Parametric coverage interval

2007 Metrologia 44 L7

(http://iopscience.iop.org/0026-1394/44/2/N01)

IOP PUBLISHING METROLOGIA
Metrologia **44** (2007) L7–L9 doi:10.1088/0026-1394/44/2/N01

SHORT COMMUNICATION

Parametric coverage interval

Lin-An Chen[1], **Jing-Ye Huang**[2] **and Hung-Chia Chen**[1]

[1] Institute of Statistics, National Chiao Tung University, Hsinchu, Taiwan
[2] Department of Statistics, National Taichung Institute of Technology, Taichung, Taiwan

Received 9 January 2007
Published 14 March 2007
Online at stacks.iop.org/Met/44/L7

Abstract
Parametric estimation of coverage interval is useful since the parametric intervals are generally narrower than the non-parametric ones; however, it has been considered only for the measurement variable with normal distribution. Here we propose a general technique for constructing parametric coverage intervals that may deal with all distributions, both symmetric and asymmetric, in measurement science.

1. Introduction

The coverage interval, called the reference interval in laboratory chemistry, refers to population-based reference values obtained from a well-defined group of reference individuals. This is an interval with two confidence limits which covers the measurement values in the population in some probabilistic sense. It is useful for determining what should be the future measurement values, based upon present or past data. It assists in making major decisions, such as those pertaining to policy on environmental and occupational health about maximum acceptable chemical exposures. With its fundamental importance in clinical chemistry, toxicology, environmental health, quality control, etc (Holst and Christensen 1992), there are published standards for the requisite statistical methodology, e.g. the International Organization for Standardization (ISO 3534-1 1993, 3534-2 1993), the International Federation of Clinical Chemists (IFCC) (Solberg 1987a, 1987b, Dybkær and Solberg 1987).

The coverage interval can be estimated either parametrically or non-parametrically. The parametric method generally assumes that the underlying distribution of the measurement variable is normal, whereas the non-parametric approach estimates the quantiles directly. In general, the parametric intervals are narrower than the non-parametric intervals. Since the distribution of biological measurements is generally non-normal, constructing appropriate coverage intervals for non-normal distributions, particularly for asymmetric ones, is important in measurement science.

To develop the standards in the science of measurements for the application of statistical methods, Perruchet (2004) pointed out that possible damage may be caused to laboratories due to inconsistency in using concepts of statistics and metrology. From a metrological point of view, it has been agreed that collaboration should be established between professional organizations, industry and health authorities to achieve common reference intervals for the users. Thus, if the coverage interval's derivation with respect to various measurands is not consistent because they have different distributions, this may confuse users in the interpretation of these measurands. In fact, the choice of the coverage interval at a given coverage probability is not unique. The present paper proposes a technique that constructs the shortest one. We will show that this shortest coverage interval is consistent to deal with all distributions, symmetric or asymmetric.

2. Development of coverage interval

Fundamentally, a parametric coverage interval is an estimate of a fractile interval of fixed coverage probability γ (usually it is 0.95). The IFCC defines a 0.95 fractile interval as the interval between the 0.025 and 0.975 fractiles. In general, letting $F^{-1}(\delta)$ be the δth fractile for the measurement variable, the following interval

$$[F^{-1}(\delta),\ F^{-1}(\gamma+\delta)] \tag{1}$$

for some $\delta \in (0, 1-\gamma)$ is a fractile interval of coverage probability γ. Then, any estimate of (1) may serve as a γ coverage interval. Among the choices of fractile interval of (1) in terms of δ, we consider if there is an adaptive way to define a coverage interval appropriate for our consideration of consistency.

This proposal is clarified by its use for an application. In clinical chemistry, the coverage interval is also called a reference interval that refers to population-based measurement

values obtained from a group of normal or healthy people. Let f be the density function corresponding to the distribution function F for a measurement variable X of a normal person. Suppose that x_a is a measurement value, for person A, contained in a considered fractile interval. Then a value x_b, for person B, with $f(x_b) \geqslant f(x_a)$ indicates that B is more likely than A to be a normal person. Then, why should not we insert value x_b in this fractile interval? This scientific consideration for fractile interval of highest density values results in the shortest fractile interval. This proposal considers the shortest fractile interval in the form as

$$C(\gamma) = [F^{-1}(\delta^*),\, F^{-1}(\gamma + \delta^*)], \tag{2}$$

where $\delta^* = \delta^*(\gamma) = \arg_\delta \min_{0<\delta<1-\gamma}\{F^{-1}(\gamma+\delta) - F^{-1}(\delta)\}$. Here we assume a set of measurement variables $X_1, ..., X_n$ from this distribution with the distribution function F such that appropriate fractile estimates $\hat{F}^{-1}(\delta^*)$ and $\hat{F}^{-1}(\gamma + \delta^*)$ are available. Then,

$$\hat{C}(\gamma) = [\hat{F}^{-1}(\delta^*),\, \hat{F}^{-1}(\gamma + \delta^*)]$$

serves as a choice of coverage interval, called the shortest coverage interval for distribution F.

3. Consistency property of coverage interval

We then investigate the consistency property for this new coverage interval. Suppose that the measurement variable X has a unimodal distribution with symmetric density function f. Then the shortest fractile interval searches for δ that minimizes

$$\ell = F^{-1}(\gamma + \delta) - F^{-1}(\delta).$$

Now, solving δ for $0 = \frac{\partial \ell}{\partial \delta}$, we have

$$0 = \frac{\partial F^{-1}(\gamma+\delta)}{\partial \delta} - \frac{\partial F^{-1}(\delta)}{\partial \delta} = \frac{1}{f(F^{-1}(\gamma+\delta))} - \frac{1}{f(F^{-1}(\delta))},$$

which indicates that $f(F^{-1}(\gamma + \delta)) = f(F^{-1}(\delta))$. The symmetry assumption yields $\delta = \frac{1-\gamma}{2}$, which leads to the fractile interval $C(\gamma) = [F^{-1}(\frac{1-\gamma}{2}), F^{-1}(\frac{1+\gamma}{2})]$. When X has normal distribution $N(\mu, \sigma^2)$, the shortest γ fractile interval is $C(\gamma) = [\mu - z_{\frac{1+\gamma}{2}}\sigma, \mu + z_{\frac{1+\gamma}{2}}\sigma]$ which is identical to the classical normal fractile interval. The shortest and classical coverage intervals are then identical when parameters μ and σ have the same estimates. This verifies the consistency of using the same technique when X has a symmetric distribution.

4. Estimation of coverage interval

We now consider fractile estimates for the shortest coverage interval. With this setting in (2), although the coverage interval exists for any distribution, we will introduce only an important family of distributions as an example and several specific distributions for advanced study. Many distributions allow a fractile represented in an explicit form as $F^{-1}(\delta) = a(\delta, \theta)$ with a known function a so that the shortest coverage interval of (2) is $\hat{C}(\gamma) = [a(\delta^*, \hat{\theta}), a(\gamma + \delta^*, \hat{\theta})]$. We first consider a general distribution family that does have this desirable property. A measurement variable X with location-scale distribution has the density of the form $f(x, \theta_1, \theta_2) = \frac{1}{\theta_2} f_0(\frac{x-\theta_1}{\theta_2})$ with parameters $\theta_1 \in R$ and $\theta_2 > 0$ where f_0 is a parameter-free function. The δth fractile $F^{-1}(\delta)$ which satisfies $\delta = \int_{-\infty}^{F^{-1}(\delta)} \frac{1}{\theta_2} f_0(\frac{x-\theta_1}{\theta_2})\,\mathrm{d}x$ is $F^{-1}(\delta) = \theta_1 + \theta_2 F_0^{-1}(\delta)$ where F_0 is the distribution function with density f_0. This further yields the shortest quartile interval as

$$C(\gamma) = [\theta_1 + \theta_2 F_0^{-1}(\alpha^*),\, \theta_1 + \theta_2 F_0^{-1}(\gamma + \alpha^*)],$$

where

$$\alpha^* = \arg_\alpha \min_{0<\alpha<1-\gamma}(F_0^{-1}(\alpha+\gamma) - F_0^{-1}(\alpha)).$$

The shortest γ coverage interval is simply

$$\hat{C}(\gamma) = [\hat{\theta}_1 + \hat{\theta}_2 F_0^{-1}(\alpha^*),\, \hat{\theta}_1 + \hat{\theta}_2 F_0^{-1}(\gamma + \alpha^*)],$$

where $\hat{\theta}_1$ and $\hat{\theta}_2$ are estimates of θ_1 and θ_2 if they are available. We note that fractiles $F_0^{-1}(\alpha^*)$ and $F_0^{-1}(\gamma + \alpha^*)$ are evaluated from density f_0 and then they are free of parameters.

The measurement variable following a location-scale distribution has a fractile linear in location and scale parameters θ_1 and θ_2. Hence, a coverage interval may be simply obtained by replacing these two parameters by their estimates. On the other hand, the location-scale distribution family is a rich class of distributions that allows development of coverage intervals for many non-normal distributions to be accomplished. We do not go further on this point except to study several examples.

The two-parameter exponential distribution with density function $f(x, \theta, \lambda) = \lambda \mathrm{e}^{-\lambda(x-\theta)} I(x \geqslant \theta)$. This is a location-scale family distribution with location parameter θ and scale parameter $\frac{1}{\lambda}$ with fractile function $F^{-1}(u) = \theta - \lambda^{-1}\log(1-u)$, $0 < u < 1$. The shortest width γ fractile interval is

$$C(\gamma) = [\theta, \theta - \lambda^{-1}\log(1-\gamma)].$$

Based on maximum likelihood estimates of θ and λ with $\hat{\theta} = X_{n,1}$ and $\hat{\lambda} = \frac{1}{\bar{X} - X_{n,1}}$ where $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $X_{n,1}$ is the smallest variable of $X_1, ..., X_n$, the shortest coverage interval is

$$\hat{C}(\gamma) = [X_{n,1}, X_{n,1} - (\bar{X} - X_{n,1})\log(1-\gamma)].$$

In the case that θ is a known constant, we have

$$\hat{C}(\gamma) = [\theta, \theta - (\bar{X} - \theta)\log(1-\gamma)].$$

Furthermore, when $\theta = 0$, this involves the usual one-parameter exponential distribution and we have

$$\hat{C}(\gamma) = [0, -\bar{X}\log(1-\gamma)].$$

The Gamma distribution with density $f(x, \beta, \kappa) = \beta^\kappa x^{\kappa-1}\exp(-\beta x)/\Gamma(\kappa)$, $x > 0$, $\beta, \kappa > 0$ is also the $\chi^2_{2\kappa}/2\beta$ distribution. The γ fractile interval is

$$C(\gamma) = \left[\frac{1}{2\beta}G_{2\kappa}^{-1}\{\delta^*(\kappa)\},\, \frac{1}{2\beta}G_{2\kappa}^{-1}\{\gamma + \delta^*(\kappa)\}\right],$$

where $\delta^*(\kappa) = \arg_\delta \min_{0<\delta<1-\gamma}\{G_{2\kappa}^{-1}(\gamma+\delta) - G_{2\kappa}^{-1}(\delta)\}$ and G_ν is the distribution function of the chi-squared distribution with ν degrees of freedom.

The coverage interval based on the method of moments is of the form

$$\hat{C} = \left[\frac{\bar{X}}{2\hat{\kappa}} G^{-1}_{2\hat{\kappa}}\{\delta^*(\hat{\kappa})\}, \frac{\bar{X}}{2\hat{\kappa}} G^{-1}_{2\hat{\kappa}}\{\gamma + \delta^*(\hat{\kappa})\}\right], \qquad (3)$$

with $\delta^*(\kappa) = \arg_\delta \inf_{0<\delta<1-\gamma}\{G^{-1}_{2\kappa}(\gamma + \delta) - G^{-1}_{2\kappa}(\delta)\}$ and $\hat{\kappa} = \bar{X}^2/(n^{-1}\sum_{i=1}^n X_i^2 - \bar{X}^2)$. If κ is known, the coverage interval is in the form of (3), replacing $\hat{\kappa}$ by κ.

The Weibull distribution with density $f(x, \theta, \kappa) = \theta\kappa x^{\kappa-1}\exp(-\theta x^\kappa)$, $x > 0$, $\theta, \kappa > 0$, has the fractile function $F^{-1}(u) = \{-\theta^{-1}\log(1-u)\}^{1/\kappa}$. The γ fractile interval is

$$C(\gamma) = \left[\left[-\frac{1}{\theta}\log\{1 - \delta^*(\kappa)\}\right]^{\frac{1}{\kappa}}, \left[-\frac{1}{\theta}\log\{1 - \gamma - \delta^*(\kappa)\}\right]^{\frac{1}{\kappa}}\right],$$

where $\delta^*(\kappa) = \arg_\delta \inf_{0<\delta<1-\gamma}[\{-\log(1-\gamma-\delta)\}^{\frac{1}{\kappa}} - \{-\log(1-\delta)\}^{\frac{1}{\kappa}}]$. The coverage interval is

$$\hat{C}(\gamma) = \left[\left[-\frac{1}{\hat{\theta}}\log\{1 - \delta^*(\hat{\kappa})\}\right]^{\frac{1}{\hat{\kappa}}}, \left[-\frac{1}{\hat{\theta}}\log\{1 - \gamma - \delta^*(\hat{\kappa})\}\right]^{\frac{1}{\hat{\kappa}}}\right],$$

where $\delta^*(\kappa) = \arg_\delta \inf_{0<\delta<1-\gamma}[\{-\log(1-\gamma-\delta)\}^{\frac{1}{\kappa}} - \{-\log(1-\delta)\}^{\frac{1}{\kappa}}]$ and $\hat{\theta}$ and $\hat{\kappa}$ are estimates of θ and κ, respectively. The maximum likelihood estimates of θ and κ may be handled by many software packages.

The parametric coverage intervals may be derived for these distributions such that their fractile intervals of (2) are available. Although we have not studied them, most distributions appearing in the literature are satisfied with this condition. If the fractile interval of (2) is not available for a specific distribution, a non-parametric type shortest coverage interval may be derived if we substitute distribution function F in (1) by the empirical distribution function F_n which is free of parameters. We will not do this further in this paper.

Acknowledgments

The authors are grateful to the Editor and a referee for comments which greatly improved the presentation of this paper. This research work was partially supported by the National Science Council of Taiwan, Grant No NSC 95-2118-M-009-007.

References

Dybkær R and Solberg H E (International Federation of Clinical Chemistry (IFCC)) 1987 Approved recommendation (1987) on the theory of reference values: 6. Presentation of observed values related to reference values *Clin. Chim. Acta* **170** 33–42; *J. Clin. Chem. Clin. Biochem.* **25** 657–62

Holst E and Christensen J M 1992 Intervals for the description of the biological level of a trace element in a reference population *Statistician* **41** 233–42

ISO 3534-1 1993 Statistics—vocabulary and symbols—part 1: basic statistical terms and concepts (Geneva: International Organization for Standardization)

ISO 3534-2 1993 Statistics—vocabulary and symbols—part 2: statistical quality control (Geneva: International Organization for Standardization)

Perruchet C 2004 Some differences between the applied statistical approach for measurement uncertainty theory and the traditional approach in metrology and testing *Advanced Mathematical and Computational Tools in Metrology VI* ed P Ciarlini and F Pavese (Singapore: World Scientific)

Solberg H E (International Federation of Clinical Chemistry (IFCC)) 1987a Approved recommendation (1986) on the theory of reference values: 1. The concept of reference values *Ann. Biol. Clin.* **45** 237–41; *Clin. Chim. Acta* **165** 111–18; *J. Clin. Chem. Clin. Biochem.* **25** 337–42

Solberg H E (International Federation of Clinical Chemistry (IFCC)) 1987b Approved recommendation (1987) on the theory of reference values: 5. Statistical treatment of collected reference values. Determination of reference limits *Clin. Chim. Acta* **170** 13–32; *J. Clin. Chem. Clin. Biochem.* **25** 645–56