where $\mathcal{P}_0$ and $\mathcal{P}_1$ are disjoint subsets of $\mathcal{P}$. If

$$\lim_{N\to\infty} \sup \alpha_{T_N}(P) \le \alpha \tag{2.59}$$

for any $P \in \mathcal{P}_0$, then α is called an asymptotic significance level *of T_N.*

Definition 2.7. *Consider sample $\mathbf{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N)$ from population $P \in \mathcal{P}$. Let θ be a parameter vector for P, and let $C(\mathbf{X})$ be a confidence set for $\boldsymbol{\theta}$. If $\liminf_{N\to\infty} P[\boldsymbol{\theta} \in C(\mathbf{X})] \ge 1 - \alpha$ for any $P \in \mathcal{P}$, then $1 - \alpha$ is called an* asymptotic confidence level *of $C(\mathbf{X})$.*

Definition 2.8. *If $\lim_{N\to\infty} P[\boldsymbol{\theta} \in C(\mathbf{X})] = 1 - \alpha$ for any $P \in \mathcal{P}$, then $C(\mathbf{X})$ is a* $1 - \alpha$ asymptotically correct *confidence set.*

There are many possible approaches, for example, one can look for "Asymptotically Pivotal" quantities; or invert acceptance regions of "Asymptotic Tests".

2.2.4 Profile Likelihood

We may compute approximate confidence intervals, in the sense of coverage, using the "profile likelihood". Consider likelihood $L(\boldsymbol{\mu}, \boldsymbol{\eta})$, based on observation $\boldsymbol{X} = \boldsymbol{x}$. Let

$$L_P(\boldsymbol{\mu}) = \sup_{\boldsymbol{\eta}} L(\boldsymbol{\mu}, \boldsymbol{\eta}) . \tag{2.60}$$

$L_P(\boldsymbol{\mu}) = L(\boldsymbol{\mu}, \boldsymbol{\eta}(\boldsymbol{\mu}))$ is called the *Profile Likelihood* for $\boldsymbol{\mu}$. This provides a lower bound on coverage. Users of the popular fitting package MINUIT (James and Roos, 1975) will recognize that the MINOS interval uses the idea of the profile likelihood. We remind the reader that, for Gaussian sampling, intervals obtained with the profile likelihood have exact coverage (Section 2.2).

The profile likelihood has good asymptotic behavior: let $\dim(\boldsymbol{\mu}) = k$. Consider the likelihood ratio:

$$\lambda(\boldsymbol{\mu}) = \frac{L_P(\boldsymbol{\mu})}{\max_{\boldsymbol{\theta}'} L(\boldsymbol{\theta}')} , \tag{2.61}$$

where $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\eta})$. The set

$$C(X) = \{\boldsymbol{\mu} : -2\log\lambda(\boldsymbol{\mu}) \le c_\alpha\} , \tag{2.62}$$

where c_α is the χ^2 corresponding to the $1 - \alpha$ probability point of a χ^2 with k degrees of freedom, is an $1 - \alpha$ asymptotically correct confidence set. It may however not provide accurate coverage for small samples. Corrections to the profile likelihood exist that improve the behavior for finite samples (Reid, 2003).

2.2.5 Conditional Likelihood

Consider likelihood $L(\boldsymbol{\mu}, \boldsymbol{\eta})$. Suppose $T_{\boldsymbol{\eta}}(\boldsymbol{X})$ is a sufficient statistic for $\boldsymbol{\eta}$ for any given $\boldsymbol{\mu}$. Then, conditional distribution $f(\boldsymbol{X}|T_{\boldsymbol{\eta}}; \boldsymbol{\mu})$ does not depend on $\boldsymbol{\eta}$. The likelihood function corresponding to this conditional distribution is called the *conditional likelihood*. Note that estimates (e.g., MLE for $\boldsymbol{\mu}$) based on conditional likelihood may be different than for those based on full likelihood. This eliminates the nuisance parameter problem, if it can be done without too high a price.

For example, suppose we want to test the consistency of two Poisson distributed numbers. Such a question might arise concerning the existence of a signal in the presence of background. Our sampling distribution is

$$P(m, n) = \frac{\mu^m \nu^n}{m!n!} e^{-(\mu+\nu)} . \tag{2.63}$$

The null hypothesis is $H_0 \colon \mu = \nu$, to be tested against alternative $H_1 \colon \mu \neq \nu$. We are thus interested in the difference between the two means; the sum is effectively a nuisance parameter. A sufficient statistic for the sum is $N = m + n$. That is, we are interested in

$$\begin{aligned} P(n|m+n=N) &= \frac{P(N|n)P(n)}{P(N)} \\ &= \frac{\mu^{N-n}e^{-\mu}}{(N-n)!}\frac{\nu^n e^{-\nu}}{n!} \Bigg/ \frac{(\mu+\nu)^N e^{-(\mu+\nu)}}{N!} \\ &= \binom{N}{n}\left(\frac{\nu}{\mu+\nu}\right)^n \left(\frac{\mu}{\mu+\nu}\right)^{N-n} . \end{aligned} \tag{2.64}$$

This probability now permits us to construct a uniformly most powerful test of our hypothesis (Lehmann and Romano, 2005). Note that it is simply a binomial distribution, for given *N*. The uniformly most powerful property holds independently of *N*, although the probabilities cannot be computed without *N*.

The null hypothesis corresponds to $\mu = \nu$, that is

$$P(n|m+n=N) = \binom{N}{n}\left(\frac{1}{2}\right)^N . \tag{2.65}$$

For example, with $N = 916$ and $n = 424$, the *p*-value is 0.027, assuming a two-tailed probability is desired. This may be compared with an estimate of 0.025 in the normal approximation. Note that for our binomial calculation we have included the endpoints (424 and 492). If we try to mimic more closely the normal estimate by subtracting one-half the probability at the endpoints, we obtain 0.025, essentially the normal number. We have framed this in terms of a hypothesis test, but confidence intervals on the difference $\nu - \mu$ may likewise be obtained. The estimation of the ratio of Poisson means is a frequently encountered problem that can be addressed similarly (Reid, 2003).

2.3 Fits for Small Statistics

Often we are faced with extracting parametric information from data with only a few samplings, that is, in the case of "small statistics". At large statistics, the central