrm{pl}_x(y)$, we have $\overline{F}(y) = \mathrm{pl}_x(y)$. In these terms, in the plausibility-based forecasting method, as the lower bound $\underline{\theta}_\alpha$ of the confidence interval, we select the value $\underline{\theta}_\alpha$ for which

$$\mathrm{pl}_x(\underline{\theta}_\alpha) = \frac{\alpha}{2}.$$

Thus, the confidence interval obtained by using the plausibility method is the interval $[\underline{\theta}_\alpha, \overline{\theta}_\alpha]$ between the two values $\underline{\theta}_\alpha < \hat{\theta} < \overline{\theta}_\alpha$ for which

$$\mathrm{pl}_x(\underline{\theta}_\alpha) = \mathrm{pl}_x(\underline{\theta}_\alpha) = \frac{\alpha}{2}.$$

The confidence interval $[\underline{\theta}_\alpha, \overline{\theta}_\alpha]$ consists of all the values θ for which

$$\mathrm{pl}_x(\theta) \geq \frac{\alpha}{2}.$$

In terms of the likelihood function $L_x(\theta)$, this means that, as the confident interval, we select the set of all the values θ for which

$$\frac{L_x(\theta)}{L_x\left(\hat{\theta}\right)} \geq \frac{\alpha}{2},$$

i.e., equivalently, for which

$$\ln\left(L_x(\theta)\right) \geq \ln\left(L_x\left(\hat{\theta}\right)\right) - \left(\ln(2) + |\ln(\alpha)|\right). \qquad (1)$$

Let us compare the resulting confident interval with the traditional likelihood-based confidence interval. In traditional statistics, one of the methods to estimate the confidence interval based on the likelihood function – based on Wilks's theorem – is to select the set of all possible values θ for which

$$\ln\left(L_x(\theta)\right) \geq \ln\left(L_x\left(\hat{\theta}\right)\right) - \frac{1}{2} \cdot \chi^2_{1,1-\alpha}; \qquad (2)$$

see, e.g., [2], where $\chi^2_{1,1-\alpha}$ is the threshold for which, for the χ^2_1-distribution – i.e., for the square of the standard normally distributed random variable, with 0 means and standard deviation 1, we have $\text{Prob}(\chi^2 \leq \chi^2_{1,1-\alpha}) = 1 - \alpha$.

The corresponding confidence interval (2) is somewhat different from the interval (1) obtained by using plausibility-based forecasting. It is known that Wilks's theorem provides an *asymptotically accurate* description of the confidence region when the number of observations n increases.

It is desirable to modify the plausibility-based forecasting method to make it asymptotically optimal. It is desirable to modify the plausibility-based forecasting method to make its results asymptotically optimal.

3 How to Best Modify the Current Plausibility-Based Forecasting Method: Analysis of the Problem

Problem: reminder. In the previous section, we have shown that the use of a (heuristically selected) unform distribution for the variable ω, while empirically efficient, does not always lead us to asymptotically optimal estimates. Let us therefore try to find an alternative distribution for ω for which, in the above case, the resulting confidence interval will be asymptotically optimal.

Which distribution for ω we should select: analysis of the problem. In the general case, we still have $\overline{F}(y) = \text{Prob}(\omega \leq p_x(y))$. We want to make sure that for the Wilks's bound, this probability is equal to $\frac{\alpha}{2}$.

For the Wilks's bound, by exponentiating both sides of the formula (2), we conclude that

$$p_x(y) = \frac{L_x(\theta)}{L_x\left(\hat{\theta}\right)} = \exp\left(-\frac{1}{2} \cdot \chi^2_{1,1-\alpha}\right);$$

thus, we conclude that

$$\text{Prob}\left(\omega \le \exp\left(-\frac{1}{2}\cdot \chi^2_{1,1-\alpha}\right)\right) = \frac{\alpha}{2}. \tag{3}$$

By definition of $\chi^2_{1,1-\alpha}$, if we take a variable n which is normally distributed with 0 mean and standard deviation 1, then we have:

$$\text{Prob}\left(n^2 \le \chi^2_{1,1-\alpha}\right) = 1-\alpha.$$

Thus, for the opposite event, we have

$$\text{Prob}\left(n^2 \ge \chi^2_{1,1-\alpha}\right) = (1-(1-\alpha)) = \alpha.$$

The inequality $n^2 \ge \chi^2_{1,1-\alpha}$ occurs in two equally probable situations:

- when n is positive and $n \ge \sqrt{\chi^2_{1,1-\alpha}}$ and
- when n is negative and $n \le -\sqrt{\chi^2_{1,1-\alpha}}$.

Thus, the probability of each of these two situations is equal to $\frac{\alpha}{2}$; in particular, we have:

$$\text{Prob}\left(n \le -\sqrt{\chi^2_{1,1-\alpha}}\right) = \frac{\alpha}{2}. \tag{4}$$

Let us transform the desired inequality (3) to this form. The inequality

$$\omega \le \exp\left(-\frac{1}{2}\cdot \chi^2_{1,1-\alpha}\right)$$

is equivalent to

$$\ln(\omega) \le -\frac{1}{2}\cdot \chi^2_{1,1-\alpha},$$

hence to

$$-2\ln(\omega) \ge \chi^2_{1,1-\alpha},$$

$$\sqrt{-2\ln(\omega)} \ge \sqrt{\chi^2_{1,1-\alpha}},$$

and

$$-\sqrt{-2\ln(\omega)} \le -\sqrt{\chi^2_{1,1-\alpha}}.$$

Thus, the desired inequality (3) is equivalent to

$$\text{Prob}\left(-\sqrt{-2\ln(\omega)} \le -\sqrt{\chi^2_{1,1-\alpha}}\right) = \frac{\alpha}{2}.$$

In view of the formula (4), this equality is attained if we have $n = -\sqrt{-2\ln(\omega)}$. In this case, $-2\ln(\omega) = n^2$, hence

$$\ln(\omega) = -\frac{n^2}{2},$$

and thus,

$$\omega = \exp\left(-\frac{n^2}{2}\right). \quad (5)$$

So, we arrive at the following conclusion.

Conclusion. Instead of a uniformly distributed random variable ω, we need to use a variable (5), where n is a random variable distributed according to the standard normal distribution – with 0 means and standard deviation 1.

What is the probability density function of this distribution? In general, if we have a random variable with a probability density function $\rho_X(x)$, then for any function $f(x)$, for the random variable $Y = f(X)$, we can determine its probability density function $\rho_Y(y)$ from the condition that for $y = f(x)$, we have

$$\text{Prob}(f(x) \le Y \le f(x+dx)) = \text{Prob}(x \le X \le x+dx) = \rho_X(x) \cdot dx.$$

Here, $f(x) = y$ band

$$f(x+dx) = f(x) + f'(x) \cdot dx = y + f'(x) \cdot dx,$$

hence

$$\text{Prob}(f(x) \le Y \le f(x+dx)) = \text{Prob}(y \le Y \le y + f'(x) \cdot dx) = \rho_Y(y) \cdot |f'(x)| \cdot dx.$$

Equating these two expressions, we conclude that for $y = f(x)$, we have

$$\rho_Y(y) = \frac{\rho_X(x)}{|f'(x)|}.$$

In our case,

$$\rho_X(x) = \frac{1}{\sqrt{2\pi}} \cdot \exp\left(-\frac{x^2}{2}\right)$$

and

$$f(x) = \exp\left(-\frac{x^2}{2}\right),$$

hence

$$f'(x) = -\exp\left(-\frac{x^2}{2}\right) \cdot x.$$

Thus,

$$\rho_Y(y) = \frac{1}{\sqrt{2\pi} \cdot |x|}.$$

From $y=\exp\left(-\frac{x^2}{2}\right)$, we conclude that $\frac{x^2}{2}=-\ln(y)$, thus, $|x|=\sqrt{2\cdot|\ln(y)|}$. So, the probability distribution function for $y=\omega$ has the form

$$\rho(\omega)=\frac{1}{\sqrt{2\pi}\cdot\sqrt{2\cdot|\ln(\omega)|}}=\frac{1}{2\cdot\sqrt{\pi}\cdot\sqrt{|\ln(\omega)|}}.$$

This distribution is indeed close to uniform. The value $\ln(\omega)$ is changing very slowly, so, in effect, the resulting probability density function is close to a constant, and thus, the corresponding probability distribution is close to the uniform one.

4 Resulting Recommendations

As a result of the above analysis, we arrive at the following modification of the plausibility-based forecasting algorithm.

In this modification, first, we define the likelihood function $L_x(\theta)$ and then find its largest possible value $L_x\left(\hat{\theta}\right)=\max\limits_{\theta} L_x(\theta)$.

Then, we define the plausibility function as

$$\text{pl}_x(\theta)=\frac{L_x(\theta)}{\sup\limits_{\theta'} L_x(\theta')}=\frac{L_x(\theta)}{L_x\left(\hat{\theta}\right)}.$$

Based on this plausibility function, we define, for each real number $\omega\in[0,1]$, a plausibility region

$$\Gamma_x(\omega)=\{\theta:\text{pl}_x(\theta)\ge\omega\}.$$

We then represent a probability distribution for y as $y=g(\theta,z)$ for an auxiliary variable z which is uniformly distributed on the interval $[0,1]$.

Based on the plausibility regions, we then compute the belief and plausibility of each set A of possible values of θ as follows:

$$\text{Bel}(A)=\text{Prob}(g(\Gamma_x(\omega),z)\subseteq A)$$

and

$$\text{Pl}(A)=\text{Prob}(g(\Gamma_x(\omega),z)\cap A\neq\emptyset),$$

where both z is uniformly distributed on the interval $[0,1]$, and ω is distributed in accordance with the probability density

$$\rho(\omega)=\frac{1}{2\cdot\sqrt{\pi}\cdot\sqrt{|\ln(\omega)|}}.$$

The corresponding random variable can be simulated as

$$\omega = \exp\left(-\frac{n^2}{2}\right),$$

where n is a standard normally distributed random variable, with 0 mean and standard deviation 1.

After that, we compute the lower and upper bounds on the cdf $F(y)$ for y as

$$\underline{F}(y) = \text{Bel}((-\infty, y])$$

and

$$\overline{F}(y) = \text{Pl}((-\infty, y]).$$

Then, for any given small value $\alpha > 0$, we predict that y is, with confidence $1-\alpha > 0$, contained in the interval $[\underline{y}_\alpha, \overline{y}_\alpha]$, where $\underline{y}_\alpha = (\overline{F})^{-1}\left(\frac{\alpha}{2}\right)$ and $\overline{y}_\alpha = (\underline{F})^{-1}\left(1-\frac{\alpha}{2}\right)$.

Acknowledgments

We acknowledge the partial support of the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand. This work was also supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721, and by an award "UTEP and Prudential Actuarial Science Academy and Pipeline Initiative" from Prudential Foundation.

The authors are greatly thankful to Hung T. Nguyen for valuable suggestions.

References

1. N. B. Abdallah, N. M. Voyeneau, and T. Denoeux, "Combining Statistical and Expert Evidence within the D-S Framework: Application to Hydrological Return Level Estimation", In: T. Denoeux and M.-H. Masson (eds.), *Belief Functions: Theory and Applications: Proceedings of the 2nd International Conference on Belief Functions, Compiègne, France, May 9–11, 2012*, Springer Verlag, Berlin, Heidelberg, New York, 2012, pp. 393–400.
2. F. Abramovich and Y. Ritov, *Statistical Theory: A Concise Introduction*, CRC Press, Boca Raton, Florida, 2013.
3. O. Kanjanatarakul, S. Sriboonchitta, and T. Denoeux, "Forecasting using belief functions: an application to marketing econometrics", *International Journal of Approximate Reasoning*, 2014, Vol. 55, pp. 1113–1128.
4. N. Thianpaen, J, Liu, and S. Sriboonchitta, "Time series using AR-belief approach", *Thai Journal of Mathematics*, to appear.