IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 84–94

DOI: 10.1214/10-IMSCOLL709

A note on reference limits

Jing-Ye Huang [1,*] **Lin-An Chen** [2,†] **and A. H. Welsh** [3,§,‡]

National Taichung Institute of Technology, National Chiao Tung University† and The Australian National University‡*

Abstract: We introduce a conceptual framework within which the problem of setting reference intervals is one of estimating population parameters. The framework enables us to broaden the possibilities for inference by showing how to create confidence intervals for population intervals. We propose a new kind of interval (the γ-mode interval) as the population parameter of interest and show how to estimate and make optimal inference about this interval. Finally, we clarify the relationship between our reference intervals and other types of intervals.

1. Introduction

Reference limits are fundamentally important in clinical chemistry, toxicology, environmental health, metrology (the study of measurement), quality control, engineering and industry (Holst & Christensen citer8) and there are published standards for their statistical methodology; see for example the International Standards Organisation (ISO 3534-1, 1993; 3534-2, 1993), the International Federation of Clinical Chemists (IFCC) (Solberg, [19], [20], Peticlerc & Solberg [16], Dybkær & Solberg [4], National Committee for Clinical Laboratory Standards (NCCLS C28-A2 [12]) and the International Union of Pure and Applied Chemistry (IUPAC) (Poulsen, Holst & Christensen [17]). The purpose of this paper is to discuss reference limits from a more statistical perspective.

Suppose that we have a sample $X_1, \ldots, X_n$ of size $n \geq 1$ of independent observations from the distribution $F(\cdot; \theta)$ with unknown parameter θ. The reference limit problem is to use the sample to construct an interval for an unobserved statistic $w = w(Z_1, \ldots, Z_m)$, $m \geq 1$, which has distribution function $F_w(\cdot; \theta)$ when $Z_1, \ldots, Z_m$ have the same distribution $F(\cdot; \theta)$ as $X_1, \ldots, X_n$. The statistic w is often the sample mean $\bar{Z} = m^{-1} \sum_{i=1}^{m} Z_i$ or, when $m = 1$, a single observation, but the general formulation is useful.

The IFCC standard (γ-content) reference interval for w is an estimate of the inter-fractile interval $C^{if}_{w,\gamma}(\theta) = [F_w^{-1}\{(1-\gamma)/2; \theta\}, F_w^{-1}\{(1+\gamma)/2; \theta\}]$, often with $\gamma = 0.95$. This target interval ensures the intuitive requirement that a reference

[1]Department of Statistics, National Taichung Institute of Technology, 129 Sanmin Rd., Sec. 3, Taichung, Taiwan, e-mail: hjy@ntit.edu.tw
[2]Institute of Statistics, National Chiao Tung University, 1001 Ta Hsueh Rd. Hsinchu, Taiwan, e-mail: lachen@stat.nctu.edu.tw
[3]Centre for Mathematics and its Applications, The Australian National University, Canberra ACT 0200, Australia, e-mail: Alan.Welsh@anu.edu.au
§Research supported by Australian Research Council DP0559135.
AMS 2000 subject classifications: Primary 62F25; secondary 62G15.
Keywords and phrases: confidence interval, coverage interval, inter-fractile interval, mode interval, reference interval, reference limits, tolerance interval.

interval represent a specified proportion of the central values obtained in the reference population is satisfied. The standard requires either applying a known (possibly identity) transformation to the data, estimating the normal version of $C^{if}_{w,\gamma}(\theta)$ and then retransforming to obtain the interval, or a nonparametric approach which estimates $C^{if}_{w,\gamma}(\theta)$ directly. The IFCC recommends that the reference interval be reported with $1-\alpha$ (usually $1-\alpha = 0.95$) confidence intervals for the endpoints of $C^{if}_{w,\gamma}(\theta)$.

It is useful to see how the IFCC standard works in a simple example. Suppose that w is a single observation from an exponential distribution with mean θ and the estimation sample is from the same distribution. For the parametric approach, there is no transformation (not depending on θ) that produces exact normality but we can apply transformations which stabilise the variance ($g(x) = \log(x)$) or symmetrise the distribution ($g(x) = x^{1/3}$). In either case, let $A_g = n^{-1}\sum_{i=1}^n g(X_i)$ and $S_g = (n-1)^{-1}\sum_{i=1}^n \{g(X_i) - A_g\}^2$ be the sample mean and variance of the transformed data. Then the IFCC 95% reference interval is $[g^{-1}(A_g - 1.96S_g), g^{-1}(A_g + 1.96S_g)]$. Since $\mathrm{E}\{g(X)\} \approx g(\theta)$ and $\mathrm{var}\{g(X)\} \approx \theta^2 g'(\theta)^2$, the reference interval is estimating $[0.14080\theta, 7.099\theta]$ when we use the logarithmic transformation and $[0.0416\theta, 4.5194\theta]$ when we use the cube root transformation. The actual coverage of these intervals is 0.868 (length $= 6.9582\theta$) and 0.948 (length 4.478θ) respectively. The nonparametric approach produces an estimate of the 95% inter-fractile interval $[0.0253\theta, 3.6889\theta]$ (length $= 3.664\theta$). None of these intervals includes the region around zero, the region of highest probability for the exponential distribution.

The exponential example shows that we need a conceptual framework to evaluate reference intervals and unambiguous, interpretable methods for constructing reference intervals with desirable properties. Our approach developed in Section 2 is to treat underlying population intervals (such as $C^{if}_{w,\gamma}(\theta)$ or $[\mu_w - k\sigma_w, \mu_w + k\sigma_w]$) as parameters and then consider estimating and making inference about them. In this framework, reference intervals are 'point estimates' of underlying intervals so we can use well-established ideas to evaluate and interpret them. The only new issue is that the unknown parameter is an interval rather than a familiar vector. The treatment of an interval as an unknown parameter is arguably implicit in the statistical literature (for example in Carroll & Ruppert [1]) but it is useful to make it explicit in the present context because it enables us to separate discussion of the choice of parameter from discussion of alternative estimators and methods of inference. As we discuss in Section 3, it also allows us to relate reference intervals to well-known intervals for future observations such as prediction and tolerance intervals.

In this paper, we also propose that reference intervals be based on a new γ-content interval $C_{w,\gamma}(\theta)$ defined in Section 2 which we call the γ-mode interval, rather than the inter-fractile interval $C^{if}_{w,\gamma}(\theta)$. The γ-mode interval is the same as the inter-fractile interval when w has a unimodal, symmetric distribution; it is a more appropriate and useful interval when w has an asymmetric distribution which cannot be transformed directly to normality. For a single observation from an exponential distribution, the 95%-mode interval is $[0, \lambda^{-1}2.9957]$ which is shorter than the other intervals we examined and includes the mode of the distribution. The γ-mode interval contains the highest density points in the sample space so has the highest-density property used as a starting point by Eaton et al.[5] for their discussion of multivariate reference intervals for the multivariate normal distribution. Even for the multivariate normal distribution, multivariate reference intervals are difficult to obtain; for recent results, see for example Trost [21] and Eaton et al. [5].

We define the intervals and present some results on optimal confidence intervals in Section 2. We discuss in detail the relationship between reference and confidence intervals for γ-mode intervals and prediction and tolerance intervals in Section 3. We illustrate the methodology and explore the relationships between the different kinds of intervals further in the Gaussian and Gamma cases in Sections 4 and 5 respectively. We restrict ourselves to these simple cases so that we can obtain explicit results and make comparisons with other methods in the literature: the results can be extended to other statistics w and other models such as regression and generalized linear models in which one or more model parameters are functions of known covariates.

Although our present focus is on parametric methods, we have developed a nonparametric approach (using order statistics) when w is a single observation (so $F_w = F$). However, the approach is difficult to apply with complex, structured data, when w is a more general statistic, is less efficient than the parametric methods, and the confidence intervals perform poorly in small samples (because tail quantiles are difficult to estimate). Parametric methods overcome these difficulties at the cost of requiring more careful model examination (including diagnostics) and consideration of robustness. At least when the model holds, parametric and nonparametric methods should estimate the same interval. This is the case with our methodology but not with the IFCC method where parametric estimation can lead to estimating a different interval from the one we have specified (which we can interpret as bias) and does not necessarily yield efficient estimators (in the sense that their variance is larger than necessary).

2. Definitions and results

A random interval $\hat{C} = [\hat{a}, \hat{b}]$ is an *unbiased estimator* of a nonrandom interval $C(\theta) = [a(\theta), b(\theta)]$ if $E_\theta[\text{Length}\{(\hat{C} \cap C(\theta)^c) \cup (\hat{C}^c \cap C(\theta))] = 0$ and a *consistent estimator* of $C(\theta)$ if $\Pr_\theta[\text{Length}\{(\hat{C} \cap C(\theta)^c) \cup (\hat{C}^c \cap C(\theta))\} > \epsilon] \to 0$ for all $\epsilon > 0$. That is, the length of the region in which the intervals do not overlap has expectation zero or tends to zero in probability. We can show that an interval is unbiased or consistent if $\ell\hat{a} + (1-\ell)\hat{b}$ is unbiased or consistent for $\ell a(\theta) + (1-\ell)b(\theta)$, $0 \le \ell \le 1$. Thus the discussion of separate maximum likelihood and uniformly minimum variance unbiased estimation of the endpoints of the normal inter-fractile interval in Trost [21] immediately applies to estimation of that interval as a single parameter.

A $100(1-\alpha)\%$ *confidence interval* for $C(\theta)$ is a realisation of a random interval $\hat{C}_\alpha = [\hat{a}_\alpha, \hat{b}_\alpha]$ which satisfies

$$P_\theta(\hat{a}_\alpha \le a(\theta) < b(\theta) \le \hat{b}_\alpha) = P_\theta\{\hat{C}_\alpha \supseteq C(\theta)\} = 1 - \alpha \quad \text{for all } \theta.$$

To develop an optimality theory based on the concept of *uniformly most accurate (UMA)* confidence intervals, we define a $100(1-\alpha)\%$ confidence interval $\hat{C}_\alpha$ for $C(\theta)$ to be *type I UMA* if

$$P_\theta\{\hat{C}_\alpha \supseteq C(\theta')\} \le P_\theta\{\hat{C}^*_\alpha \supseteq C(\theta')\}, \qquad \text{for all } \theta' < \theta,$$

type II UMA if

$$P_\theta\{\hat{C}_\alpha \supseteq C(\theta')\} \le P_\theta\{\hat{C}^*_\alpha \supseteq C(\theta')\}, \qquad \text{for all } \theta' > \theta,$$

for any other $100(1-\alpha)\%$ confidence interval $\hat{C}^*_\alpha$ for $C(\theta)$. A $100(1-\alpha)\%$ confidence interval $\hat{C}_\alpha$ for $C(\theta)$ is *unbiased* if

$$P_\theta\{\hat{C}_\alpha \supseteq C(\theta')\} \le 1 - \alpha, \qquad \text{for all } \theta \neq \theta',$$

and *UMA unbiased* if it is unbiased and

$$P_\theta\{\hat{C}_\alpha \supseteq C(\theta')\} \le P_\theta\{\hat{C}^*_\alpha \supseteq C(\theta')\}, \quad \text{for all } \theta \ne \theta',$$

for any other $100(1-\alpha)\%$ unbiased confidence interval $\hat{C}^*_\alpha$ for $C(\theta)$.

The following theorem shows how to construct optimal confidence intervals for a wide class of fixed intervals, including many of the intervals of interest to us.

Theorem 2.1. Consider the interval $C(\theta) = [a(\theta), b(\theta)]$, where θ is a scalar unknown parameter and a and b are increasing functions of θ. Let $\hat{T} = [\hat{\theta}_1, \hat{\theta}_2]$ be an interval with $\hat{\theta}_1 < \hat{\theta}_2$ and define $\hat{C}_T = [a(\hat{\theta}_1), b(\hat{\theta}_2)]$.

i) If $\hat{T}$ is a $100(1-\alpha)\%$ confidence interval for θ, then $\hat{C}_T$ is a $100(1-\alpha)\%$ confidence interval for $C(\theta)$.
ii) If $\hat{T}$ is a $100(1-\alpha)\%$ unbiased confidence interval for θ, then $\hat{C}_T$ is a $100(1-\alpha)\%$ unbiased confidence interval for $C(\theta)$.
iii) If $\hat{T}$ is a $100(1-\alpha)\%$ UMA unbiased confidence interval for θ, then $\hat{C}_T$ is a $100(1-\alpha)\%$ UMA unbiased confidence interval for $C(\theta)$.

Proof. As a and b are monotone increasing, we have that for any θ'

$$\{\hat{C}_T \supseteq C(\theta')\} = \{a(\hat{\theta}_1) \le a(\theta') < b(\theta') \le b(\hat{\theta}_2)\} \Leftrightarrow \{\hat{\theta}_1 \le \theta' \le \hat{\theta}_2\}$$

so

$$P_\theta\{\hat{C}_T \supseteq C(\theta')\} = P_\theta(\hat{\theta}_1 \le \theta' \le \hat{\theta}_2).$$

The results i) and ii) follow from the definitions of confidence intervals and unbiased confidence intervals. For iii), suppose that $\hat{C}^*_\alpha$ is a $100(1-\alpha)\%$ confidence interval for $C(\theta)$ and consider the set $\hat{T}^* = \{\theta : C(\theta) \subset \hat{C}^*_\alpha\}$. Then, for any θ',

$$P_\theta(\theta' \in \hat{T}^*) = P_\theta\{C(\theta') \subset \hat{C}^*_\alpha\}$$

so setting $\theta' = \theta$, we see that $\hat{T}^*$ is a $100(1-\alpha)\%$ confidence set for θ and setting $\theta' \ne \theta$, we see that $\hat{T}^*$ is an unbiased $100(1-\alpha)\%$ confidence set for θ whenever $\hat{C}^*_\alpha$ is a unbiased $100(1-\alpha)\%$ confidence interval for $C(\theta)$. Since $\hat{T}$ is a $100(1-\alpha)\%$ UMA unbiased confidence set for θ

$$P_\theta\{\hat{C}_T \supseteq C(\theta')\} = P_\theta(\hat{\theta}_1 \le \theta' \le \hat{\theta}_2) \le P_\theta(\theta' \in \hat{T}^*) = P_\theta\{\hat{C}^*_\alpha \supseteq C(\theta')\}$$

and the result obtains. □

The theorem can be applied with a and b decreasing if we reparametrize the model and write a and b as increasing functions of the transformed parameter.

A slightly different approach is required for the case that one endpoint of the interval $C(\theta)$ is known.

Theorem 2.2. Consider the interval $C(\theta) = [a, b(\theta)]$, where a is known and b is a monotone increasing function of a scalar unknown parameter θ, or $C(\theta) = [a(\theta), b]$, where a is a monotone increasing function of a scalar unknown parameter θ and b is known. Let $\hat{T} = (-\infty, \hat{\theta}_2]$ be an upper interval or $\hat{T} = [\hat{\theta}_1, \infty)$ be a lower interval according to whether a is known or b is known, and define $\hat{C}_T = [a, b(\hat{\theta}_2)]$, if a is known, or $\hat{C}_T = [a(\hat{\theta}_1), b]$, if b is known.

i) If $\hat{T}$ is a $100(1-\alpha)\%$ upper/lower confidence interval for θ, then $\hat{C}_T$ is a $100(1-\alpha)\%$ confidence interval for $C(\theta)$ with a/b known.
ii) If $\hat{T}$ is a $100(1-\alpha)\%$ UMA upper/lower confidence interval for θ, then $\hat{C}_T$ is a $100(1-\alpha)\%$ type I/type II UMA confidence interval for $C(\theta)$ with a/b known.

Proof. The proof is similar to that of Theorem 1 using the relations $\{a \leq b(\theta) \leq b(\hat{\theta}_2)\} \Leftrightarrow \{\theta \leq \hat{\theta}_2\}$ when a is known and $\{a(\hat{\theta}_1) \leq a(\theta) \leq b\} \Leftrightarrow \{\hat{\theta}_1 \leq \theta\}$ when b is known. □

A much simpler but more restricted theory for optimal confidence intervals based directly on the length or the length on the log scale can be constructed in particular cases.

Theorem 2.3. Suppose that the interval $C(\theta) = [a(\theta),\, b(\theta)]$ is a location interval so that $a(\theta) = \theta + k_1$ and $b(\theta) = \theta + k_2$ or a scale interval so that $a(\theta) = k_1\theta$ and $b(\theta) = k_2\theta$ with $k_1, k_2 \neq 0$. Then in the location/scale case, if $\hat{T}$ is the shortest/log-shortest $100(1-\alpha)\%$ confidence interval for θ, it follows that $\hat{C}_T$ is the shortest/log-shortest $100(1-\alpha)\%$ confidence interval for $C(\theta)$.

Proof. For the location family, the length of $\hat{C}$ is

$$\text{length}(\hat{C}_T) = b(\hat{\theta}_2) - a(\hat{\theta}_1) = \text{length}(\hat{T}) + k_2 - k_1$$

and for the scale family

$$\text{length}_{\log}(\hat{C}_T) = \log b(\hat{\theta}_2) - \log a(\hat{\theta}_1) = \text{length}_{\log}(\hat{T}) + \log k_2 - \log k_1$$

and the result follows from the fact that $k_2 - k_1$ and $\log k_2 - \log k_1$ are fixed. □

The intuitive meaning of the above results is that good confidence intervals for $C(\theta)$ are obtained from good confidence intervals for θ. Not surprisingly, the case in which θ is a vector parameter is much more difficult to handle; exact intervals can only be constructed in particular cases (for an example, see Section 4) but we can construct asymptotic intervals.

The above results apply to any kind of interval; we now turn our attention to a particular type of interval. A γ*-content interval* for w is a nonrandom interval $C_{w,\gamma}(\theta) = [a_{w,\gamma}(\theta), b_{w,\gamma}(\theta)]$ which satisfies $\Pr_\theta\{w \in C_{w,\gamma}(\theta)\} = F_w\{b_{w,\gamma}(\theta);\theta\} - F_w\{a_{w,\gamma}(\theta);\theta\} = \gamma$. Note that $C_{w,\gamma}(\theta)$ is non-random so tolerance intervals are not γ-content intervals in this sense; see Section 3 for further discussion.

A *reference interval* for w is an estimate of a γ-content interval for w. A confidence interval for an interval captures the uncertainty in estimating the interval and provides an estimate with the same content as the interval with confidence $1-\alpha$. i. e. a $1-\alpha$ *confidence interval for a* γ*-content interval* is a γ-content interval with confidence $1-\alpha$.

Consider the class of γ-content intervals $C_{w,\gamma,\delta}(\theta) = [F_w^{-1}(\delta;\theta), F_w^{-1}(\gamma+\delta;\theta)]$, $0 < \delta < 1-\gamma$, where δ is a location constant to be chosen by the user. These intervals include the inter-fractile intervals when $\delta = (1-\gamma)/2$ but are more flexible. A γ*-mode interval* is the shortest interval in the class $C_{w,\gamma,\delta}(\theta)$, namely $C_{w,\gamma}(\theta) = C_{w,\gamma,\delta^*}(\theta)$, where $\delta^* = \delta^*(\gamma,\theta) = \arg_\delta \min_{0<\delta<1-\gamma} F_w^{-1}(\gamma+\delta;\theta) - F_w^{-1}(\delta,\theta)$. A γ-mode interval always contains the highest density points in the sample space and, if it is unique, the mode of F_w (c.f. Eaton et al.[5] . We propose that reference intervals be based on γ-mode intervals instead of inter-fractile intervals.

3. Relationships with other intervals

Reference intervals and confidence intervals for population intervals are related to prediction, expectation tolerance and tolerance intervals. These are realisations of random intervals (L, U) which satisfy

$P_\theta(L \leq w \leq U) = \gamma$ (prediction interval),
$EP_\theta\left(F_w(U;\theta) - F_w(L;\theta)\right) = \gamma$ (expectation tolerance interval) or
$P_\theta\left(F_w(U;\theta) - F_w(L;\theta) \geq \gamma\right) = 1 - \alpha$ (γ-level tolerance interval)
respectively. Prediction intervals are expectation tolerance intervals because

$$E_\theta\left(F_w(U;\theta) - F_w(L;\theta)\right) = E_\theta\left(P_\theta(L \leq w \leq U|L,U)\right) = P_\theta(L \leq w \leq U) = \gamma$$

although the converse is not true. Prediction intervals are interpreted in this way in the IUPAC recommendations where they are called coverage intervals (Poulsen et al. 1997). Tolerance intervals (see for example Wilks [25], Wald [23], Paulson [15], Guttman [8], Patel [14] and Krishnamoorthy and Mathew [11]) are conceptually more complicated. These definitions do not involve a non-stochastic population interval so they are not γ-content intervals in the sense used in this paper. We have the following result.

Theorem 3.1. Suppose that $[a_{w,\gamma}(\theta), b_{w,\gamma}(\theta)]$ is a γ-content interval for w. If F_w is continuous at $a_{w,\gamma}(\theta)$ and $b_{w,\gamma}(\theta)$, then a reference interval which is consistent for $[a_{w,\gamma}(\theta), b_{w,\gamma}(\theta)]$ is an asymptotic γ-level prediction and expectation tolerance interval for w. A $100(1-\alpha)\%$ confidence interval for $[a_{w,\gamma}(\theta), b_{w,\gamma}(\theta)]$ is a $100(1-\alpha)\%$ γ-level tolerance interval for w.

Proof. Suppose that $[\hat{a}, \hat{b}]$ is a consistent estimator of $[a_{w,\gamma}(\theta), b_{w,\gamma}(\theta)]$. Then

$$P_\theta\left(\hat{a} \leq w \leq \hat{b}\right) = E_\theta\{P_\theta\left(\hat{a} \leq w \leq \hat{b}|\hat{a}, \hat{b}\right)\} = E_\theta\left(F_w(\hat{b};\theta) - F_w(\hat{a};\theta)\right) \to \gamma,$$

as $n \to \infty$ and the first part obtains. Next, suppose that $[\hat{a}_\alpha, \hat{b}_\alpha]$ is a $100(1-\alpha)\%$ confidence interval for $[a_{w,\gamma}(\theta), b_{w,\gamma}(\theta)]$. Then

$$\begin{aligned}
& P_\theta\left(F_w(\hat{b}_\alpha;\theta) - F_w(\hat{a}_\alpha;\theta) \geq \gamma\right) \\
&= P_\theta\left(F_w(\hat{b}_\alpha;\theta) - F_w(\hat{a}_\alpha;\theta) \geq F_w(b_{w,\gamma}(\theta);\theta) - F_w(a_{w,\gamma}(\theta);\theta)\right) \\
&\geq P_\theta\left(F_w(\hat{a}_\alpha;\theta) \leq F_w(a_{w,\gamma}(\theta);\theta) < F_w(b_{w,\gamma}(\theta);\theta) \leq F_w(\hat{b}_\alpha;\theta)\right) \\
&\geq 1 - \alpha
\end{aligned}$$

so $[\hat{a}_\alpha, \hat{b}_\alpha]$ is a $100(1-\alpha)\%$ γ-level tolerance interval for w. □

Reference intervals are good prediction intervals (appropriate for making one or a few predictions) because, as pointed out by Carroll & Ruppert [1], adjustments for estimation uncertainty in prediction intervals are typically of order $1/n$. The confidence intervals adjust for estimation uncertainty at order $1/n^{1/2}$ so it is interesting that these relate to tolerance intervals. Tolerance intervals cannot generally be interpreted as confidence intervals for a population γ-content interval C (Willink [24], Chen and Hung [2]) because there are shorter tolerance intervals which do not have the coverage property of the confidence intervals.

Poulsen et al. [17] recommended that their coverage (prediction or expectation tolerance) intervals be reported with the coverage uncertainty, the value of β making

$$P_\theta\left(\gamma - \beta \leq F_w(U;\theta) - F_w(L;\theta) \leq \gamma + \beta\right) = 1 - \alpha.$$

The coverage uncertainty is the adjustment required to make the coverage interval a $(\gamma-\beta)$-level $100(1-\alpha)\%$ tolerance interval. It seems more useful to construct directly intervals which achieve a chosen level. For coverage intervals, this leads to reporting tolerance intervals analogously to the way we recommend reporting confidence intervals; for reference levels, it leads to reporting confidence intervals for population intervals.

More insight can be achieved by comparing reference intervals and $100(1-\alpha)\%$ confidence intervals for γ-mode intervals to prediction and tolerance intervals in some simple cases. These calculations are presented in the following Sections. Some other examples of reference intervals but without confidence intervals are given by Chen et al. [3].

4. The Gaussian distribution

In this section, we derive the γ-mode, reference and confidence intervals for the Gaussian model.

4.1. The γ-mode interval

The mean $\bar{Z}$ of m independent $\mathrm{N}(\mu, \sigma^2)$ random variables has a $\mathrm{N}(\mu, m^{-1}\sigma^2)$ distribution so the quantile function is $F_{\bar{Z}}^{-1}(u) = \mu + m^{-1/2}\sigma\Phi^{-1}(u)$, where Φ is the standard Gaussian cumulative distribution function. The location constant δ^* in the γ-mode interval satisfies the estimating equation $\phi\left(\Phi^{-1}(\gamma+\delta^*)\right) = \phi\left(\Phi^{-1}(\delta^*)\right)$, where ϕ is the standard Gaussian density function. Since ϕ is symmetric, δ^* satisfies $\Phi^{-1}(\gamma+\delta^*) = -\Phi^{-1}(\delta^*) = \Phi^{-1}(1-\delta^*)$, so $\delta^* = (1-\gamma)/2$. It follows that the γ-mode interval for the mean of $m \geq 1$ observations is

$$(1) \quad C_{\bar{Z},\gamma}(\mu,\sigma) = \left[\mu - \Phi^{-1}\left\{(1+\gamma)/2\right\}\sigma/m^{1/2},\ \mu + \Phi^{-1}\left\{(1+\gamma)/2\right\}\sigma/m^{1/2}\right]$$

which is centered at the mode μ.

4.2. The reference interval

We construct the reference interval by estimating (1). Suppose that $X_1, \ldots, X_n$ are independent $\mathrm{N}(\mu, \sigma^2)$ random variables. Then in (1) we can replace μ by the sample mean $\bar{X}$ and σ by the scaled sample standard deviation $c_n S$, where c_n is a non-stochastic function of n. The maximum likelihood estimator of $C_{\bar{Z},\gamma}(\mu,\sigma)$ has $c_n = \{(n-1)/n\}^{1/2}$; the uniformly minimum variance unbiased estimator has $c_n = \{(n-1)/2\}^{1/2}\Gamma\{(n-1)/2\}/\Gamma(n/2)$ etc.

4.3. The confidence interval with known variance

When the underlying variance σ^2 known, it follows from Theorems 2.1 and 2.3 that a $100(1-\alpha)\%$ UMA unbiased and shortest confidence interval for (1) with $m = 1$ is

$$(2) \qquad [\bar{X} - k_n^*(\gamma, 1-\alpha)\sigma,\ \bar{X} + k_n^*(\gamma, 1-\alpha)\sigma],$$

where $k_n^*(\gamma, 1-\alpha) = \Phi^{-1}\left\{(1+\gamma)/2\right\} + n^{-1/2}\Phi^{-1}(1-\alpha/2)$.

The interval (2) is also the mean-based γ-level $100(1-\alpha)\%$ two-sided tolerance interval constructed by Owen [13] to control both tails. On the other hand, a widely used mean-based γ-level $100(1-\alpha)\%$ two-sided tolerance interval (see for example

Proschan, 1953 p. 560) is of the same form as (2) but with $k_n^* = k_n^*(\gamma, 1-\alpha)$ satisfying

$$\gamma = \Phi\left(n^{-1/2}\Phi^{-1}(1-\alpha/2) + k_n^*\right) - \Phi\left(n^{-1/2}\Phi^{-1}(1-\alpha/2) - k_n^*\right).$$

Comparison of the values of $k_n^*(\gamma, 1-\alpha)$ in this interval and (2) shows that the confidence interval is wider than this tolerance interval (so the tolerance interval undercovers the γ-mode interval). When $\gamma = 1-\alpha$, we can write $k_n^*(\gamma, 1-\alpha)$ in (2) as $k_n^* = \Phi^{-1}(1-\alpha/2)(1+n^{-1/2})$. This interval resembles but is wider than the $100(1-\alpha)\%$ prediction interval in which $1+n^{-1/2}$ is replaced by $(1+n^{-1})^{1/2}$. These intervals are both wider than the naive prediction interval ($\alpha = 1$) which effectively replaces $1+n^{-1/2}$ by 1. These calculations confirm the general relationships between the different intervals.

4.4. The confidence interval with unknown variance

Suppose now that the underlying distribution has both parameters unknown. Let $T_\nu(\cdot;\eta)$ be the distribution function of the noncentral t-distribution with ν degrees of freedom and noncentrality parameter η. Then we can show that a $100(1-\alpha)\%$ confidence interval for (1) with $m = 1$ is

$$\left[\bar{X} - k_n(\gamma, 1-\alpha)S, \bar{X} + k_n(\gamma, 1-\alpha)S\right], \tag{3}$$

where $k_n(\gamma, 1-\alpha) = n^{-1/2}T_{n-1}^{-1}\left[1-\alpha/2; n^{1/2}\Phi^{-1}\{(1+\gamma)/2\}\right]$.

The confidence interval (3) is the mean and variance based γ-level $100(1-\alpha)\%$ two-sided tolerance interval controlling both tails. Alternative mean and variance based γ-level $100(1-\alpha)\%$ tolerance intervals have been given by Wald & Wolfowitz [23] and Howe [10]. The $100\gamma\%$ prediction interval is of the same form as (3) with $k_n(\gamma) = T_{n-1}^{-1}((1+\gamma)/2)(1+n^{-1})^{1/2}$ and the relationships between these intervals are the same as when σ is known.

5. The Gamma distribution

In this section, we derive the γ-mode, reference and confidence intervals for the Gamma model.

5.1. The γ-mode interval

The Gamma distribution $\gamma(\kappa, \theta)$ with density $f(x, \theta, \kappa) = x^{\kappa-1}\exp(-x/\theta)/\theta^\kappa\Gamma(\kappa)$, $x > 0$, θ, $\kappa > 0$ is also the $\theta\chi^2_{2\kappa}/2$ distribution. The mean $\bar{Z}$ of $m \geq 1$ independent observations from this distribution has a $\theta\chi^2_{2m\kappa}/2m$ distribution so the γ-mode interval for the mean of $m > 1/\kappa$ observations is

$$C_{\bar{Z},\gamma}(\theta) = \left[\theta G_{2m\kappa}^{-1}\{\delta^*(\kappa)\}/2m, \theta G_{2m\kappa}^{-1}\{\gamma + \delta^*(\kappa)\}/2m\right], \tag{4}$$

where $\delta^*(\kappa) = \arg_\delta\inf_{0<\delta<1-\gamma} G_{2m\kappa}^{-1}(\gamma+\delta) - G_{2m\kappa}^{-1}(\delta)$ and G_ν is the cumulative distribution function of the chi-squared distribution with ν degrees of freedom. The mode is $\theta(m\kappa - 1)/m$ when $m > 1/\kappa$ and zero when $m \leq 1/\kappa$. Provided $\kappa > 1$, when $m = 1$, (4) is also the γ-mode interval for a single observation. However, when $\kappa = 1$ (i. e. the exponential distribution), (4) with $\kappa = 1$ gives the γ-mode interval for the sample mean of $m \geq 2$ observations but the γ-mode interval for a single observation is

$$C_{Z,\gamma}(\theta) = [0,\ -\theta \log(1-\gamma)]. \tag{5}$$

The mode 0 is always in this interval.

5.2. The reference interval

Suppose that $X_1, \ldots, X_n$ are independent $\gamma(\kappa, \theta)$ random variables. The maximum likelihood estimator $\hat{\kappa}$ of κ satisfies $\psi(\kappa) - \log(\kappa) = n^{-1} \sum_i^n \log(X_i/\bar{X})$ with $\psi(\cdot)$ the digamma function and, the method of moments estimator, $\hat{\kappa} = n\bar{X}^2 / \sum_i^n (X_i - \bar{X})^2$. In either case, we estimate θ by $\bar{X}/\hat{\kappa}$. If κ is known, both estimators are obtained by replacing $\hat{\kappa}$ by κ in (4). The maximum likelihood estimator of (5) is $\hat{C}_{Z,\gamma}(\theta) = [0,\ -\bar{X} \log(1-\gamma)]$.

5.3. Confidence intervals

Suppose initially that the shape parameter $\kappa > 1$ is known so the γ-mode interval is (4). Choose g and h to satisfy $1 - \alpha = \Pr(g < \chi^2_{2n\kappa} < h)$. Then, from Theorem 2.1, a $100(1-\alpha)\%$ confidence interval for (4) with $m = 1$ is

$$\left[n\bar{X} G^{-1}_{2\kappa}(\delta^*)/h,\ n\bar{X} G^{-1}_{2\kappa}(\gamma + \delta^*)/g\right]. \tag{6}$$

From Theorem 2.1, for the UMA unbiased confidence interval, g and h also satisfy $G'_{2n\kappa}(g) = G'_{2n\kappa}(h)$; from Theorem 2.3, for the log-shortest confidence interval based on the pivot $2n\bar{X}/\theta$, g and h also satisfy $gG'_{2n\kappa}(g) = hG'_{2n\kappa}(h)$.

A two-sided γ-level $100(1-\alpha)\%$ tolerance interval for the gamma distribution with known shape parameter was given by Guenther [7]. The interval is $(\bar{X}c_1, \bar{X}c_2)$, where, for large $2n\kappa$, c_1 and c_2 satisfy the two equations

$$G_{2\kappa}(hc_2/n) - G_{2\kappa}(hc_1/n) = G_{2\kappa}(gc_2/n) - G_{2\kappa}(gc_1/n) = \gamma,$$

where g and h satisfy $1 - \alpha = \Pr(g < \chi^2_{2n\kappa} < h)$. The tolerance interval is close to but not the same as the confidence interval for the γ-mode interval.

If $\kappa = 1$, the γ-mode interval for a single observation is (5) and from Theorem 2.2, a $100(1-\alpha)\%$ type II UMA confidence interval for (5) is

$$[0, -2n\bar{X} \log(1-\gamma)/G^{-1}_{2n}(\alpha)]. \tag{7}$$

The confidence interval (7) is constructed as a two-sided interval but is numerically the same as the one-sided γ-level $100(1-\alpha)\%$ tolerance interval. The two-sided γ-level $100(1-\alpha)\%$ tolerance interval obtained by Goodman & Madansky [6] by controlling both tails like Owen [13], is the same as the $100(1-\alpha)\%$ confidence interval for the inter-fractile interval, namely

$$\left[-2n\bar{X} \log\{(1+\gamma)/2\}/G^{-1}_{2n}(1-\alpha/2),\ -2n\bar{X} \log\{(1-\gamma)/2\}/G^{-1}_{2n}(\alpha/2)\right]. \tag{8}$$

We argue that the confidence interval for the mode interval is the more meaningful interval and question the value of the standard two-sided tolerance interval (8) which omits the highest density region. Prediction intervals can be constructed from the normalized spacings between order statistics but these do not relate in a simple way to the estimated γ mode interval.

When the shape parameter κ is also unknown, exact intervals are not available. However, it is straightforward to use large sample approximations based on Taylor series expansions of the endpoints of the reference interval to construct approximate confidence intervals for (4).

References

[1] Carroll, R. J. and Ruppert, D. (1991). Prediction and tolerance intervals with transformation and/or weighting. *Technometrics* **33** 197–210.

[2] Chen, L.-A. and Hung, N.-H. (2006). Extending the discussion on coverage intervals and statistical coverage intervals. *Metrologia* **43** L43–L44.

[3] Chen, L.-A., Huang, J.-Y., and Chen, H.-C. (2007). Parametric coverage interval. *Metrologia* **44** L7–L9.

[4] Dybkær, R. and Solberg, H. E. (1987). International Federation of Clinical Chemistry (IFCC). Approved recommendation (1987) on the theory of reference values. Part 6. Presentation of observed values related to reference values. *Clinica Chimica Acta* **170** 33–42; *J. Clinical Chemistry and Clinical Biochemistry* **25** 657–662.

[5] Eaton, M. L, Muirhead, R. J., and Pickering, E. H. (2006). Assessing a vector of clinical observations. *J. Statist. Plan. Inf.* **136** 3383–3414.

[6] Goodman, L. A. and Madansky, A. (1962). Parameter-free and nonparametric tolerance limits: the exponential case. *Technometrics* **4** 75–96.

[7] Guenther, W. C. (1972). Tolerance intervals for univariate distributions. *Naval Research Logistics Quarterly* **19** 310–333.

[8] Guttman, I. (1970). *Statistical tolerance regions: Classical and Bayesian*. Griffin, London.

[9] Holst, E. and Christensen, J. M. (1992). Intervals for the description of the biological level of a trace element in a reference population. *The Statistician* **41** 233–242.

[10] Howe, W. G. (1969). Two-sided tolerance limits for normal populations - some improvements. *J. Amer. Statist. Assoc.* **64** 610–620.

[11] Krishnamoorthy, K. and Mathew, T. (2009). *Statistical Tolerance Regions: Theory, Application and Computation*. Wiley, New York.

[12] NCCLS C28-A2 (1995). *How to define and determine reference intervals in the clinical laboratory: Approved Guideline*. Second edition. Villanova, PA, National Committee for Clinical Laboratory Standards.

[13] Owen, D. B. (1964). Control of percentage in both tails of the normal distribution. *Technometrics* **6** 377–387.

[14] Patel, J. K. (1986). Tolerance limits – A review. *Communications in Statistics – Theory and Methods* **15** 2719–2762.

[15] Paulson, E. (1943). A note on tolerance limits. *Ann. Math. Statist.* **14** 90–93.

[16] Petitclerc, C. and Solberg, H. E. (1987). International Federation of Clinical Chemistry (IFCC). Approved recommendation (1987) on the theory of reference values. Part 2. Selection of individuals for the production of reference values. *Clinica Chimica Acta* **170** 1–12; *J. Clinical Chemistry and Clinical Biochemistry* **25** 639–644.

[17] Poulsen, O. M., Holst, E., and Christensen, J. M. (1997). Calculation and application of coverage intervals for biological reference values (technical report). *Pure and Applied Chemistry* **69** 1601–1611.

[18] Proschan, F. (1953). Confidence and tolerance intervals for the normal distribution. *J. Amer. Statist. Assoc.* **48** 550–564.

[19] Solberg, H. E. (1987). International Federation of Clinical Chemistry (IFCC). Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *Annales de Biologie Clinique* **45** 237–241; *Clinica Chimica Acta* **165** 111–118; *J. Clinical Chemistry and Clinical Biochemistry* **25** 337–42.

[20] Solberg, H. E. 1987. International Federation of Clinical Chemistry (IFCC). Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. *Clinica Chimica Acta* **170** 13–32; *J. Clinical Chemistry and Clinical Biochemistry* **25** 645–656.

[21] Trost, D. C. (2006). Multivariate probability-based detection of drug-induced hepatic signals. *Toxicol. Review* **25** 37–54.

[22] Wald, A. (1943). An extension of Wilks' method for setting tolerance limits. *Ann. Math. Statist.* **14** 45–55.

[23] Wald, A. and Wolfowitz, J. (1946). Tolerance limits for a normal distribution. *Ann. Math. Statist.* **17** 208–218.

[24] Willink, R. (2004). Coverage intervals and statistical coverage intervals. *Metrologia* **41** L5–L6.

[25] Wilks, S. S. (1941). Determination of sample sizes for setting tolerance limits. *Ann. Math. Statist.* **12** 91–96.