

Unknown cumulative distribution function research proposal

[Law](#), [Evidence](#)



Introduction

The various setbacks to developing and authenticating input distributions are well known in any institutions that deal with risk assessment and analysis.

Due to all these, it has been given a considerable attention more recently.

There are various strategies of estimating probability distribution from either empirical data or standard approaches, which are limit practical

effectiveness especially whenever such data available are limited. Whenever such cases presents themselves, certain mechanisms and strategies are employed.

Such include maximum entropy criteria which actually

synthesizes distribution from priori constraints. The other mechanism is to

employ the use of ' default' such include the triangle or exponential

distributions, which are applicable whenever very little empirical information is available (Diebold, Doherty & Herring, 2012).

There has always been an argument that the use of ' default' input

distributions should at whatever circumstance to be avoided since they are purely based on thoughtful or wishful thinking which results to scientific

deduction or retardation. As suggested by Wittgenstein's very last

proposition, we should at the very least be silent about things we cannot say anything about. Empirical methods are always preferred to the other

previously mentioned alternative but subjected to various assumptions. In

practice, practioners of such (analysts) often have limited empirical evidence or justification to anchor in some of the distributions they use as inputs or

variables for probabilistic assessments. This results in the analyses requiring assumptions as mentioned earlier that cannot be justified by appeal to

evidence. The repacaution to this is substantial due to the fact that the

results of probabilistic analyses are sometimes known to be very sensitive to the distributions of choice. The final result of any given analysis will only be as good as the variables it is based on (Ibrahim, Chen & Sinha, 2001).

Statement of the research problem

The research problem is on how to compute an unknown cumulative distribution, provided with certain parameters. This of course has to be based on certain assumptions and hypothesis that need to be investigated. This topic is very vital in companies and organization that deals in risk assessment and analysis, so as not to assume that certain parameters follow a certain distribution when in the real sense it does not.

Context of the research Problem

Whenever several inputs or variables and assumptions are at question, a combinatorial explosion effectively restricts any comprehensive analysis in any sensitive study. For an example, whenever an analysts admits total ignorance about the variates of any particular distribution except that they must be larger than minimum (min) and smaller than the maximum (max). It would be dictated that in such a scenario, a uniform distribution be used instead. The convention is dated back to the commencement of probability theory. It was argued by Laplace that the use of an alternative distribution would result in additional information concerning the relative likelihood of the possible inputs which by hypothesis is lacking. The idea itself is nowadays referred to as the 'insufficient principle of reasoning' and maximum entropy criterion is its generalization (Diebold, Doherty & Herring, 2012). The mentioned criterion allow further constraint information to be

incorporated into the argument to select an appropriate distribution that should express only the constraints at the same time neglecting ceteris paribus.

The approach should therefore provide the way to correctly, characterize the uncertainty in input distributions which are usually used in risk analyses, as has argued scientifically as well as ethically (Kolassa, 2006).

Audience

The target groups of such a proposal are the students that are studying courses in line with this topic, managers in either big or small institutions, governments and even free-lancers that carry out risk assessment and analysis on behalf of other companies or institutions. Others are law enforcers by simply understanding the trend of crime and formulating a distribution that would assist them reduce such; instead of acting on just mere suspicion and information from other reliable sources. A total of 100 participants are to be included in the proposal writing so as to get conclusive methods and ways of finding unknown cumulative distributions. The 100 groups were grouped into small groups of 5 so as to create numerous sample (Lafaye, Drouilhet & Liquet, 2013).

Purpose of the study

The purpose is to help manage risks and make good use of distribution instead of making ambiguous assumptions just to suit one's interests or needs. This would also curb the issue of food insecurity, corruption, water supply by just simply trying to understand what actually these parameters or

variable given mean so as to formulate their correct distribution(Lafaye, Drouilhet & Lique, 2013).

Research questions

- Can one actually compute unknown distribution given certain parameters?
- Is finding the unknown distribution vital in any company, country or an organization?
- How is one going to go about this?
- How is one going.

Statement of method and design section

Non-parametric models

In such scenarios, cases or circumstance where distribution shapes cannot be determined reliably and where the empirical information is limited, one can also derive bounded probability regions. We already went through the case in which the {min, max} are the only available information which generalizes the uniform distribution suggested by a maximum entropy criterion (Diebold, Doherty & Herring, 2012).

Cases where only {min, max, mean} are known, mean of any given value can always be used to constrained to a much smaller region (Diebold, Doherty & Herring, 2012). The derivation of these bounds is quite simple to comprehend. Let us for instance consider the range of x-values between minimum (min) and mean (μ). The value p of a CDF at x surely represents probability mass to the left of x . However much mass there is, it must be balanced by mass on the right of the mean. The greatest mass would definitely be balanced by entirely assuming that the rest of the probability (1

- p) is generally based at maximum. The same applies to the mass arrangement on the other side (left) requires the smallest balance whenever it is all concentrated at point x. Thus, aforementioned considerations lead to the following expression;

$$px + (1 - p)\max = \text{mean}$$

The above equation is manipulated or calculated to yield the following expression

$p = (\max - \text{mean}) / (\max - x)$, which specifies the greatest value of the CDF for the value x.

Assumed distributions

Most certainly, an analyst may be confident about the parameters and shapes of some input distributions on account of good empirical evidence. This confidence can be represented by using osculating bounds describing the distribution function (Grous, 2013).

$$d(p) = F^{-1}(p) = U(p)$$

. Implementation of several distributions are now possible through soft wares, including Most of the commonly known distributions such as the log normal, logistics, Laplace, exponential, Cauchy, discrete and many other available distributions; Pareto, power function, Rayleigh, triangular, uniform, and Wei-bull have as well been implemented in software. The software permits distributions with theoretically infinite supports, that automatically truncates the $100\alpha\%$ percentile and $100(1 - \alpha)\%$ percentile; where α is selected by the analyzing person (analyst) (Diebold, Doherty & Herring, 2012).

Parametric models

It is simple to compute probability bounds from any such cases where the distribution family is known (given), but only interval estimates can be given for the parameters. Let us base this from previous knowledge; we are always willing to assume the fact that a given distribution is log normal, but we can never be completely convinced that the exact parameter values that would define this mentioned distribution (Diebold, Doherty & Herring, 2012). If for a matter of fact that there are bounds on mean (μ) and σ (standard deviation), computing the interval of all possible log normal distributions that have parameters in specified intervals would result in bounds on the distribution.

The bounds are

$$d(p) = \max_{\alpha} L_{\alpha^{-1}}(p)$$

$$u(p) = \min_{\alpha} L_{\alpha^{-1}}(p)$$

Where

$$\alpha \in \{ (\mu, \sigma) \mid \mu \in [\mu_1, \mu_2], \sigma \in [\sigma_1, \sigma_2] \}$$

and L is the CDF of a log normal distribution.

Justification

Basically, such calculations principally might prove to be a difficult task since 3α (α) indexes a set of distribution which is infinite (Grous, 2013).

However, in real life perspective, obtaining the bounds either lower or upper requires computing the envelope over only four (4) distributions: which corresponds to the following parameter sets as indicated below,

(μ_1, σ_1) , (μ_2, σ_1) , (μ_1, σ_2) , and (μ_2, σ_2) . It's straight forwardness and clarity

is as a result of how the distribution families happens to be parameterized by

mean (μ) and standard deviation (σ). Nevertheless, distribution families such as the normal, uniform, exponential, Cauchy, and many others are just as easy to find probability bounds as explained above (Millar, 2011). The approach can also be applied empirical information is given. Grosf suggested that standard confidence interval procedures can be used to deduce probability bounds (Diebold, Doherty & Herring, 2012).

Qualifying remarks

For instance, instead of selecting the log normal distribution of which their parameters are known to be the best estimates from an empirical study which is limited, one can alternatively incorporate some of the sampling uncertainty from the study by using bounds calculated from the confidence intervals around the given parameters. For instance, suppose that strong convincing evidence implies that a distribution is log normal inform, with its μ and σ known only within interval ranges (Grous, 2013).

Apparatus/ Materials

We are going to use a simple random sample to evaluate all the created sample from the number of people that were interviewed.

Procedure and Results

Computing with probability bounds

One would require to be extra attentive and keen to spot the difference between an algorithm for subtraction and that of pluses except that the pluses between the d's and the u are replaced by minuses. The same applies to Multiplication and division, only that they use their own respective

operators, provided that both variables are strictly positive (Chihara & Hesterberg, 2011). A more comprehensive and self explanatory algorithm is normally a basis in the general case, even though division is not defined whenever the divisor support includes zero. Besides from the basic arithmetical operations, it is known to possibly handle other functions as much, integral powers, logarithms, maximum, minimum and exponentiation. There are some people who as well describe such methods (numerical) for calculating bounds without necessarily using independence assumption that exists between variables (Millar, 2011).

Bounds on the sum of A and B, for example, are where (i) varies between 0 and m. The following bounds are assured to enclose the correct and true answer whatsoever the correlation or statistical dependency that may exists between either A or B. More similar expressions may be used for multiplication, division, subtraction and other mathematical operations. These methods generally constitute a more comprehensive dependency analysis of bounds. The algorithms are very general and can therefore be used for practically any given distribution that has a countable or finite Support (which may mean that, as for the Monte Carlo method, uncountable distributions must always be truncated to countable or finite range). The resultant bounds may as well be used in calculations which are subsequent. For example, the bounds of A+B for instance may be brought together with C where C is a definite CDF (Chihara & Hesterberg, 2011). This numerical method can certainly therefore be used in analysis of risk so as to compute bounds.

Limitations of the study

- Technical limitations: Algorithms given do not most of the time apply for multiplication and division whenever distributions have ranges that include non-positive ($x < 0$) numbers.
- Algorithms cannot make use of information about non-zero correlations among the variables despite the fact that such algorithms can either assume independence or make no assumption about dependency.
- Optically narrow solutions are no longer guaranteed whenever repeated variables appear in the risk expressions.

The above limitations however require further methodological research even though they may be surmountable (Grous, 2013).

- Uncertainty treatment of Non-graded.

Treatment with a different uncertainty interval is achieved by the uncertainty about the probability distributions of available inputs. This suggests that one whatsoever makes no statement about precisely where within the bounds the actual distribution lies, not even in which areas it is more likely to be.

The approach is not able to handle 2nd order probability distributions (Chihara & Hesterberg, 2011).

Conclusion

The approach suggested in this paper is quite flexible such that it allows analysts to make risks which are probabilistic. A probabilistic risk refers to risks that are actually calculate from the concept of probability theory. This approach will definitely be useful whenever empirical information is available

because it does not restrict sampling that defines them to be used during mathematical computations. The level of specificity is not normally affected by how the variables are combined in subsequent arithmetic operations, and the result thereby represents the appropriate level uncertainty rather than a false precision gained by unjustified assumptions. This approach therefore results positively in a way to honestly as well as fully characterize the uncertainty involved in the input distributions.

References

Diebold, F. X., Doherty, N. A., & Herring, R. (2012). *The known, the unknown, and the unknowable in financial risk management: Measurement and theory advancing practice*. Princeton, N. J: Princeton University Press.

Millar, R. B. (2011). *Maximum likelihood estimation and inference: With examples in R, SAS and ADMB*.

Meucci, A. (2005). *Risk and asset allocation*. Berlin [etc.]: Springer.

Chihara, L., & Hesterberg, T. (2011). *Mathematical statistics with resampling and R*. Hoboken, N. J: J. Wiley & Sons.

Grous, A. (2013). *Fracture mechanics: 1*.

Lafaye, M. P., Drouilhet, R., & Liqueur, B. (2013). *The R software: Fundamentals of programming and statistical analysis*.

Kolassa, J. E. (2006). *Series approximation methods in statistics*. (Springer e-books.) New York: Springer.

Ibrahim, J. G., Chen, M.-H., & Sinha, D. (2001). *Bayesian survival analysis*. New York, NY [u. a.]: Springer.

Millar, R. B. (2011). *Maximum likelihood estimation and inference: With*

examples in R, SAS and ADMB.

Cullen, A. C., & Frey, H. C. (1999). Probabilistic techniques in exposure assessment: A handbook for dealing with variability and uncertainty in models and inputs. New York, NY [u. a.: Plenum Press.

De, B. P. (2004). Explanatory item response models: A generalized linear and nonlinear approach. New York, NY [u. a.: Springer.