Confidence distributions in statistical inference

Sergei I. Bityukov, Vera V. Smirnova
Institute for High Energy Physics, Protvino, Russia
Nikolai V. Krasnikov
Institute for Nuclear Research RAS, Moscow, Russia
Saralees Nadarajah
University of Manchester, Manchester M13 9PL, United Kingdom

Plan

- **Motivation**
- **A bit of history**
- **Confidence distributions**
- **Examples and inferential information contained in a CD**
- **Inference: a brief summary**
- **CDs and pivots**
- **Applications**
- **Conclusion**
- **References**

Motivation (I) Common sense

If a procedure states one-to-one correspondence between the observed value of a random variable and the confidence interval of any level of significance then we can reconstruct a unique confidence density of the parameter and, correspondingly, a unique confidence distribution.

The confidence distribution is interpreted here in the same way as confidence interval. From the duality between testing and confidence interval estimation, the cumulative confidence distribution function for parameter μ evaluated at μ_0, $F(\mu_0)$, is the $p-$value of testing $H_0 : \mu \geq \mu_0$ against its one-side alternative. It is thus a compact format of representing the information regarding μ contained in the data and given the model. Before the data have been observed, the confidence distribution is a stochastic element with quantile intervals $(F^{-1}(\alpha_1), F^{-1}(1-\alpha_2))$ which covers the unknown parameter with probability $1-\alpha_1-\alpha_2$. After having observed the data, the realized confidence distribution is not a distribution of probabilities in the frequentist sense, but of confidence attached to interval statements concerning μ.

Motivation (II) Construction by R.A. Fisher

A following example shows the construction by R.A. Fisher (B. Efron (1978))

A random variable x **with parameter** μ

$$x \sim \mathcal{N}(\mu, 1). \tag{1}$$

Probability density function (pdf) here is

$$\varphi(x|\mu) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2}}. \tag{2}$$

We can write

$$x = \mu + \epsilon, \tag{3}$$

where $\epsilon \sim \mathcal{N}(0, 1)$ **and** μ **is a constant.**

Let $\hat{x}$ **be a single realization of** x**. For normal distribution it is an unbias estimator of parameter** μ**, i.e.** $\hat{\mu} = \hat{x}$**, therefore**

$$\mu|\hat{x} = \hat{x} - \epsilon. \tag{4}$$

As is known $(-\epsilon) \sim \mathcal{N}(0, 1)$ **due to symmetry of the bell-shaped curve about its central point, i.e.**

$$\mu|\hat{x} \sim \mathcal{N}(\hat{x}, 1). \tag{5}$$

Motivation (II)

Thus we construct the confidence density of the parameter

$$\tilde{\varphi}(\mu|\hat{x}) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(\hat{x}-\mu)^2}{2}} \qquad (6)$$

uniquely for each value of $\hat{x}$.

As pointed in paper (Hampel 2006) "Fisher (Fisher 1930; 1933) gave correct interpretation of this "tempting" result. But starting in 1935 (Fisher 1935), he really believed he had changed the status of μ from that of a fixed unknown constant to that of a random variable on the parameter space with known distribution". The history and generalization of the last approach can be found in paper (Hannig 2006).

The fiducial argument is very attractive notion and sometimes it reopen (see, as an example, (Hassairi 2005) and corresponding critique (Mukhopadhyay 2006)).

In principle, the parameter μ can be a random variable in the case of the random origin of parameter. We will not discuss here this possibility.

Motivation (III) The presence of invariant

The construction above is a direct consequence of the following identity

$$\int_{-\infty}^{\hat{x}-\alpha_1} \varphi(x|\hat{x})dx + \int_{\hat{x}-\alpha_1}^{\hat{x}+\alpha_2} \tilde{\varphi}(\mu|\hat{x})d\mu + \int_{\hat{x}+\alpha_2}^{\infty} \varphi(x|\hat{x})dx = 1, \quad (7)$$

where $\hat{x}$ is the observed value of random variable x, and $\hat{x}-\alpha_1$ and $\hat{x}+\alpha_2$ are confidence interval bounds for location parameter μ.

The presence of the identities of such type (Eq.7) is a property of statistically self-dual distributions (Bityukov 2004; 2005):

normal and normal

$$\varphi(x|\mu,\sigma) = \tilde{\varphi}(\mu|x,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \sigma = const$$

Cauchy and Cauchy

$$\varphi(x|\mu,b) = \tilde{\varphi}(\mu|x,b) = \frac{b}{\pi(b^2+(x-\mu)^2)}, \quad b = const$$

Laplace and Laplace

$$\varphi(x|\mu,b) = \tilde{\varphi}(\mu|x,b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}, \quad b = const$$

and so on.

Motivation (IV) The invariant in the case of asymmetric distributions

In the case of Poisson and Gamma-distributions we can also exchange the random variable and the parameter, preserving the same formula for the probability distribution:

$$f(i|\mu) = \tilde{f}(\mu|i) = \frac{\mu^i e^{-\mu}}{i!}$$

In this case we can use another identity to relate the pdf of random variable and confidence density of the parameter for the unique reconstruction of confidence density (Bityukov 2000; 2002) (any another reconstruction is inconsistent with the identity and, correspondingly, breaks the probability conservation):

$$\sum_{i=\hat{x}+1}^{\infty} \frac{\mu_1^i e^{-\mu_1}}{i!} + \int_{\mu_1}^{\mu_2} \frac{\mu^{\hat{x}} e^{-\mu}}{\hat{x}!} d\mu + \sum_{i=0}^{\hat{x}} \frac{\mu_2^i e^{-\mu_2}}{i!} = 1 \qquad (8)$$

for any real $\mu_1 \geq 0$ and $\mu_2 \geq 0$ and non-negative integer $\hat{x}$, i.e.

$$\sum_{i=\hat{x}+1}^{\infty} f(i|\mu_1) + \int_{\mu_1}^{\mu_2} \tilde{f}(\mu|\hat{x}) d\mu + \sum_{i=0}^{\hat{x}} f(i|\mu_2) = 1, \qquad (9)$$

where $f(i|\mu) = \tilde{f}(\mu|i) = \dfrac{\mu^i e^{-\mu}}{i!}$. Confidence density $\tilde{f}(\mu|i)$ is the pdf of Gamma-distribution $\Gamma_{1,i+1}$ and $\hat{x}$ is the number of observed events.

A bit of history

The basic notion of CDs traces back to the fiducial distribution of Fisher (1930); however, it can be viewed as a pure frequentist concept. Indeed, as pointed out in Schweder (2002) the CD concept is "Neymannian interpretation of Fisher's fiducial distribution" [Neyman (1941)]. Its development has proceeded from Fisher (1930) through various contributions, just to name a few, of Kolmogorov (1941), Pitman (1957), Efron (1993; 1998), Fraser (1991; 1996), Lehmann (1993), Singh (2001; 2007), Schweder (2002; 2003A) and others. Bityukov (2002; 2005) developed the approach for reconstruction of the confidence distribution densities by using the corresponding identities.

Another useful application of CD is for meta-analysis. Meta-analysis is the modern term for combining results from different experiments or trials (see, for example, (Hedges 1985)). The consecutive theory of combining information from independent sources through CD is proposed in paper (Singh 2005). Recently (Bickel 2006), the method for incorporating expert knowledge into frequentist approach by combining generalized confidence distributions is proposed.

Confidence distributions

Suppose $X_1, X_2, \ldots, X_n$ are n independent random draws from a population $\mathcal{F}$ and χ is the sample space corresponding to the data set $\mathcal{X}_n = (X_1, X_2, \ldots, X_n)^T$. Let θ be a parameter of interest associated with $\mathcal{F}$ ($\mathcal{F}$ may contain other nuisance parameters), and let Θ be the parameter space.

Definition 1 (Singh 2005): A function $H_n(\cdot) = H_n(X_n, (\cdot))$ on $\chi \times \Theta \to [0, 1]$ is called a confidence distribution (CD) for a parameter θ if

(i) for each given $\mathcal{X}_n \in \chi$, $H_n(\cdot)$ is a continuous cumulative distribution function;

(ii) at the true parameter value $\theta = \theta_0$, $H_n(\theta_0) = H_n(\mathcal{X}_n, \theta_0)$, as a function of the sample $\mathcal{X}_n$, has the uniform distribution $U(0, 1)$.

The function $H_n(\cdot)$ is called an asymptotic confidence distribution (aCD) if requirement *(ii)* above is replaced by *(ii)'*: at $\theta = \theta_0, H_n(\mathcal{X}_n, \theta_0) \xrightarrow{W} U(0, 1)$ as $n \to +\infty$, and the continuity requirment on $H_n(\cdot)$ is dropped.

Item *(i)* basically requires the function $H_n(\cdot)$ to be a distribution function for each given sample.

Item *(ii)* basically states that the function $H_n(\cdot)$ contains the right amount of information about the true θ_0.

Confidence distributions

We call, when it exists, $h_n(\theta) = H_n^{'}(\theta)$ a confidence density or CD density.

It follows from the definition of CD that if $\theta < \theta_0,\ H_n(\theta) \overset{sto}{\leq} 1 - H_n(\theta)$, and if $\theta > \theta_0,\ 1 - H_n(\theta) \overset{sto}{\leq} H_n(\theta)$. Here $\overset{sto}{\leq}$ is a stochastic comparison between two random variables; i.e. for two random variable Y_1 and Y_2, $Y_1 \overset{sto}{\leq} Y_2$, if $P(Y_1 \leq t) \geq P(Y_2 \leq t)$ for all t. Thus a CD works, in a sense, like a compass needle. It points towards θ_0, when placed at $\theta \neq \theta_0$, by assigning more mass stochastically to that side (left or right) of θ that contains θ_0. When placed at θ_0 itself, $H_n(\theta) = H_n(\theta_0)$ has the uniform $U[0, 1]$ distribution and thus it is noninformative in direction.

Definition 1 **is very convenient for the purpose of verifying if a particular function is a CD or an aCD.**

Examples and inferential information contained in a CD

Example 1 **Normal mean and variance (Singh 2005): Suppose $X_1, X_2, \ldots, X_n$ is a sample from $\mathcal{N}(\mu, \sigma^2)$, with both μ and σ^2 unknown. A CD for μ is $H_n(y) = F_{t_{n-1}}(\frac{y - \bar{X}}{s_n/\sqrt{n}})$, where $\bar{X}$ and s^2 are, respectively, the sample mean and variance, and $F_{t_{n-1}}(\cdot)$ is a cumulative distribution function of the Student t_{n-1}-distribution. A CD for σ^2 is $H_n(y) = 1 - F_{\chi^2_{n-1}}(\frac{(n-1)s_n^2}{y})$ for $y \geq 0$, where $F_{\chi^2_{n-1}}(\cdot)$ is the cumulative function of the χ^2_{n-1}-distribution.**

Example 2 **p-value function (Singh 2005): For any given $\tilde{\theta}$, let $p_n(\tilde{\theta}) = p_n(\mathcal{X}_n, \tilde{\theta})$ be a p-value for a one-sided test $K_0 : \theta \leq \tilde{\theta}$ versus $K_0 : \theta > \tilde{\theta}$. Assume that the p-value is available for all $\tilde{\theta}$. The function $p_n(\cdot)$ is called a p-value function. Typically, at the true value $\theta = \theta_0$, $p_n(\theta_0)$ as a function of $\mathcal{X}_n$ is exactly (or asymptotically) $U(0, 1)$-distributed. Also, $H_n(\cdot) = p_n(\cdot)$ for every fixed sample is almost always a cumulative distribution function. Thus, usually $p_n(\cdot)$ satisfies the requirements for a CD.**

Inference: a brief summary (Singh 2005)

- *Confidence interval.* **From the definition, it is evident that the intervals $(-\infty, H_n^{-1}(1-\alpha)], [H_n^{-1}(\alpha), +\infty)$ and $(H_n^{-1}(\alpha/2), H_n^{-1}(1-\alpha/2))$ provide $100(1-\alpha)$%-level confidence intervals of different kinds for θ, for any $\alpha \in (0,1)$.**

- *Point estimation.* **Natural choices of point estimators of the parameter θ, given $H_n(\theta)$, include the median $M_n = H_n^{-1}(1/2)$, the mean $\bar{\theta} = \int_{-\infty}^{\infty} t dH_n(t)$ and the maximum point of the CD density $\hat{\theta} = arg\ max_{\theta} h_n(\theta),\ h_n(\theta) = H_n^{'}(\theta)$.**

- *Hypothesis testing.* **From a CD, one can obtain p-values for various hypothesis testing problems. The work (Fraser 1991) developed some results on such a topic through p-value functions. The natural line of thinking is to measure the support that $H_n(\cdot)$ lends to a null hypothesis $K_0 : \theta \in C$. There are possible two types of support:**

  1. **Strong-support $p_s(C) = \int_C dH_n(\theta)$.**
  2. **Weak-support $p_w(C) = sup_{\theta \in C} 2min(H_n(\theta), 1-H_n(\theta))$.**

  If K_0 is of the type $(-\infty, \theta_0]$ or $[\theta_0, +\infty)$ or a union of finitely many intervals, the strong-support $p_s(C)$ leads to the classical p-values.

  If K_0 is a singleton, that is, K_0 is $\theta = \theta_0$, then the weak-support $p_w(C)$ leads to the classical p-values.

Remarks

Confidence distributions can be viewed as "distribution estimators" and are convenient for constructing point estimators, confidence intervals, $p-$values and more.

Confidence distributions can be interpreted as objective Bayesian posteriors (Bayesian posteriors based on objective priors) that have the desirable properties of good coverage and invariance to transformations. In this case, the distribution for frequentist (confidence distribution) and Bayesian (posterior distribution) are the same, the uncertainty intervals: confidence interval and credible interval, are the same, and the uncertainty intervals are usually interpreted in the same way: **95%** probability that true value is in the interval. Other methods, such as profile likelihood, are not dependent on priors, but may be more difficult to interpret.

In the one parameter model, in paper (Schweder 2002) is defined the confidence distribution notion that summarizes a family of confidence intervals.

CDs and pivots (Schweder 2003B)

Consider the statistical model for the data X. The model consists of a family of probability distributions for X, indexed by the vector parameter (ψ, χ), where ψ is a scalar parameter of primary interest, and χ is a nuisance parameter (vector).

Definition 2 : *A univariate data-dependent distribution for ψ, with cumulative distribution function $C(\psi; X)$ and with quantile function $C^{-1}(\alpha; X)$ is an exact confidence distribution if $P_{\psi\chi}(\psi \leq C^{-1}(\alpha; X)) = P_{\psi\chi}(C(\psi; X) \leq \alpha) = \alpha$ for all $\alpha \in (0, 1)$ and for all probability distributions in the statistical model.*

By definition, the stochastic interval $(\infty, C^{-1}(\alpha; X))$ covers ψ with probability α, and is a one-sided confidence interval method with coverage probability α. The interval $(C^{-1}(\alpha; X), C^{-1}(\beta; X))$ will for the same reason cover ψ with probability $\beta - \alpha$, and is a confidence interval method with this coverage probability. When data have been observed as $X = x$, the realized numerical interval $(C^{-1}(\alpha; x), C^{-1}(\beta; x))$ will either cover or not cover the unknown true value of ψ. The degree of confidence $\beta - \alpha$ that is attached to the realized interval is inherited from the coverage probability of the stochastic interval.

CDs and pivots

The confidence distribution has the same dual property. *Ex ante* **data, the confidence distribution is a stochastic entity with probabilistic properties.** *Ex post* **data, however, the confidence distribution is a distribution of confidence that can be attached to interval statement.**

The realized confidence (degree of confidence) $C(\psi; x)$ **is a** $p-$**value of the one-sided hypothesis** $H_0 : \psi \leq \psi_0$ **versus** $\psi > \psi_0$ **when data have been observed to be** x**. The** *ex ante* **confidence,** $C(\psi; X)$ **is by definition uniformly distributed. The** $p-$**value is just a transformation of the test statistic to the common scale of the uniform distributions (***ex ante***). The realized** $p-$**value when testing the two-sided hypothesis** $H_0 : \psi = \psi_0$ **versus** $\psi \neq \psi_0$ **is** $2\ min\{C(\psi_0), 1 - C(\psi_0)\}$**.**

CDs and pivots

Confidence distributions are easily found when pivots (Barndorff 1994) can be identified.

A function of the data and the interest parameter, $p(X, \psi)$, is a pivot if the probability distribution of $p(X, \psi)$ is the same for all (ψ, χ), and the function $p(X, \psi)$ is increasing in ψ for almost all x.

If based on a pivot with cumulative distribution function F, the cumulative confidence distribution is $C(X, \psi) = F(p(X, \psi))$.

From the definition, a confidence distribution is exact if and only if $C(X, \psi) \sim U$ is a uniformly distributed pivot.

The self-duality in Eq. 7 is equivalent to the existence of a linear and symmetrically distributed pivot.

Applications: signal and expected background (Bityukov 2000; 2007)

The confidence density is more informative notion than the confidence interval. For example, the Gamma-distribution $\Gamma_{1,\hat{n}+1}$ is the confidence density of the parameter of Poisson distribution in the case of the $\hat{n}$ observed events from the Poisson flow of events (Bityukov 2000; 2007). It means that we can reconstruct any confidence intervals (shortest, central, ...) by the direct calculation of the pdf of a Gamma-distribution. The following example illustrates the advantages of the confidence density construction.

Let us consider the Poisson distribution with two components: the signal component with a parameter μ_s and background component with a parameter μ_b, where μ_b is known. To construct confidence intervals for the parameter μ_s in the case of observed value $\hat{n}$, we must find the distribution $\tilde{f}(\mu_s|\hat{n})$.

First let us consider the simplest case $\hat{n} = \hat{s} + \hat{b} = 1$. Here $\hat{s}$ is the number of signal events and $\hat{b}$ is the number of background events among the observed number $\hat{n}$ of events.

$\hat{b}$ can be equal to **0** and **1**.

Applications: signal and expected background

We know that the $\hat{b}$ is equal to 0 with probability

$$p_0 = P(\hat{b} = 0) = \frac{\mu_b^0}{0!} e^{-\mu_b} = e^{-\mu_b} \quad (10)$$

and the $\hat{b}$ is equal to 1 with probability

$$p_1 = P(\hat{b} = 1) = \frac{\mu_b^1}{1!} e^{-\mu_b} = \mu_b e^{-\mu_b}. \quad (11)$$

Correspondingly,
$P(\hat{b} = 0|\hat{n} = 1) = P(\hat{s} = 1|\hat{n} = 1) = \dfrac{p_0}{p_0 + p_1}$ **and**
$P(\hat{b} = 1|\hat{n} = 1) = P(\hat{s} = 0|\hat{n} = 1) = \dfrac{p_1}{p_0 + p_1}$**.**

It means that the distribution of the confidence density $\tilde{f}(\mu_s|\hat{n} = 1)$ is equal to the weighted sum of distributions

$$P(\hat{s} = 1|\hat{n} = 1)\tilde{f}(\mu_s|\hat{s} = 1) + P(\hat{s} = 0|\hat{n} = 1)\tilde{f}(\mu_s|\hat{s} = 0), \quad (12)$$

where the confidence density $\tilde{f}(\mu_s|\hat{n} = 0)$ is the Gamma distribution $\Gamma_{1,1}$ with the pdf $\tilde{f}(\mu_s|\hat{s} = 0) = e^{-\mu_s}$
and the confidence density $\tilde{f}(\mu_s|\hat{n} = 1)$ is the Gamma distribution $\Gamma_{1,2}$ with the pdf $\tilde{f}(\mu_s|\hat{s} = 1) = \mu_s e^{-\mu_s}$.

Applications: signal and expected background

As a result, we have the confidence density of the parameter μ_s

$$\tilde{f}(\mu_s|\hat{n}=1) = \frac{\mu_s + \mu_b}{1+\mu_b} e^{-\mu_s}. \tag{13}$$

Using this formula for $\tilde{f}(\mu_s|\hat{n}=1)$, we can construct the shortest confidence interval of any confidence level trivially.

In this manner we can construct the confidence density $\tilde{f}(\mu_s|\hat{n})$ for any values of $\hat{n}$ and μ_b. From Eq. 9 we use the confidence densities $\tilde{f}(\mu_s|\hat{s}=i),\ i=0,\hat{n}$. Mixing together the confidence densities with corresponding conditional probability weights (in analogy with Eq. 12) yields the confidence density

$$\tilde{f}(\mu_s|\hat{n}) = \frac{(\mu_s+\mu_b)^{\hat{n}}}{\hat{n}! \sum\limits_{i=0}^{\hat{n}} \frac{\mu_b^i}{i!}} e^{-\mu_s}. \tag{14}$$

We have obtained the known formula (Helene 1988; Zech 1989; D'Agostini 2003). The numerical results of the calculations of shortest confidence intervals by the using of this confidence density coincide with Bayesian confidence intervals constructed by the using the uniform prior.

Aplications: quality of planned experiment (Bityukov 2003)

Let us consider the estimation of quality of planned experiments as another example of the use of confidence density. The approach is based on the analysis of uncertainty, which will take place under the future hypotheses testing about the existence of a new phenomenon in Nature.

We consider the Poisson distribution with parameter μ and we preserve the notation pf the previous application. We test a simple statistical hypothesis
H_0**:** *new physics is present in Nature* **(i.e.** $\mu = \mu_s + \mu_b$ **)**
against a simple alternative hypothesis
H_1**:** *new physics is absent* **(**$\mu = \mu_b$**).**

The value of uncertainty is determined by the values of the probability to reject the hypothesis H_0 when it is true (Type I error α) and the probability to accept the hypothesis H_0 when the hypothesis H_1 is true (Type II error β). This uncertainty characterizes the distinguishability of the hypotheses under the given choice of critical area.

Aplications: quality of planned experiment

Let both values μ_s and μ_b, which are defined in the previous application, be exactly known. In this simplest case the errors of Type I and II, which will take place in testing of hypothesis H_0 versus hypothesis H_1, can be written as follows:

$$\begin{cases} \alpha = \sum\limits_{i=0}^{n_c} f(i|\mu_s + \mu_b), \\ \beta = 1 - \sum\limits_{i=0}^{n_c} f(i|\mu_b) \end{cases}, \tag{15}$$

where f is a Poisson probability function and n_c is a critical value.

Let the values $\hat{\mu}_s = \hat{s}$ and $\hat{\mu}_b = \hat{b}$ be known, for example, from Monte Carlo experiment with integrated luminosity which is exactly the same as the data luminosity later in the planned experiment. It means that we must include the uncertainties in values μ_s and μ_b to the system of the equations above.

Aplications: quality of planned experiment

As is shown in ref. (Bityukov 2002) (see, also, the generalized case in the same reference and in my poster here) we have the system

$$\begin{cases} \alpha = \int_0^\infty \tilde{f}(\mu|\hat{s}+\hat{b}) \sum_{i=0}^{n_c} f(i|\mu) d\mu = \sum_{i=0}^{n_c} \frac{C^i_{\hat{s}+\hat{b}+i}}{2^{\hat{s}+\hat{b}+i+1}}, \\ \beta = 1 - \int_0^\infty \tilde{f}(\mu|\hat{b}) \sum_{i=0}^{n_c} f(i|\mu) d\mu = 1 - \sum_{i=0}^{n_c} \frac{C^i_{\hat{b}+i}}{2^{\hat{b}+i+1}} \end{cases}, \qquad (16)$$

where n_c is a critical value of the hypotheses testing about the observability of signal and C^i_N is $\frac{N!}{i!(N-i)!}$.

Note, here the Poisson distribution is a prior distribution of the expected probabilities and the negative binomial (Pascal) distribution is a posterior distribution of the expected probabilities of the random variable. This is transformation of the estimated confidence densities $\tilde{f}(\mu|\hat{s}+\hat{b})$ and $\tilde{f}(\mu|\hat{b})$ (pdfs of the corresponding $\Gamma-$distributions) to the space of the expected values of the random variable.

Conclusion

The notion of confidence distribution, an enterely frequentist concept, is in essence a Neymanian interpretation of Fisher's fiducial distribution. It contains information related to every kind of frequentist inference. The confidence distribution is a direct generalization of the confidence interval, and is a useful format of presenting statistical inference.

The follow quotation from Efron(1998) on Fisher's contribution of the fiducial distribution seems quite relevant in the context of CDs:

"... but here is a safe prediction for the 21st century: statisticians will be asked to solve bigger and more complicated problems. I believe there is a good chance that objective Bayes methods will be developed for such problem, and that something like fiducial inference will play an important role in this development. Maybe Fisher's biggest blunder will become a big hit in the 21st centure!"

References

O.E. Barndorff-Nielsen, D.R. Cox (1994), Inference and Asymptotics, Chapman & Hall, London, 1994.

D.R. Bickel (2006), Incorporating expert knowledge into frequentist interface by combining generalized confidence distributions, e-Print: math/0602377, 2006.

S.I. Bityukov, N.V. Krasnikov, V.A. Taperechkina (2000), Confidence intervals for Poisson distribution parameter, Preprint IFVE 2000-61, Protvino, 2000; also, e-Print: hep-ex/0108020, 2001.

S.I.Bityukov (2002), Signal Significance in the Presence of Systematic and Statistical Uncertainties, JHEP 09 (2002) 060; e-Print: hep-ph/0207130;

S.I. Bityukov, N.V. Krasnikov (2003), NIM A502 (2003) 795-798.

S.I. Bityukov, V.A. Taperechkina, V.V. Smirnova (2004), Statistically dual distributions and estimation of the parameters, e-Print: math.ST/0411462, 2004.

S.I. Bityukov, N.V. Krasnikov (2005), Statistically dual distributions and conjugate families, in Proceedings of MaxEnt'05, August 2005, San Jose, CA, USA.

S.I. Bityukov, N.V. Krasnikov, V.V. Smirnova, V.A. Taperechkina (2007), The transform between the space of observed values and the space of possible values of the parameter, Proceedings of Science (ACAT) 062 (2007) 1-9.

G. D'Agostini (2003), Bayesian Reasoning in Data Analysis, a Critical Introduction. World Scientific, Hackensack, NJ, 2003.

B. Efron (1978), Controversies in the Foundations of Statistics, The American Mathematical Monthly, 85(4) (1978) 231-246.

B. Efron (1993), Bayes and likelihood calculations from confidence intervals. Biometrika 80 (1993) 3-26. MR1225211.

B. Efron (1998), R.A. Fisher in the 21st Century. Stat.Sci. 13 (1998) 95-122.

R.A. Fisher (1930), Inverse probability. Proc. of the Cambridge Philosophical Society 26 (1930) 528-535.

R.A. Fisher (1933), The concepts of inverse probability and fiducial probability referring to unknown parameters, Proceedings of the Royal Society, London, A139 (1933) 343-348.

R.A. Fisher (1935), The fiducial argument in statistical inference, Annals of Eugenics 6 (1935) 391-398.

D.A.S. Fraser (1991), Statistical inference: Likelihood to significance. J. Amer. Statist. Assoc. 86 (1991) 258-265. MR1137116.

D.A.S. Fraser (1996), Comments on "Pivotal inference and the fiducial argument", by G. A. Barnard. Internat. Statist. Rev. 64 (1996) 231-235.

F. Hampel (2006), The proper fiducial argument, Lecture Notes in Computer Science 4123 (2006) 512-526.

J. Hannig, H. Iyer, P. Patterson (2006), Fiducial generalized confidence intervals, Journal of the American Statistical Association 101 (2006) 254-269.

A. Hassairi, A. Masmoudi, C.C. Kokonendji (2005), Implicit distributions and estimation, Communications in Statistics—Theory and Methods 34 (2005) 245-252.

L.V. Hedges, I. Olkin (1985), Statistical Methods for Meta-analysis, Academic Press, Orlando, 1985.

A.N. Kolmogorov (1942), The estimation of the mean and precision of a finite sample of observations (in Russian), Bulletin of the Academy of Sciences, USSR, Mathematics Series 6 (1942) 3-32.

E.L. Lehmann (1993), The Fisher, Neyman-Pearson theories of testing hypotheses:One theory or two? J. Amer. Statist. Assoc. 88 (1993) 1242-1249. MR1245356.

N. Mukhopadhyay (2006), Some comments on Hassairi et al.'s "Implicit distributions and estimation", Communications in Statistics—Theory and Methods 35 (2006) 293-297.

J. Neyman (1941), Fiducial argument and the theory of confidence intervals. Biometrika 32 (1941) 128-150. MR5582.

P.J.G. Pitman (1957), Statistics and science, Journal of the American Statistical Association 52 (1957) 322-330.

T. Schweder and N.L. Hjort (2002), Confidence and likelihood, Scand. J. Statist. 29 (2002) 309-332.

T. Schweder, N.L. Hjort (2003A), Frequentist analogies of priors and posteriors, in: Econometrics and the Philosophy of Economics, Princeton University Press, pp.285-317, 2003.

T. Schweder (2003B), Integrative fish stock assessment by frequentist methods: confidence distributions and likelihoods for howhead whales, Scientia Marina 67 (2003) 89-07.

K. Singh, M. Xie, W. Strawderman. (2001), Confidence distributions - concept, theory and applications, Technical report, Dept.Statistics, Rutgers Univ., Revised 2004

K. Singh, M. Xie, W.E. Strawderman (2005), Combining information from independent sources through confidence distributions, The Annals of Statistics 33 (2005) 159-183.

K. Singh, M. Xie, W. Strawderman. (2007), Confidence distributions (CD) - Distribution Estimator of a Parameter. Yehuda Vardi Volume. IMS-LNMS. (to appear).

G. Zech (1989), Upper limits in experiments with background or measurement errors, Nuclear Instruments and Methods in Physics Research A277 (1989) 608.