Confidence distribution: a sample-dependent distribution function for inference

Min-ge Xie
Rutegrs University, Piscataway, New Jersey, USA Corresponding author: Min-ge Xie, email: mxie@stat.rutgers.edu

Abstract

Distributional inference aims to define a sample-dependent distribution on the parameter space that can provide meaningful answers for all sorts of questions related to statistical inference. In recent years, there have been several breakthroughs coming from different communities studying in their turn objective Bayesian inference, confidence distributions, generalized fiducial inference and belief functions, leading to renewed interest in the field. This talk will review some recent developments of confidence distributions, along with a modern definition and interpretation of the concept. A confidence distribution uses a distribution function to estimate a parameter of interest. Some researchers have suggested that a confidence distribution is a frequentist analogue of a Bayesian posterior, although the notion of confidence distribution, especially in its asymptotic form, is much broader. This talk will provide a unified framework for statistical inference under a frequentist criterion. It will also provide real data examples that highlight the potential and added values of confidence distributions as an effective tool for modern statistical inference.

Keywords: Estimation theory; frequestist/fiducial/Bayesian methods; statistical inference.

1 Introduction

In a frequentist inference, we often use a single point (point estimator) and an interval ("interval estimator"/confidence interval) to estimate a parameter of interest. A simple questions is:

> *Can we also use a distribution function ("distribution estimator") to estimate a parameter of interest in a frequentist inference?*

The answer is affirmative, and *confidence distribution* (CD) is such a "distribution estimator" that can be fully defined and interpreted in the frequentist (repetition) framework. The CD concept has a long history, especially with its interpretation in and connection to fiducial inference (see, e.g., Fisher, 1973, Efron, 1993). Historically, it has been long misconstrued as a fiducial concept, and has not been fully developed under the frequentist framework — maybe partly due to Fisher's "stubborn insistence" and his "unproductive dispute" with Neyman (Zabell, 1992). In recent years, the CD concept has attracted a surge of renewed attention, where the new developments have re-defined the CD as a *purely frequentist concept, without any fiducial reasoning*. The goal is on providing useful statistical inference tools for problems where frequentist methods with good properties were previously unavailable. The nice thing of treating CD as a purely frequentist concept is that the CD is now a clean and coherent frequentist concept (similar to a point estimator) and it frees from those restrictive, if not controversial, constraints set forth by Fisher on fiducial distributions.

The CD concept has often been loosely referred to as a sample-dependent distribution function that can represent confidence intervals of all levels for a parameter of interest (see, e.g., Efron, 1993). As a distribution function, it contains a wealth of information for inferences; much more than a point estimator or a confidence interval. It is more than a"frequentist analogue of a Bayesian posterior". Recent research has shown that the new CD concept encompasses and unifies a wide range of examples. A main theme that was articulated in Xie and Singh (2013) is —"any approach, regardless of being frequentist, fiducial or Bayesian, can potentially be unified under the concept of confidence distributions, as long as it can be used to build confidence intervals of all levels, exactly or asymptotically." Efron (2013) considered the development of confidence distribution as "a grounding process" to solve "perhaps the most important unresolved problem in statistical inference" on "the use of Bayes theorem in the absence of prior information." Cox (2013) stated that a confidence distribution "provides simple and interpretable summaries of what can reasonably be learned from data (and an assumed model)." It is our hope that recent emerging developments on confidence distributions, along with recent surge of publications on generalized fiducial distributions and objective Bayes, will stimulate further explorations that can enrich statistical sciences.

The CD concept is relatively unfamiliar to a general audience of statistics. We provide a brief review of the CD concept, as long with some comparison with fiducial and Bayesian inference, in this conference article. Most materials are from Xie and Singh (2013). More elaborated discussions can be found in Xie and Singh (2013) and the references therein.

2 CD concept, history, new definition and highlights

The concept of "confidence" was first introduced by Neyman (1934, 1937) in his seminal papers on confidence intervals, where frequentist repetition properties for confidence were clarified. According to Fraser (2011), the seed idea of a confidence distribution can be even traced back before Neyman (1934, 1937) to Bayes (1763) and Fisher (1922). The earliest use of the terminology confidence distribution in a formal publication is perhaps Cox (1958). But the terminology appears to have been used before, as early as in 1927; c.f., David and Edwards (2000), p. 191, for an excerpt of a letter from E.S. Pearson to W.S. Gossett. Schwede and Hjort (2002) and Hampel (2006) suggested that confidence distributions are "the Neymanian interpretation of Fishers fiducial distributions", although Fisher furiously disputed this interpretation.

From a new viewpoint of estimation, a CD function uses a sample-dependent distribution function, instead of a single point (point estimator) or an interval (confidence interval), to estimate the parameter of interest. A CD function is a function of both the parameter and the random sample. It is a sample-dependent distribution function on the parameter space of the parameter of interest, and it satisfies certain requirement so that it can provide appropriate inferences for the parameter. The following CD definition is formulated in Schweder and Hjort (2002) and Singh, Xie and Strawderman (2005). In the definition, Θ is the parameter space of the unknown parameter of interest θ and $\mathcal{X}$ is the sample space corresponding to data $\mathbf{X}_n = \{X_1, \ldots, X_n\}$. Singh, Xie and Strawderman (2005) demonstrated that this version of CD definition is consistent with the classical CD definition which is compiled from confidence intervals of varying confidence levels (c.f., Efron, 1993), and it is easier to use in many practices.

Definition 1. *A function $H_n(\cdot) = H_n(\mathbf{X}_n, \cdot)$ on $\mathcal{X} \times \Theta \to [0, 1]$ is called a confidence distribution (CD) for a parameter θ, if it follows two requirements: R1) For each given $\mathbf{X}_n \in \mathcal{X}$, $H_n(\cdot)$ is a continuous cumulative distribution function on Θ; R2) At the true parameter value $\theta = \theta_0$, $H_n(\theta_0) \equiv H_n(\mathbf{X}_n, \theta_0)$, as a function of the sample $\mathbf{X}_n$, follows the uniform distribution $U[0, 1]$. Also, the function $H(\cdot)$ is an asymptotic CD (aCD), if the $U[0, 1]$ requirement is true only asymptotically and the continuity requirement on $H_n(\cdot)$ is dropped.*

Based on the definition, any sample-dependent distribution function on the parameter space can potentially be used to estimate the parameter, but the $U[0, 1]$ requirement in *R2* is imposed to ensure that the statistical inferences (e.g., point estimates, confidence intervals, p-values, etc) derived from the CD have desired frequentist property. This practice of two requirements has an analogue in point estimation: any single point (a real value or a statistic) on the parameter space can potentially be used to estimate a parameter, but we often impose restrictions so that the point estimator can have certain desired properties, such as unbiasedness, consistency etc.

A simple example of a CD, that has been broadly used in statistical practice, is a bootstrap distribution. Efron (1998) explicitly stated that a bootstrap distribution is typically a "distribution estimator" and a "confidence distribution" (CD) function of the parameter that it targets. Singh, Xie and Strawderman (2005, 2007) showed that a bootstrap distribution typically satisfies the CD (asymptotic CD) definition. In fact, in any situations where one can construct a bootstrap distribution, one can construct a CD (or asymptotic CD) function as well.

Another simple example of a CD, which is also used by Fisher (1922, 1973) to illustrate his fiducial function, is from the normal mean inference problem with sample $X_i \sim N(\mu, \sigma^2)$, $i = 1, \ldots, n$. The basic CD's for μ are $\Phi(\sqrt{n}(\mu - \bar{X})/\sigma)$ when σ is known and $T_{n-1}(\sqrt{n}(\mu - \bar{x})/s)$ when σ is not known, and furthermore $\Phi(\sqrt{n}(\mu - \bar{X})/s)$ is an asymptotic CD when $n \to \infty$. Here, $\bar{X}$ and s^2 are the sample mean and variance, respectively, and T_{n-1} stands for the cumulative distribution function of the t-distribution with $n-1$ degrees of freedom. In other words, $N(\bar{X}, \sigma^2)$ is a "distribution estimator" (CD) of μ, when σ^2 is known. The distribution functions $T_{n-1}(\sqrt{n}(\mu - \bar{x})/s)$ or $N(\bar{X}, s^2)$ can be used to estimate μ, when σ^2 is not known.

A confidence distribution is a probability distribution function on the parameter space. It contains a wealth of information for inference; much more than a point estimator or a confidence interval. Xie and Singh (2013) provided a detailed review on how to obtain various types of inference (point estimate, confidence interval, p-value, etc) under the frequentist framework from a confidence CD. It reveals that the CD-based inference is similar in style to that of a Bayesian posterior.

3 CD versus fiducial inference and belief function

Both the CD approach and Fisher's fiducial approach share the same goal of providing "distribution estimator" for parameters, and are sometimes misconstrued as equivalent. According to Hannig (2009), the Fiducial inference can be "viewed as a procedure that obtains a measure on a parameter space" and "statistical methods designed using the fiducial reasoning have typically very good statistical properties" as measured by frequentist inference. These two considerations are similar to those behind the new CD definition, expressed as the two requirements in Definition 1.

However, the new CD development is not a part of any fiducial development — this is different from the classical CD development in which CD is a notion affiliated with fiducial distributions (Fisher, 1973). The new CD definition is purely a frequentist concept and its development does not involve any fiducial reasoning. One such example is the bootstrap distribution. Bootstrap distributions are CDs but no fiducial reasoning is needed for the development of bootstrap.

In addition, the recent development has also made the CD notion substantially broader than the fiducial distribution concept. The relation of the new CD concept versus the fiducial distribution can be described easily by an analogy using two point estimators: a consistent estimator versus a maximum likelihood estimator (MLE). The CD concept is analogous to a consistent estimator which is defined to ensure a certain desired property for inference; a fiducial concept is analogous to the MLE which provides a standard procedure to find an estimator which often happens to possess the desired frequentist property. A consistent estimator does not have to be an MLE; but, under some regularity conditions, the MLE typically has consistency and the MLE method thus provides a standard procedure to obtain a consistent estimator. In the context of distribution estimation, a CD does not have to be a fiducial distribution or involve any fiducial reasoning; but, under suitable conditions, a fiducial distribution may satisfy the required coverage property (see, e.g., Hannig, 2009), which establishes it as a CD function.

There is also renewed interest in fiducial inference and its extensions, notably the recent developments on "generalized fiducial inference" by S. Weerahandi, J. Hannig, H. Iyer, and their colleagues and the developments of belief functions under Dempster-Shafer theory by A. Dempster, G. Shafer and others. See, e.g., review articles by Hannig (2009), Dempster (2008) and Martin, Zhang and Liu (2010). These developments, together with the new developments of CDs, represent an emerging new research field to address problems where frequentist methods with good properties were previously unavailable. However, the recent development on CDs is also distinct from the other two developments in that it is developed strictly within the frequentist domain and resides entirely within the frequentist logics. There is no involvement of any new theoretical framework such as fiducial reasoning or Dempster-Shafer theory. Unlike belief functions and Dempster's rule of recombination, the CD approaches are easy to compute and can be directly related to a vast collection of practical examples used in the classical and current statistical practices.

4 CD versus Bayesian inference

A CD approach is a frequetntist procedure and it inherits the properties of frequentist procedures. Wasserman (2007) re-examined Efron's 1986 paper titled"Why not everyone is a Bayesian?" He summarized four fundamental differences of the frequestist and Bayesian approaches. Due to space limitation, we only exam here one of the four items "*division of labor*", which has also been highlighted in some recent research on CDs.

Under the item "division of labor", Wasserman (2007) stated that "statistical problems need to be solved as one coherent whole in a Bayesian approach", including assigning priors and conducting analyses with "nuisance parameters", while "statistical problems do not have to be solved as one coherent whole" is a "liberating" aspect of "frequentist inference". Efron (1986) and Wasserman (2007) illustrated this point by using the example of population quantile, which can be directly estimated in a frequentist setting by its corresponding sample quantile without any modeling effort or involving other (nuisance) parameters. One example in the CD developments that highlights the phenomenon of "division of labor" is Xie, Liu, Damaraju and Olson (2013).

The research of Xie et al. (2013) stemmed from a consulting project on a real clinical trial of a migraine therapy in Johnson and Johnson Pharmaceuticals. In the clinical trial, both the treatment and control responses are binary: $X_{1i} \sim \text{Bernoulli}(p_1)$, $i = 1$ to n_1, and $X_{0j} \sim \text{Bernoulli}(p_0)$, $j = 1$ to n_0. The parameter of interest is $\delta = p_1 - p_0$. Prior to the clinical trial and following the design of Parmar et al. (1994) and Spiegelhalter et al. (1994), expert opinions on the improvement δ were solicited and the prior information was aggregated to obtain a prior distribution $\delta \sim \pi(\delta)$. The goal is to incorporate the information from the experts with the clinical trial data, and make inference about δ.

Note that, it is not possible to find a conditional density function $f(data|\delta)$ in this binomial trail, thus we can not directly apply Bayes formula $f(\delta|data) \propto \pi(\delta)f(data|\delta)$ focusing only on δ. (This is exactly Efron and Wasserman's argument that a Bayesian approach is not fit for "division of labor"!). A sound Bayesian solution is a full Bayesian model that can jointly model p_0 and p_1 or their re-parametrizations. Joseph, et al (1997) presented such a full Bayesian approach using independent beta priors for p_0 and p_1. Xie et al. (2013) extended the full Bayesian approach by including three more priors: independent hierarchical beta priors, dependent bivariate beta priors (Olkin and Liu, 2003), and dependent hierarchical bivariate beta priors.

However, in the case of having a skewed histogram of expert opinions (Fig 1(a)), the bivariate prior is also skewed (blue contours in Fig 1(b)). When we project (obtain the marginals of) the

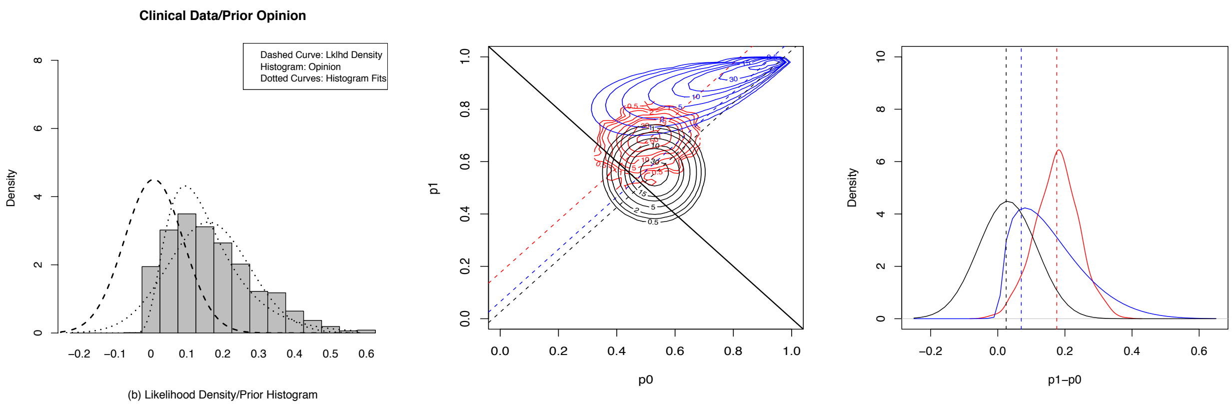

Fig 1. (a) Histogram (prior) of expert opinions on $\delta = p_1 - p_0$; (b) Contours of a bi-beta distributed prior $\pi(p_0, p_1)$ (in blue), density function $f(data|p_0, p_1)$ (in black), and posterior $f(p_0, p_1|data)$ (in red); (c) Projections (marginals) of $\pi(p_0, p_1)$ (in blue), density function $f(data|p_0, p_1)$ (in black), and posterior $f(p_0, p_1|data)$ (in red) onto the direction of $\delta = p_1 - p_0$.

joint prior $\pi(p_0, p_1)$, the joint likelihood function $\ell(p_0, p_1|data)$ and the joint posterior $f(p_0, p_1|data)$ on to the direction of $\delta = p_1 - p_0$, the posterior distribution $f(\delta|data)$ is on the right side of both the marginal prior $\pi(\delta)$ and the projection of the likelihood function (Fig 1(c)). This contradicts our intuition and it amounts to a "paradox". Further examination suggests that this is a genuine mathematical phenomenon and it can happen in many Bayesian analysis involving skewed distributions.

Alternatively, Xie et al. (2013) proposed a univariate frequentist approach to model directly the parameter of interest δ (not jointly (p_0, p_1)). But, it is clear that a standard frequentist approach is not equipped to deal with external information such as expert opinions, since opinions are not actual observed data from the clinical trials. To overcome this difficulty, Xie et al. (2013) took advantage of the CD development. In particular, a CD function was used to summarize the expert opinions and, subsequently, combined it with the estimates from the clinical trial. This CD approach can be viewed as a compromise between the Bayesian and frequentist paradigms. It is a frequentist approach, since the parameter is treated as a fixed value and not a random entity. It nonetheless also has a Bayesian flavor, since the prior expert opinions represent only the relative experience or prior knowledge of the experts but not any actual observed data. The CD approach is easy to implement and computationally cheap (no MCMC algorithm). More importantly, it seems to have avoided some difficult issues alluded to the Bayesian approaches.

Finally, combining evidence from disparate sources of information has long been considered a weak point of frequentist theory, while being the strong suit of Bayesian analysis. Recent publications on CDs seem to affirm that the CD development may be able to fill such a void. On the other hand, the CD development may also assist the development of objective Bayesian approaches. For instance, a CD approach could potentially be a good way to obtain an objective prior where we can use a variety of frequentist tool kits, including robust approaches.

References

Cox, D.R. (1958). Some problems with statistical inference. Ann. Math. Stat., 29, 357-372.

Cox, D. R. (2013), Discussion of Confidence Distribution, the Frequentist Distribution Estimator of a Parameter a Review by Xie and Singh. International Statistical Review, 81, 40-41.

Dempster, A.P. (2008). The Dempster-Shafer calculus for statisticians. Internat. J. Approx. Rea-

son., 48, 365-377.
Efron, B. (1986). Why isnt everyone a Bayesian? Amer. Statist., 40, 262-266.
Efron, B. (1993). Bayes and likelihood calculations from confidence intervals. Biometrika, 80, 3-26.
Efron, B. (1998). R.A.Fisher in the 21st century. Stat. Sci., 13, 95-122.
Efron, B. (2013), Discussion of Confidence Distribution, the Frequentist Distribution Estimator of a Parameter a Review by Xie and Singh. International Statistical Review, 81, 41-42.
Fisher, R.A. (1922). On the mathematical foundations of theoretical statistics. Philos. Trans. R. Soc. Lond. A, 222, 309-368.
Fisher, R.A. (1973). Statistical Methods and Scientific Inference, 3rd ed. New York: Hafner Press.
Fraser, D.A.S. (1991). Statistical inference: Likelihood to significance. J. Amer. Statist. Assoc., 86, 258-265.
Fraser, D.A.S. (2011). Is Bayes posterior just quick and dirty confidence? Stat. Sci., 26, 299-316.
Hannig, J. (2009). On generalized fiducial inference. Statist. Sinica, 19, 491-544.
Joseph, L., du Berger, R. & Belisle, P. (1997). Bayesian and mixed Bayesian/likelihood criteria for sample size determination. Stat. Med., 16, 769-781.
Martin, R., Zhang, J. & Liu, C. (2010). Dempster-Shafer theory and statistical inference with weak beliefs. Statist. Sci., 25, 72-87.
Neyman, J. (1934). On the two different aspects of representative method: The method of stratified sampling and the method of purpose selection. J. Roy. Statist. Soc. Ser. A, 97, 558-625.
Neyman, J. (1937).Outline of a theory of statistical estimation based on the classical theory of probability. Philos.Trans. R. Soc. A., 237, 333-380.
Parmar, M.K.B., Spiegelhalter, D.J, Freedman, L.S. & Chart Steering Committee (1994). The chart trials: Bayesian design and monitoring in practice. Stat. Med., 13, 1297-1312.
Schweder, T. & Hjort, N.L. (2002). Confidence and likelihood. Scand. J. Stat., 29, 309-332.
Schweder, T & Hjort, N.L (2003). Frequentist analogues of priors and posteriors. In Econometrics and the Philosophy of Economics, Ed. B.P. Stigum, pp. 285-317. Princeton, New Jersey: Princeton University Press.
Singh, K. & Xie, M. (2012). CD-posterior combining prior and data through confidence distributions. In Contemporary Developments in Bayesian Analysis and Statistical Decision Theory: A Festschrift for William E. Strawderman. IMS Collection, Eds. D. Fourdrinier, E. Marchand & A.L. Rukhin, Vol. 8, pp. 200-214. Baltimore: Mattson Publishing Services.
Singh, K.,Xie,M. &Strawderman,W.E. (2005).Combining information from independent sources through confidence distributions. Ann. Statist. 33, 159-183.
Singh, K., Xie, M. & Strawderman, W.E. (2007). Confidence distribution (CD)-distribution estimator of a parameter. In Complex Datasets and Inverse Problems. IMS Lecture Notes-Monograph Series, 54, 132-150.
Spiegelhalter, D.J., Freedman, L.S. & Parmar, M.K.B. (1994). Bayesian approaches to randomized trials. J. Roy. Statist. Soc. Ser. A, 157, 357-416.
Wasserman, L. (2007). Why isnt everyone a Bayesian? In The Science of Bradley Efron, Eds. C.N. Morris & R. Tibshirani, pp. 260-261. New York: Springer.
Xie, M., Liu, R.Y., Damaraju, C.V. & Olson, W.H. (2013). Incorporating external information in analyses of clinical trials with binary outcomes. The Annals of Applied Statistics, 7, 342-368.
Xie, M., Singh, K. & Strawderman,W.E. (2011). Confidence distributions and a unifying framework formeta-analysis. J. Amer. Statist. Assoc., 106, 320-333.
Xie, M. & Singh, K. (2013). Confidence Distribution, the Frequentist Distribution Estimator of a Parameter a Review. International Statistical Review (with discussion). 81, 3-39.
Zabell, S.L. (1992). R.A. Fisher and fiducial argument. Stat. Sci., 7, 369-387.