**Manuscript ID:** AOS1807-038

**Manuscript Title:** Integrated Likelihood Ratio Test

This paper introduces integrated likelihood ratio test. It aims to develop a test when likelihood function could be unbounded or have disturbing behaviors. The idea is to be a bit Bayesian and use a distribution over the parameter to smooth out the ill-behaving region. For example, The paper suggests using posterior distribution as this smoothing distribution. The paper also considers an extension to using Lq likelihood and, correspondingly, fractional posterior.

The main theoretical result is around the Wilks phenomenon for the integrated likelihood ratio test and its extensions. The condition is weaker for the integrated likelihood ratio test than the likelihood ratio test, especially with its Lq likelihood extension. The authors study two examples within the normal mixture model, where the likelihood function could have undesirable behaviors but the integrated likelihood ratio test is less affected than the classical likelihood ratio test.

The paper is overall clear. The theoretical results are sensible. However, the motivation of the paper did not convince me. For example, the paper hinges on the unboundedness of the likelihood function, which occurs only with one single data point; however, it studies asymptotic properties. Also the paper points out that the likelihood function is hard to optimize when it is not concave. But using Lq likelihood or variational inference or MCMC does not resolve this issue.

# 1 Major Comments

1. **Unbounded likelihood.** The main motivation of integrated likelihood ratio test is the claim that the likelihood can be unbounded. If so, the paper claims that the likelihood ratio test is undefined. However, this claim did not convince me. In fig 1., the likelihood function goes to infinity when there is only one data point. However, if we consider two or more data points, this unboundedness does not occur any more. The likelihood will not go to infinity even when $\sigma^2$ becomes very small, unless the two data points coincide with each other precisely.

   Maybe we could stick with the one data point case, where the pathology of likelihood function occurs. In this case, it is hard to conceive of a reason why we want to perform likelihood ratio test with a data set that contains only one data point.

   Maybe we could trust the paper that a similar bad behavior of the likelihood function can occur when the data set is very small and clustered. However, the main result derived in the paper is in the large data limit. This seems to be at odds at the motivation.

2. **Nonconcavity of likelihood functions.** The paper also poses the dissatisfaction when the likelihood function is not concave and hard to optimize. This non-concavity could happen but the solution proposed does not obviously resolve this concern. Is the integrated likelihood ratio test guaranteed to be concave? If rigorous conditions and proofs can show the guaranteed concavity, it could be quite appealing.

3. **Conditions on the $\Pi$ function.** The integrated likelihood ratio test is equivalent to the classical likelihood ratio test when the $\Pi$ function is taken as a point mass. However, the paper requires the consistency of the $\Pi$ function eventually, that is it

converges to some point mass around the truth under a root-n rate. It seems at odds with the goal of having integrated likelihood ratio resolve issues with the classical likelihood ratio test.

4. **Lq likelihood extensions.** The paper shows that the good asymptotic properties of integrated likelihood ratio test requires fewer conditions than the classical likelihood ratio test. Classically, it is known that the fractional posterior requires weaker assumptions for posterior consistency. The corresponding Lq likelihood used is also known to enjoy robustness properties (Ferrari et al., 2010; Qin and Priebe, 2013b,a).

   However, it is unclear what is the motivation to use the Lq likelihood in the integrated likelihood ratio test here. It does make the integrated likelihood ratio test require fewer conditions for asymptotics than the classical likelihood ratio test. Does it resolve the non-concave concern? Does it induce robustness properties that is needed for the likelihood ratio test to avoid pathologies?

5. **Posterior predictive check.** Integrated likelihood ratio test aligns pretty closely with posterior predictive check (Gelman et al., 1996) when the likelihood function is the test statistic. Connections between the integrated likelihood ratio test and the posterior predictive check can be further explored.

   One possible way to strengthen the motivation is to convince the readers if the integrated likelihood ratio test can address limitations of the posterior predictive check.

   Finally, one classical criticism of the posterior predictive check is its "double-dipping". It uses the data to compute a posterior and then use the same data to do a test on the model.

   Due to the closeness connection between the integrated likelihood ratio test and the posterior predictive check, the same criticism also applies. This is not a drawback of the proposal. However, this issue could be worth some discussion in the paper.

## 2 Minor Comments

1. abstract: "frequency properties" should be "frequentist properties".
2. notation: using $\phi_n$ for test functions and $\phi$ for the normal density can be confusing.

## References


Ferrari, D., Yang, Y., et al. (2010). Maximum lq-likelihood estimation. *The Annals of Statistics*, 38(2):753–783.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica sinica*, pages 733–760.

Qin, Y. and Priebe, C. E. (2013a). Maximum l q-likelihood estimation via the expectation-maximization algorithm: A robust estimation of mixture models. *Journal of the American Statistical Association*, 108(503):914–928.

Qin, Y. and Priebe, C. E. (2013b). Robust hypothesis testing via lq-likelihood. *arXiv preprint arXiv:1310.7278*.