UCLA

Department of Statistics Papers

Title

Geometric Understanding of Likelihood Ratio Statistics

Permalink

<https://escholarship.org/uc/item/3rk4q8hv>

Authors

Fan, Jianqing

Hung, Hui-Nien

Wong, Wing-Hung

Publication Date

1998-03-11

Geometric Understanding of Likelihood Ratio Statistics*

Jianqing Fan, Hui-Nien Hung and Wing-Hung Wong

March 11, 1998

Abstract

It is well-known that twice a log-likelihood ratio statistic follows asymptotically a χ^2 -distribution. The result is usually understood and proved via Taylor's expansions of likelihood functions and by assuming asymptotic normality of maximum likelihood estimators. We contend that more fundamental insights can be obtained for likelihood ratio statistics: the Wilks type of results hold as long as likelihood contour sets are of fan-shape. The classical Wilks theorem corresponds to the situations where the likelihood contour sets are ellipsoid. This provides an insightful geometric understanding and a useful extension of the likelihood ratio theory. As a result, even if the MLEs are not asymptotically normal, the likelihood ratio statistics can still be asymptotically χ^2 -distributed. Our technical arguments are simple and can easily be understood.

Jianqing Fan and Wing-Hung Wong are professors, Department of Statistics, University of California, Los Angeles, CA 90095. Hui-Nien Hung is associate professor, Institute of Statistics, National Chiao-Tong University, Taiwan. Fan's research was partially supported by NSF grant DMS-9803200 and NSA grant 96-1-0015. Wong's work was partially supported by the Mathematical Sciences Research Institute of the Chinese University of Hong Kong and by NSF grant DMS-9703918.

1 Introduction

One of the most celebrated folk theorems in statistics is that twice the logarithm of a maximum likelihood ratio statistic is asymptotically χ^2 -distributed. This result is due to Wilks (1938) and is proved via Taylor's expansions of likelihood functions and by assuming that the maximum likelihood estimator (MLE) is asymptotically normal. See also Wald (1941), Wilks (1962) and heuristics given in popular textbooks such as Cox and Hinkley (1974), Kendall and Stuart (1979), among others. While this understanding is insightful, it has three drawbacks. First of all, the likelihood function has to be smooth enough in order to admit a Taylor's expansion. Secondly, the MLE has to be asymptotically normal and this itself relies on Taylor's expansions and the central limit theorem. Thirdly, assumptions on the independence of observations are typically made. Rigorous technical proofs of the first two steps above are by no mean simple. This is probably why rigorous statements and heuristic proofs are suppressed in many popular graduate textbooks. See for example page 229 of Bickel and Doksum (1977), page 486 of Lehmann (1986) and page 381 of Casella and Berger (1990).

We contend that much simpler fundamental insight to the Wilks theorem is available: if the contour sets of a likelihood function around an MLE are of fan shape, then the Wilks type of results hold. The classical Wilks' theorem corresponds to the situations where the contour sets are ellipsoid. In general, the asymptotic normality of the MLE is not required, neither does the asymptotic distribution of the MLE have to exist. One can easily construct an example where the MLE is not asymptotically normal, but Wilks' type of results hold. See Examples 1 & 2 in Section 3. An additional benefit is that our technical arguments are simple and can be understood without much probability background.

We begin with the simplest case, where the null hypothesis consists of only one point:

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0, \quad (1)$$

with θ being a vector of unknown parameters in an Euclidean space. Let $\mathbf{X} \sim f(\mathbf{x}; \theta)$ be a random vector from which a sample of data is drawn and let $\ell(\theta; \mathbf{x}) = \log f(\mathbf{x}; \theta)$ be the log-likelihood function. Denote by $\hat{\theta}$ the maximum likelihood estimator. Set

$$W(\theta_0, \mathbf{X}) = \ell(\hat{\theta}, \mathbf{X}) - \ell(\theta_0, \mathbf{X}), \quad (2)$$

which is the log-likelihood ratio statistic for testing the hypotheses (1). Our idea is simple. It uses some simple tools of Bayesian statistics. Let us assign a continuous prior density $\pi(\cdot)$ for the parameter θ . One can then easily show that the posterior distribution of $W(\theta, \mathbf{X})$

given $\mathbf{X}$ is asymptotically a gamma distribution, independent of the prior distribution. This implies that the marginal distribution of W is also asymptotically a gamma distribution. Since the result holds for every continuous prior distribution, it must follow that the distribution of W given θ is asymptotically a gamma distribution. A similar kind of arguments were used before by, for example, Bickel and Ghosh (1990) and Dawid (1991). In particular, when the shape parameter of the gamma distribution is a half of an integer, the random variable $2W$ follows asymptotically a χ^2 -distribution. In other words, the Wilks theorem is a specific case of our generalized results.

The above arguments can readily be extended to the cases where the null hypothesis contains nuisance parameters. The key to our success relies on the regenerating property of gamma distributions.

The paper is organized as follows. Section 2 derives the posterior distribution of likelihood ratio statistics. In Section 3, we obtain the sampling distribution of the likelihood ratio statistic from the posterior distribution. The arguments are extended in Section 4 to the cases where the null hypothesis contains nuisance parameters.

2 Posterior distribution

Assume that $\theta \in R^p$ has a prior density $\pi(\theta)$. Then, the posterior density of θ given $\mathbf{X} = \mathbf{x}$ is given by

$$\frac{\exp\{\ell(\theta, \mathbf{x})\}\pi(\theta)}{\int_{\Theta} \exp\{\ell(\theta, \mathbf{x})\}\pi(\theta)d\theta} = \exp\{-W(\theta, \mathbf{x})\}\pi(\theta)/g_n(\mathbf{x}),$$

where

$$g_n(\mathbf{x}) = \int_{\Theta} \exp\{-W(\theta, \mathbf{x})\}\pi(\theta)d\theta.$$

Let $S_w = \{\theta \in \Theta : W(\theta, \mathbf{x}) = w\}$ be a likelihood contour set. Our aim is to show that the posterior distribution of $W(\theta, \mathbf{X})$ given $\mathbf{X} = \mathbf{x}$ is asymptotically Gamma-distributed if the likelihood contour set can be approximated as

$$S_w \approx \hat{\theta} + a_n w^r S \tag{3}$$

for a sequence of $a_n \rightarrow 0$, $r > 0$ and a surface S in R^p . This is an extension of classical conditions on the Wilks theorem where the likelihood contour set is approximated by an ellipsoid.

The condition (3) is not rigorous. To formally state the result, we assume that there exists a function h in R^p such that

$$h(t\theta) = t^{1/r}h(\theta), \forall t > 0, \quad \text{and } S = \{\theta : h(\theta) = 1\}. \tag{4}$$

This and (3) imply heuristically that

$$W(\theta, \mathbf{x}) \approx h(a_n^{-1}(\theta - \hat{\theta})) = a_n^{-1/r} h(\theta - \hat{\theta}).$$

Let $W_n^*(\theta, \mathbf{x}) = a_n^{\frac{1}{r}} W(\theta, \mathbf{x})$. The formal conditions can be expressed as follows:

(A1) There exist a function $m(\cdot)$ and a constant N such that when $n > N$,

$$\inf\{W_n^*(\theta, \mathbf{x}) : \|\theta - \hat{\theta}\| > \delta\} \geq m(\delta) > 0 \quad \text{for all } \delta > 0.$$

Moreover the maximum likelihood estimator $\hat{\theta}$ is a stochastically bounded sequence.

(A2) There exists a function $h(\cdot)$ such that the likelihood contour sets are of fan shape in the following sense

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{\|\hat{\theta} - \theta\| \leq \delta} \left| \frac{W_n^*(\theta, \mathbf{x}) - h(\theta - \hat{\theta})}{h(\theta - \hat{\theta})} \right| = 0.$$

(A3) The function h satisfies $h(t\theta) = t^{1/r} h(\theta)$ and $\inf\{h(\theta) : \|\theta\| = 1\} > 0$.

Condition (A2) is a rigorous condition of (3). The latter can be intuitively understood as that the unit volume around surface S_w is proportional to $w^{(rp-1)}$, namely

$$\lim_{\Delta \rightarrow 0} \Delta^{-1} V\{\theta : W(\theta, \mathbf{X}) \in w \pm \Delta/2\} \text{ is proportional to } w^{(rp-1)}. \quad (5)$$

In other words, as long as the contour sets $\{S_w\}$ are rigid for all $w > 0$, the Wilks type of results hold.

Theorem 1 *Suppose $\pi(\theta)$ is a bounded positive continuous function. Then, under the regularity conditions (A1)-(A3), the conditional distribution of the log-likelihood ratio statistic W given $\mathbf{X} = \mathbf{x}$ has an asymptotic Gamma-distribution with shape parameter rp and scale parameter one, namely,*

$$\mathcal{L}(W|\mathbf{X} = \mathbf{x}) \rightarrow \text{Gamma}(rp).$$

The proof of this theorem uses the following lemma whose proof is similar to that of the Laplace Approximation (Tierney and Kadane (1986)). The proof is elementary, but somewhat tedious.

Lemma 1 *Under Conditions (A1)-(A3), we have*

$$g_n(\mathbf{x}) = \pi(\hat{\theta}) a_n^p V(\mathcal{O}) \Gamma(rp + 1) (1 + o(1)),$$

where $V(\mathcal{O})$ is the volume of the set $\mathcal{O} = [0, 1] \times S$ with S given by (4).

Proof. By Condition (A1) when $n > N$,

$$\int_{\|\theta - \hat{\theta}\| > \delta} \exp\{-W(\theta, \mathbf{x})\} \pi(\theta) d\theta \leq \exp\{-a_n^{-1/r} m(\delta)\}.$$

For any given $\varepsilon > 0$, when δ is small enough, we deduce from Conditions (A2) and (A3) that

$$\begin{aligned} & \int_{\|\theta - \hat{\theta}\| \leq \delta} \exp\{-W(\theta, \mathbf{x})\} \pi(\theta) d\theta \\ & \geq \int_{\|\theta - \hat{\theta}\| \leq \delta} \exp\{-(1 + \varepsilon) a_n^{-1/r} h(\theta - \hat{\theta})\} \pi(\theta) d\theta \\ & \geq \int_{\theta \in R^p} \exp\{-(1 + \varepsilon) a_n^{-1/r} h(\theta - \hat{\theta})\} \pi(\theta) d\theta - \exp\{-(1 + \varepsilon) a_n^{-1/r} \delta^{1/r} c\}, \end{aligned}$$

where $c = \inf\{h(\eta) : \|\eta\| = 1\}$. By a change of variables, the first term is given by

$$\begin{aligned} & a_n^p (1 + \varepsilon)^{-rp} \int_{\theta \in R^p} \exp\{-h(\theta)\} \pi(\hat{\theta} + a_n (1 + \varepsilon)^{-r} \theta) d\theta \\ & \geq \pi(\hat{\theta}) (1 - \varepsilon) a_n^p (1 + \varepsilon)^{-rp} \int_{\theta \in R^p} \exp\{-h(\theta)\} d\theta \\ & = \pi(\hat{\theta}) (1 - \varepsilon) a_n^p (1 + \varepsilon)^{-rp} \int_0^\infty \exp\{-t\} d\{V(t^r \mathcal{O})\} \\ & = \pi(\hat{\theta}) (1 - \varepsilon) a_n^p (1 + \varepsilon)^{-rp} V(\mathcal{O}) r p \int_0^\infty e^{-t} t^{rp-1} dt, \end{aligned}$$

where the inequality is obtained by invoking the dominated convergence theorem. By letting $n \rightarrow \infty$, then $\delta \rightarrow 0$ and $\varepsilon \rightarrow 0$, we have

$$\liminf_{n \rightarrow \infty} a_n^{-p} \pi(\hat{\theta})^{-1} g_n(\mathbf{x}) \geq V(\mathcal{O}) \Gamma(rp + 1).$$

In a similar vein, we can easily show that

$$\limsup_{n \rightarrow \infty} a_n^{-p} \pi(\hat{\theta})^{-1} g_n(\mathbf{x}) \leq V(\mathcal{O}) \Gamma(rp + 1).$$

This concludes the proof of Lemma 1. $\square$

Proof of Theorem 1. Note that

$$P(W < w | \mathbf{X} = \mathbf{x}) = g_n(x)^{-1} \int_{\{\theta : W(\theta, \mathbf{x}) < w\}} e^{-W(\theta, \mathbf{x})} \pi(\theta) d\theta.$$

We need only to evaluate the integral term in the above expression. The arguments follow similar lines to those in the proof of Lemma 1. More precisely, by Condition (A2), we have

$$\begin{aligned} & \int_{W(\theta, \mathbf{x}) < w} \exp\{-W(\theta, \mathbf{x})\} \pi(\theta) d\theta \leq \int_{\|\hat{\theta} - \theta\| > \delta} \exp\{-W(\theta, \mathbf{x})\} \pi(\theta) d\theta \\ & + \int_{h(\theta - \hat{\theta}) < w(1 + \varepsilon) a_n^{1/r}} \exp\{-a_n^{-1/r} (1 + \varepsilon) h(\theta - \hat{\theta})\} \pi(\theta) d\theta \end{aligned}$$

By Condition (A1), the first integral is bounded by $\exp\{-a_n^{-1/r}m(\delta)\}$. By a change of variable, the second integral equals to

$$\begin{aligned} & a_n^p(1-\epsilon)^{-rp} \int_{h(\theta) < w(1+\epsilon)(1-\epsilon)} \exp\{-h(\theta)\} \pi(\hat{\theta} + a_n^{1/r}(1-\epsilon)\theta) d\theta \\ & \leq \pi(\hat{\theta})(1+\epsilon)a_n^p(1-\epsilon)^{-rp} \int_{t=0}^{w(1-\epsilon^2)} e^{-t} d\{V(t^r \mathcal{O})\} \\ & = \pi(\hat{\theta})(1+\epsilon)a_n^p(1-\epsilon)^{-rp} rp V(\mathcal{O}) \int_{t=0}^{w(1-\epsilon^2)} e^{-t} t^{rp-1} dt. \end{aligned}$$

Using the same method, we have

$$\begin{aligned} & \int_{W(\theta, \mathbf{x}) < w} \exp\{-W(\theta, \mathbf{x})\} \pi(\theta) d\theta \\ & \geq \pi(\hat{\theta})(1-\epsilon)a_n^p(1+\epsilon)^{-rp} rp \int_{t=0}^{w(1-\epsilon^2)} e^{-t} t^{rp-1} dt. \end{aligned}$$

Thus, we have

$$\lim_{n \rightarrow \infty} a_n^{-p} \pi(\hat{\theta}) \int_{W(\theta, \mathbf{x}) < w} \exp\{-W(\theta, \mathbf{x})\} \pi(\theta) d\theta = V(\mathcal{O}) \int_{t=0}^w e^{-t} t^{rp-1} dt.$$

This together with Lemma 1 prove Theorem 1. $\square$

By noting that if $Y \sim \text{Gamma}(r)$, then $2Y \sim \chi_{2r}^2$, we have

Corollary 2.1 *Under the regularity conditions (A1)-(A3), we have*

$$\mathcal{L}(2W|\mathbf{X} = \mathbf{x}) \rightarrow \chi_{2r}^2,$$

provided that $2r$ is an integer.

3 Sampling Distribution

In this section, we derive the asymptotic distributions of likelihood ratio statistics from the frequentist point of view. Let Θ^0 be a bound open set in an Euclidean space. To stress its dependence on n , we use $P_n(W \leq w|\theta)$ to denote the sampling distribution.

Theorem 2 *If $P_n(W \leq w|\theta)$ is equicontinuous in $\theta \in \Theta^0$ for every w , then under Conditions (A1) - (A3)*

$$\mathcal{L}(W|\theta_0) \rightarrow \text{Gamma}(rp), \quad \text{for all } \theta_0 \in \Theta^0.$$

Proof. Let $\{\pi_m(\theta)\}$ be a sequence of bounded continuous prior distributions of θ which shrink to the point θ_0 . Denote the marginal distribution of W by $a_{n,m}(w) = \int P_n(W <$

$w|\theta)\pi_m(\theta)d\theta$. By Theorem 1, we have $\mathcal{L}_m(W|\mathbf{X}) \rightarrow \text{Gamma}(r)$ for all m , where $\mathcal{L}_m(W|\mathbf{X})$ is the conditional distribution of W under the prior π_m . By using the dominated convergence theorem,

$$\lim_{n \rightarrow \infty} a_{n,m}(w) = \Gamma(w, rp), \quad (6)$$

where $\Gamma(w, rp)$ is the cumulative distribution of $\text{Gamma}(rp)$. By the equicontinuity assumption, we have

$$\lim_{m \rightarrow \infty} a_{n,m}(w) = P_n(W < w|\theta_0) \quad \text{and the convergence is uniform in } n. \quad (7)$$

It follows from (6) and (7) that

$$\lim_{n \rightarrow \infty} P_n(W < w|\theta_0) = \lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} a_{n,m}(w) = \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} a_{n,m}(w) = \Gamma(w, rp).$$

This completes the proof. $\square$

Remark 3.1 *In many cases (see Examples 1 – 3 below), the sampling distribution of W is independent of θ and hence it is equicontinuous. This condition is used to show the existence of the limit distribution of W and can be replaced by the assumption that the limiting distribution of $a_n^{-1}(\hat{\theta} - \theta)$ exists. To see this, under the latter assumption, by Conditions (A2) and (A3) the likelihood ratio $W = h(a_n^{-1}(\hat{\theta} - \theta))(1 + o_P(1))$ converges weakly. Let $G(w, \theta)$ be the limit of $P_n(W < w|\theta)$. Then by (6), $EG(w, \theta) = \Gamma(w, rp)$ for every continuous prior π . Hence $G(w, \theta) = \Gamma(w, rp)$ for almost all $\theta \in \Theta^0$.*

Theorems 1 and 2 reveal that the Wilks type of results hold as long as the likelihood contour is of fan shape. It admits good geometric interpretation. The classical Wilks theorem is usually derived under the conditions similar to Cramér's ones. See for example Conditions (C1) – (C5) on page 102 of Le Cam and Yang (1990). These conditions imply that the log-likelihood function can locally be approximated by a quadratic function (See Le Cam and Yang 1990) and the likelihood contour is ellipsoid:

$$S_w \approx \hat{\theta} + (2w/n)^{1/2} S,$$

where $S = \{\theta : \theta^T \Sigma \theta = 1\}$ and Σ is the Fisher information at the true underlying parameter. Hence, this is a specific case of our results.

We contend that the Wilks type of results hold for a much larger class of likelihood contours. The shapes of the likelihood contours do not have to be ellipsoid and the radii do not have to be proportional to $w^{1/2}$. Hence, the shape parameter of the gamma distribution

does not have to be $p/2$, where p is number of parameters and the MLE does not have to be asymptotically normal.

Example 1. Suppose that we have a random sample of size n from the exponential distribution model

$$\mathbf{X} = \theta + \varepsilon,$$

where $\mathbf{X}$, θ and ε are p -dimensional vector and the components of ε are independent having the standard exponential distribution. Then, the MLE is $\hat{\theta} = \min(\mathbf{X}_1, \dots, \mathbf{X}_n)$, where the operator “min” is applied componentwise. Thus, $n(\hat{\theta} - \theta) \sim \varepsilon$, which is not asymptotically normal. The likelihood ratio

$$W(\theta, \mathbf{X}) = n \sum (\hat{\theta} - \theta) \sim \text{Gamma}(p),$$

where the operator $\sum$ is applied to components of the vector. Hence,

$$2W(\theta, \mathbf{X}) \sim \chi_{2p}^2.$$

Note that the degree of freedom is $2p$ instead of p . The likelihood contour in this case is

$$S_w = \{\theta : n \sum (\hat{\theta} - \theta) = w, \hat{\theta} \geq \theta\} = \hat{\theta} + (w/n)S,$$

where $S = \{\theta : \sum \theta_i = -1, \theta_i \leq 0\}$ is a hyper-triangle. Conditions (A1) - (A3) are satisfied with $h(\theta) = -\sum_{i=1}^p \theta_i I(\theta_i \leq 0)$ and $r = 1$, $a_n = \frac{1}{n}$. $\square$

The above example provides stark evidence that Wilks' type of results continue to hold even though the MLE is not asymptotically normal. Such a kind of example is not uncommon. We provide an additional one.

Example 2. Suppose that we have a random sample of size n from the uniform distribution on the p -dimensional hyper-rectangle $[0, \theta]$. Then, the MLE is $\hat{\theta} = \max(\mathbf{X}_1, \dots, \mathbf{X}_n)$, where the operator “max” is applied componentwise. The likelihood ratio

$$W(\theta, \mathbf{X}) = n \sum_{i=1}^p \log(\theta_i / \hat{\theta}_i) I(\theta_i \geq \hat{\theta}_i) \sim \text{Gamma}(p).$$

Again, $\hat{\theta}$ is not asymptotically normal and the degree of freedom for $2W$ is $2p$ instead of p . The likelihood contour in this example is approximately a hyper-triangle:

$$S_w \approx \hat{\theta} + (w/n)S,$$

where $S = \{\theta : \sum_{i=1}^p \frac{\theta_i}{\theta_{i0}} = 1, \theta_i \geq 0\}$ with $(\theta_{10}, \dots, \theta_{p0})^T$ being the true underlying parameters. Conditions (A1) - (A3) are satisfied with $h(\theta) = \sum_{i=1}^p \frac{\theta_i}{\theta_{i0}} I(\theta_i \geq 0)$ and $r = 1$, $a_n = \frac{1}{n}$ $\square$

The Cramér condition for the Wilks theorem depends critically on parameterization. This drawback is attenuated in our formulation.

Example 3. Suppose that we parameterize a normal population as $N(\theta^3, I_p)$, where I_p is a $p \times p$ identity matrix. Based on a random sample of size n , the MLE is given by $\hat{\theta} = \bar{X}^{1/3}$ where $\bar{X}$ is the sample mean. Let Z be the p -dimensional standard normal random vector. It is clear that when the true parameter $\theta = \mathbf{0}$, $n^{1/6} \hat{\theta} \sim Z^{1/3}$, which is not asymptotically normal. Hence, the Cramér's conditions do not hold under this parameterization. In this case, the likelihood function can be approximated as

$$W(\theta, \mathbf{X}) = \frac{1}{2} n \|\bar{X} - \theta^3\|^2 = \frac{1}{2} n \sum_{i=1}^p (\bar{X}_i^{1/3} - \theta_i)^2 \{9\theta_{i0}^{4/3} + o_P(1)\}$$

for θ in a neighborhood of $\hat{\theta}$, where $\theta_0 = (\theta_{10}, \dots, \theta_{p0})^T$ is the true underlying parameters. The likelihood contour is of fan shape

$$S_w = \{\theta : n \|\bar{X} - \theta^3\|^2 = 2w\} \approx \bar{X}^{1/3} + (2w/n)^{1/2} S$$

with S is an ellipse given by $S = \{\theta : \sum_{i=1}^p \theta_i^2 \theta_{i0}^{4/3} = 1/9\}$. Thus, Conditions (A1) - (A3) hold with $h(\theta) = \frac{9}{2} \sum_{i=1}^p \theta_i^2 \theta_{i0}^{4/3}$, $r = 1/2$ and $a_n = n^{-1/2}$. $\square$

Another advantage of our new results is that they can accommodate the situations where different parameters may have different rates of convergence. We elaborate this further in the following remark.

Remark 3.2 *In some cases, different components of θ can have different asymptotic behavior. For example, suppose that $\theta = (\theta_1, \theta_2)$ and X and Y are two independent random variables having distributions $\text{Uniform}[0, \theta_1]$ and $N(\theta_2, 1)$, respectively. To apply Theorems 1 and 2 to this kind of problems, we need to modify conditions (A1) - (A3) as follows.*

(A1') *The function $W_n^*(\theta, \mathbf{x}) = a_n W(\theta, \mathbf{x})$ satisfies*

$$\inf\{W_n^*(\theta, \mathbf{x}) : \|\theta - \hat{\theta}\| > \delta\} \geq m(\delta) > 0 \text{ for all } \delta > 0.$$

(A2') *The likelihood function can be approximated by*

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{\|\hat{\theta} - \theta\| < \delta} \left| \frac{W_n^*(\theta, \mathbf{x}) - h(\theta - \hat{\theta})}{h(\theta - \hat{\theta})} \right| = 0.$$

(A3') The function h satisfies $h(t^{r_1}\theta_1, t^{r_2}\theta_2) = th(\theta)$, where $\theta = (\theta_1, \theta_2)$ with $\theta_1 \in R^{p_1}$ and $\theta_2 \in R^{p_2}$. Further, $\inf\{h(\theta) : \|\theta\| = 1\} > 0$.

Then we only need to replace the rp in Theorem 1 by $r_1p_1 + r_2p_2$. The unit volume in (5) is now proportional to $r_1p_1 + r_2p_2 - 1$ in this case.

4 Extension to the cases with nuisance parameters

We now consider the situations with composite null hypothesis. Partition the parameter vector ξ into two parts $\xi = (\theta^T, \lambda^T)^T$, where $\theta \in R^p$ and $\lambda \in R^q$. Under the null hypothesis that $\theta = \theta_0$, the likelihood ratio test statistic is

$$W_1(\theta_0, \mathbf{X}) = \ell(\hat{\theta}, \hat{\lambda}, \mathbf{X}) - \ell(\theta_0, \hat{\lambda}_{\theta_0}, \mathbf{X}),$$

where $\hat{\lambda}_{\theta_0}$ is the MLE under the null hypothesis that $\theta = \theta_0$.

Decompose the likelihood ratio $W(\theta, \lambda, \mathbf{X}) = \ell(\hat{\theta}, \hat{\lambda}, \mathbf{X}) - \ell(\theta, \lambda, \mathbf{X})$ as

$$W(\theta, \lambda, \mathbf{X}) = W_1(\theta, \mathbf{X}) + W_2(\theta, \lambda, \mathbf{X}), \quad (8)$$

where

$$W_2(\theta, \lambda, \mathbf{X}) = \ell(\theta, \hat{\lambda}_\theta, \mathbf{X}) - \ell(\theta, \lambda, \mathbf{X}).$$

As in Section 2, we first consider the posterior distributions of $W(\theta, \lambda, \mathbf{X})$ and $W_2(\theta, \lambda, \mathbf{X})$. The regularity conditions are similar to (A1)-(A3), which are stated as follows.

(B1) The likelihood function $W(\theta, \lambda, \mathbf{X})$ satisfies Conditions (A1)-(A3).

(B2) For each given θ in a bounded open set Θ^0 , the likelihood function $W_2(\theta, \lambda, \mathbf{X})$ satisfies Conditions (A1) - (A3).

Conditions (B1) and (B2) admit similar geometric interpretation as those given in Section 2. In particular, the likelihood contour sets are required to have fan shape. Namely, these regularity conditions can be understood as

$$\lim_{\Delta \rightarrow 0} \Delta^{-1} V\{(\theta, \lambda) : W(\theta, \lambda, \mathbf{X}) \in w \pm \Delta/2\} \text{ is proportional to } w^{r(p+q)-1} \text{ and}$$

$$\lim_{\Delta \rightarrow 0} \Delta^{-1} V\{\lambda : W_2(\theta, \lambda, \mathbf{X}) \in w \pm \Delta/2\} \text{ is proportional to } w^{rq-1}.$$

Theorem 3 Suppose that $\pi(\theta, \lambda)$ is a bounded positive continuous prior on a bounded open set $\Theta^0 \times \Lambda^0$. Then, the posterior distribution of the likelihood ratio $W_1(\theta, \mathbf{X})$ has an asymptotic Gamma distribution:

$$\mathcal{L}\{W_1(\theta, \mathbf{X}) | \mathbf{X} = \mathbf{x}\} \rightarrow \text{Gamma}(rp).$$

Proof. By Theorem 1, we have

$$\mathcal{L}\{W(\theta, \lambda, \mathbf{X})|\mathbf{X}\} \rightarrow \text{Gamma}(r(p+q))$$

and

$$\mathcal{L}\{W_2(\theta, \lambda, \mathbf{X})|\mathbf{X}, \theta\} \rightarrow \text{Gamma}(rq).$$

Since r is independent of θ , conditioning on $\mathbf{X}$, $W_2(\theta, \lambda, \mathbf{X})$ is asymptotically independent of θ and hence independent of $W_1(\theta, \mathbf{X})$. It follows that the characteristic functions satisfy

$$\begin{aligned} E\{\exp(itW)|\mathbf{X}\} &= E\{\exp(itW_1)E\{\exp(itW_2)|\mathbf{X}, \theta\}|\mathbf{X}\} \\ &= E\{\exp(itW_1)|\mathbf{X}\}\phi(t, rq) + o(1), \end{aligned}$$

where $\phi(t, rq)$ is the characteristic function of $\text{Gamma}(rq)$. Thus,

$$E\{\exp(itW_1)|\mathbf{X}\} = \phi(t, r(p+q) - rq) + o(1).$$

This completes the proof. $\square$

From the posterior distribution, we can similarly obtain the sampling distribution.

Theorem 4 *If $\mathcal{L}\{W_1(\theta, \mathbf{X})|\theta, \lambda\}$ is equicontinuous in $(\theta, \lambda) \in \Theta^0 \times \Lambda^0$, an open set in the Euclidean space, or if the limiting distributions of the MLEs $(\hat{\theta}^T, \hat{\lambda}^T)^T$ and $\hat{\lambda}_\theta$ exist upon suitable normalization, then the likelihood ratio statistic has an asymptotic Gamma-distribution:*

$$\mathcal{L}\{W_1(\theta, \mathbf{X})|\theta, \lambda\} \rightarrow \text{Gamma}(rp)$$

for almost all $(\theta, \lambda) \in \Theta^0 \times \Lambda^0$.

Note that the situations that are similar to Remark 3.2 can also accommodate in our Theorem 4. For simplicity, we omit the details.

Reference

- Bickel, P.J. and Doksum, K.A. (1977). *Mathematical Statistics: Basic ideas and selected topics*. Prentice Hall, Englewood Cliffs.
- Bickel, P.J. and Ghosh, J.K. (1990). A decomposition for the likelihood ratio statistic and the Bartlett correction — A Bayesian argument, *The Annals of Statistics*, **18**, 1070-1090.

- Casella, G. and Berger, R.L. (1990), *Statistical Inference*, Duxbury, Belmont.
- Cox, D.R. and Hinkley, D.V. (1974). *Theoretical Statistics*, Chapman and Hall, London.
- David, A.P. (1991). Fisherian inference in likelihood and frequential frames of reference (with discussion), *Journal of Royal Statistics Society B*, **53**, 79-109.
- Kendall, M. and Stuart, A. (1979). *The advanced Theory of Statistics, Volume II: Inference and Relationship*, 4th edition. Macmillan, New York.
- Le Cam, L. and Yang, G.L. (1990). *Asymptotic in Statistics: Some Basic Concepts*. Springer-Verlag, New York.
- Lehmann, E. L. (1986). *Testing Statistical Hypotheses*, 2nd edition. Wiley, New York.
- Tierney, L. and Kadane, J.B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.*, **81**, 82-86.
- Wald, A. (1941). Asymptotically most powerful tests of statistical hypotheses. *Ann. Math. Statist.*, **12**, 1 – 19.
- Wilks, S.S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Math. Stat.*, **9**, 60-62.
- Wilks, S.S. (1962). *Mathematical Statistics*. Wiley, New York.