Statistical Science
2022, Vol. 37, No. 4, 628–629
https://doi.org/10.1214/22-STS869
Main article: https://doi.org/10.1214/20-STS811

Rejoinder: Confidence as Likelihood

Yudi Pawitan and Youngjo Lee

We thank M. Lavine and J. Bjørnstad (LB) for their comments. Following their notation, we start with a point of agreement that the extended likelihood of (θ, ψ) is $L_e(\theta, \psi; y) \equiv p_\theta(y, \psi)$. Many properties of Fisher's classical likelihood—most importantly, the invariance of the likelihood ratio with respect to transformation of ψ—do not apply immediately to the extended likelihood. With its new characteristics and much wider reach, $L_e(\theta, \psi; y)$ merits the qualifier "extended." B actually called it "generalized" or "general" likelihood. We further agree with their equation (6) being the extended likelihood of (θ, u), where u is the true status of the confidence interval, and with the subsequent likelihood ratios for the normal example. However, it is important to note that, without special justification, one cannot meaningfully take ratios of extended likelihoods. Allowing such ratios implies that one can find the maximum likelihood estimate by a joint maximization over the fixed and random parameters. But, due to the lack of invariance, performing such a procedure on an arbitrary scale of the random parameter can easily lead to contradictions; see Chapter 4 in Lee, Nelder and Pawitan (2017). This shows that, while the likelihood principle guarantees the sound properties of the extended likelihood in capturing evidence in the data, it does not by itself tell us what to do.

Part of LB's objection could be due to what they perceive as lack of full information. In their summary, they wrote "(4) does not account for the full model." If u is the true status of just a single confidence interval, corresponding to *a single value of* α, then there is of course a loss of information on θ. It is a major simplification, which is nonetheless commonly done, to represent the uncertainty of θ with a single confidence interval. But, in the first instance, the issue is only a recognition that $L_e(u; y)$ is an extended likelihood. If the confidence interval is based on a sufficient estimate, then the full information can be found in the full confidence density, which represents confidence intervals for *all values* of α.

Even within the classical likelihood framework, with both θ and ψ fixed parameters, one can still meaningfully set up a marginal likelihood $L(\psi; y) \equiv L(\psi; h(y))$ based on a specially chosen statistic $h(y)$ whose distribution is free of θ. The marginal likelihood is needed when ψ is the parameter of interest. There is a potential loss of information from the use of marginal likelihood, on the parameter ψ itself and certainly on θ, but there is no controversy in calling the marginal $L(\psi; y)$ a likelihood. Thus, in the same vein, the extended likelihood $L_e(u; y)$ is a marginal likelihood from $L_e(\theta, u; y)$, based on the statistic $U(y)$ whose distribution happens to be free of θ. In LB's normal example, $L_e(u = 1; y) = 0.95$ is a confidence statement that is free of θ.

In Section 3, LB state two requirements for a likelihood function, which they claim are not satisfied by (4). The first requirement is a model specification, which is naturally necessary to give meaning to the parameter and to compute any kind of likelihood. But the second requirement on "inferential aim" is qualitative and vague. For example, can we not say that our inferential aim is just to assess whether the observed confidence interval covers the true parameter, but not in the exact value of θ? LB also state that "it's experiments, not random variables, that induce likelihood functions." We find this statement self-contradictory. All random variables trivially correspond to the results of random experiments.

Consider the extended likelihood of ψ when there is no unknown fixed parameter θ. Thus, $L_e(\psi; y) \equiv p(y, \psi) = p(y|\psi)p(\psi)$. Here, the *classical* likelihood of ψ is $L(\psi; y) = p(y|\psi)$. What is the classical likelihood $L(\psi)$ *before* observing y? It should be free of ψ, which can then be set to $L(\psi) \equiv 1$, so that on observing y the classical likelihood is then a correct update of the constant likelihood. So, before observing y we have $L_e(\psi) = L(\psi)p(\psi) = p(\psi)$. Hence, if we follow the logic of extended likelihood, an *unobserved random variable* does induce an extended likelihood. Avoiding the notational and conceptual distinctions between the classical $L(\cdot)$ and the extended $L_e(\cdot)$ will only lead to confusion.

LB then present the standard example of the binomial versus the negative binomial data, where the classical likelihoods are proportional, but the P-values or the confidence distributions are different. They further write (in their notation) "$L_1(v) = L_2(v) \propto 1$." But, in this discrete case, this is incorrect. By definition, $v_1 \equiv C_1(t, \theta)$ and $v_2 \equiv C_2(t, \theta)$. Having $C_1 \neq C_2$ means $L_{e1}(v; y) \neq L_{e2}(v; y)$; the random variables V_1 and V_2 do not even

Yudi Pawitan is Professor, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (e-mail: yudi.pawitan@ki.se). Youngjo Lee is Professor Emeritus, Department Statistics, Seoul National University, Seoul, Korea (e-mail: youngjo@snu.ac.kr).