CHAPTER 8

Ancillary statistics

8.1 Introduction

In problems of parametric inference, an ancillary statistic is one whose distribution is the same for all parameter values. This is the simplest definition and the one that will be used here. Certain other authors, e.g. Cox & Hinkley (1974), require that the statistic in question be a component of the minimal sufficient statistic, and this is often a reasonable requirement if only to ensure uniqueness in certain special cases. More complicated definitions are required to deal satisfactorily with the case where nuisance parameters are present. For simplicity, nuisance parameters are not considered here.

Ancillary statistics usually arise in one of two ways. One simple example of each will suffice to show that no frequency theory of inference can be considered entirely satisfactory or complete if ancillary statistics are ignored. Even an approximate theory must attempt to come to grips with conditional inference if approximations beyond the first order are contemplated.

The first and most common way in which an ancillary statistic occurs is as a random sample size or other random index identifying the experiment actually performed. Suppose that N is a random variable whose probability distribution is

$$\mathrm{pr}(N = 1) = 0.5, \quad \mathrm{pr}(N = 100) = 0.5.$$

Conditionally on $N = n$, the random variables $Y_1, \ldots, Y_n$ are normally distributed with mean θ and unit variance. On the basis of the observed values n and $y_1, \ldots, y_n$, probabilistic statements concerning the value of θ are required.

The same problem can be posed in a number of ways. For instance, either of two machines can be used to measure a certain quantity θ. Both machines are unbiased, the first with variance 1.0, the second with variance 1/100. A machine is selected at random; a measurement is taken and the machine used is recorded.

In the first case N is ancillary: in the second case the indicator variable for the machine used is an ancillary statistic. In either case, the ancillary statistic serves as an index or indicator for the experiment actually performed.

Whatever the distribution of N, few statisticians would quarrel with the statement that $\hat{\theta} = \bar{Y}_N$ is normally distributed with mean θ and variance $1/N$, even though N is a random variable. Implicitly, we invoke the conditionality principle, which states that values of the ancillary other than that observed are irrelevant for the purposes of inference. The unconditional variance of $\hat{\theta}$, namely $E(1/N) = 1/2 + 1/200 = 0.505$, is relevant only if N is not observed and if its distribution as stated above is correct.

In fact, ancillary statistics of this kind arise frequently in scientific work where observations are often censored or unavailable for reasons not connected with the phenomenon under investigation. Accidents will happen, so the saying goes.

The second type of ancillary arises in the context of a translation model or, more generally, a group transformation model. To take a very simple example, suppose that $Y_1, \ldots, Y_n$ are independent and identically distributed uniformly on the interval $(\theta, \theta + 1)$. This is an instance of a non-regular problem in which the sample space is parameter dependent. The sufficient statistic is the pair of extreme values, $(Y_{(1)}, Y_{(n)})$, or equivalently, the minimum $Y_{(1)}$ and the range $R = Y_{(n)} - Y_{(1)}$. Clearly, by invariance, the distribution of R is unaffected by the value of θ and hence R is ancillary. Given that $R = r$, the conditional distribution of $Y_{(1)}$ is uniform on the interval $(\theta,\ \theta{+}1{-}r)$. Equivalently, the conditional distribution of $Y_{(n)} = Y_{(1)} + r$ is uniform over the interval$(\theta{+}r,\ \theta{+}1)$. Either conditional distribution leads to conditional equi-tailed confidence intervals

$$\{y_{(1)} - (1-r)(1-\alpha/2), \quad y_{(1)} - (1-r)\alpha/2\} \tag{8.1}$$

at level $1-\alpha$. Although these intervals also have the required coverage probability unconditionally as well as conditionally, their unconditional properties seem irrelevant if the aim is to summarise the information in the data concerning the value of θ.

On the other hand, if the ancillary is ignored, we may base our inferences on the pivot $Y_{(1)} - \theta$ whose density is $n(1-y)^{n-1}$ over the interval $(0, 1)$. If $n = 2$,

$$(Y_{(1)} - 0.7764, \quad Y_{(1)} - 0.0253) \tag{8.2}$$

is an exact equi-tailed 90% confidence interval for θ. Even though it is technically correct and exact, the problem with this statement is that it is inadequate as a summary of the data actually observed. In fact, if $R = 0.77$, not a particularly extreme value, the conditional coverage of (8.2) is approximately 96% rather than the 90% quoted. At the other extreme, if $R \geq 0.975$, the conditional coverage of (8.2) is exactly zero so that the unconditional 90% coverage is quite misleading.

These examples demonstrate the poverty of any frequency theory of inference that ignores ancillary statistics. Having made that point, it must be stated that the difficulties involved in establishing a coherent frequency theory of conditional inference are formidable. Current procedures, even for an approximate conditional theory, are incomplete although much progress has been made in the past decade. Two difficulties are connected with the existence and uniqueness of ancillary statistics. In general, exactly ancillary statistics need not exist. Even where they do exist they need not be unique. If two ancillary statistics exist, they are not, in general, jointly ancillary and an additional criterion may be required to choose between them (Cox, 1971). One device that is quite sensible is to require ancillaries to be functions of the minimal sufficient statistic. Even this does not guarantee uniqueness. Exercises 8.2–8.4 discuss one example where the maximal ancillary is not unique but where the conclusions given such an ancillary are unaffected by the choice of ancillary.

Given the overwhelming magnitude of these difficulties, it appears impossible to devise an exact frequency theory of conditional inference to cover the most general cases. For these reasons, we concentrate here on approximate theories based on statistics that are ancillary in a suitably defined approximate sense. It is tempting here to make a virtue out of necessity, but the approximate theory does have the advantage of greater 'continuity' between those models for which exact analysis is possible and the majority of models for which no exact theory exists.

8.2 Joint cumulants

8.2.1 *Joint cumulants of A and U*

Suppose that A has a distribution not depending on the parameter θ. The moments and cumulants do not depend on θ and hence their derivatives with respect to θ must vanish. Thus, differentiation of

$$E(A(Y)) = \int A(y) f_Y(y;\theta)dy = \text{const}$$

with respect to θ gives

$$\int A(y)u_r(\theta;y)f_Y(y;\theta)dy = \text{cov}(A, U_r) = 0.$$

This final step assumes the usual regularity condition that it is legitimate to interchange the order of differentiation and integration. In particular, the sample space must be the same for all parameter values.

Further differentiation gives

$$\mathrm{cov}(A, U_{[rs]}) = 0; \quad \mathrm{cov}(A, U_{[rst]}) = 0, \ldots$$

and so on.

If A is ancillary, all functions of A are also ancillary. Hence it follows that the joint cumulants

$$\begin{aligned} \kappa_3(A, A, U_r) = 0, &\qquad \kappa_4(A, A, A, U_r) = 0, \ldots \\ \kappa_3(A, A, U_{[rs]}) = 0, &\qquad \kappa_4(A, A, A, U_{[rs]}) = 0, \ldots \end{aligned}$$

also vanish. These results extend readily to higher-order joint cumulants provided that these cumulants involve one of U_r, $U_{[rs]}$, $U_{[rst]}, \ldots$ exactly once.

In general, although U_r is uncorrelated with A and with all functions of A, the conditional variance of U_r given A may be heavily dependent on the value of A. For instance, if A is a random sample size, the conditional covariance matrix of U_r given $A = a$ is directly proportional to a.

These results apply to *any* exact ancillary whatsoever: they also apply approximately to any approximate ancillary suitably defined.

8.2.2 *Conditional cumulants given A*

The conditional distribution of the data given $A = a$ is

$$f_{Y|A}(y|a; \theta) = f_Y(y; \theta)/f_A(a).$$

Hence, the conditional log likelihood differs from the unconditional log likelihood by a function not involving θ. Derivatives with respect to θ of the conditional log likelihood are therefore identical to derivatives of the unconditional log likelihood. With some exceptions as noted below, the conditional cumulants are different from the unconditional cumulants. In particular, it follows from differentiating the identity

$$\int f_{Y|A}(y|a; \theta) = 1$$

that the following identities hold

$$\begin{aligned} E(U_r|A) &= 0 \\ E(U_{[rs]}|A) &= 0 \\ E(U_{[rst]}|A) &= 0. \end{aligned}$$

Thus, the Bartlett identities (7.2) for the null cumulants of log likelihood derivatives are true conditionally on the value any ancillary statistic as well as unconditionally. Other cumulants or combinations of cumulants are affected by conditioning in a way that depends on the particular ancillary selected.

8.3 Local ancillarity

Exactly ancillary statistics arise usually in one of two ways, either as a random sample size or other index of the experiment actually performed, or as as a configuration statistic in location problems. These instances of exactly ancillary statistics are really rather special and in the majority of problems, no exactly ancillary statistic exists. In some cases it is known that no exact ancillary exists: in other cases no exact ancillary is known and it is suspected that none exists. For that reason, we concentrate our attention here on statistics that are approximately ancillary in some suitable sense.

Let θ_0 be an arbitrary but specified parameter value and let $A = A(Y; \theta_0)$ be a candidate ancillary statistic. If the distribution of A is approximately the same for all parameter values in some suitably chosen neighbourhood of θ_0, we say that A is approximately ancillary or locally ancillary in the vicinity of θ_0. More precisely, if the distribution of A under the density evaluated at $\theta_0 + n^{-1/2}\delta$ satisfies

$$f_A(a; \theta_0 + n^{-1/2}\delta) = f_A(a; \theta_0)\{1 + O(n^{-q/2})\}, \tag{8.3}$$

we say that A is qth-order locally ancillary in the vicinity of θ_0 (Cox, 1980). Two approximations are involved in this definition of approximate ancillarity. First, there is the $O(n^{-q/2})$ tolerance in (8.3), where n is the sample size or other measure of the extent of the observational record. Second, the definition is local, applying only to parameter values in an $O(n^{-1/2})$ neighbourhood of the target value, θ_0. The latter restriction is reasonable, at least in large samples, because if θ_0 is the true parameter point, the likelihood function eventually becomes negligible outside this $O(n^{-1/2})$ neighbourhood.

From (8.3) it can be seen that the log likelihood based on A satisfies

$$l_A(\theta_0 + n^{-1/2}\delta) = l_A(\theta_0) + O(n^{-q/2}). \tag{8.4}$$

Since $\partial l/\partial\theta = n^{1/2}\partial l/\partial\delta$, it follows that the Fisher information based on A for θ at θ_0 must be $O(n^{1-q/2})$ at most. This criterion is also used as a definition of approximate ancillarity.

If the distribution of A can be approximated by an Edgeworth series in which the cumulants are κ_1, κ_2, $\kappa_3, \ldots$, condition (8.3) is equivalent to

$$\nabla\kappa_r \equiv \kappa_r(\theta_0 + n^{-1/2}\delta) - \kappa_r(\theta_0) = O(n^{-q/2}), \quad r = 1, 2, \ldots \tag{8.5}$$

provided that A has been standardized so that $A - \kappa_1(\theta_0)$ is $O_p(1)$. Since κ_r is $O(n^{1-r/2})$, it is necessary only to check the first q cumulants of A in order to verify the property of local ancillarity. Usually (8.5) is much easier to verify than (8.3). Note, however, that A is usually a vector or array of random variables so that the rth cumulant array has a large number of components, each of which must satisfy (8.5).

We now proceed to show that approximately ancillary statistics exist and that they may be constructed from log likelihood derivatives. Since there is little merit in considering derivatives that are not tensors, we work with the tensorial derivatives V_r, $V_{rs}, \ldots$ at θ_0, as defined in (7.6). These derivatives are assumed to be based on n independent observations, to be asymptotically normal and to have all cumulants of order $O(n)$.

Under θ_0, the expectation of V_{rs} is $n\nu_{rs}$. By (7.4), the expectation of V_{rs} under the density at $\theta_0 + n^{-1/2}\delta$ is

$$n\{\nu_{rs} + \nu_{rs,i}\delta^i/n^{1/2} + \nu_{rs,[ij]}\delta^i\delta^j/(2n) + \cdots\}.$$

In terms of the standardized random variables $Z_{rs} = n^{-1/2}(V_{rs} - n\nu_{rs})$, which are $O_p(1)$ both under θ_0 and under the density at $\theta_0 + \delta/n^{1/2}$, we have

$$E(Z_{rs}; \theta_0 + \delta/n^{1/2}) = \nu_{rs,i}\delta^i + n^{-1/2}\nu_{rs,[ij]}\delta^i\delta^j/2 + O(n^{-1}).$$

Thus, since $\nu_{rs,i} = 0$, it follows from (8.5) that Z_{rs} is first-order locally ancillary in the vicinity of θ_0. Hence V_{rs}, the tensorial array of second derivatives, is first-order locally ancillary.

A similar argument shows that V_{rst} is also first-order locally ancillary but not ancillary to second order.

To construct a statistic that is second-order locally ancillary, we begin with a first-order ancillary statistic, Z_{rs}, that is $O_p(1)$ and aim to make a suitable adjustment of order $O_p(n^{-1/2})$. Thus, we seek coefficients $\beta_{rs}^{i,j}$, $\beta_{rs}^{i,jk}$ such that the cumulants of

$$A_{rs} = Z_{rs} - n^{-1/2}\beta_{rs}^{i,j} Z_i Z_j - n^{-1/2}\beta_{rs}^{i,jk} Z_i Z_{jk} \tag{8.6}$$

satisfy the ancillarity conditions up to second order. Since ancillarity is preserved under transformation, it is unnecessary in (8.6) to include quadratic terms in Z_{rs}. The differences between the cumulants of Z_r, Z_{rs} under θ_0 and under $\theta_0 + n^{-1/2}\delta$ are given by

$$\begin{aligned}
\nabla E(Z_r) &= \nu_{r,i}\delta^i + n^{-1/2}\nu_{r,[ij]}\delta^i\delta^j/2 + O(n^{-1}) \\
\nabla E(Z_{rs}) &= n^{-1/2}\nu_{rs,[ij]}\delta^i\delta^j/2 + O(n^{-1}) \\
\nabla \operatorname{cov}(Z_r, Z_s) &= n^{-1/2}\nu_{r,s,i}\delta^i + O(n^{-1}) \\
\nabla \operatorname{cov}(Z_r, Z_{st}) &= n^{-1/2}\nu_{r,st,i}\delta^i + O(n^{-1}).
\end{aligned}$$

It follows that the mean of A_{rs} changes by an amount

$$\nabla E(A_{rs}) = n^{-1/2}\{\nu_{rs,[ij]} - 2\beta_{rs}^{k,l}\nu_{i,k}\nu_{j,l}\}\delta^i\delta^j + O(n^{-1}).$$

Thus, if we choose the coefficients

$$\beta_{rs}^{i,j} = \nu_{rs,[kl]}\nu^{i,k}\nu^{j,l}/2, \tag{8.7}$$

it follows that $\nabla E(A_{rs}) = O(n^{-1})$ as required. The coefficients $\beta_{rs}^{i,j}$ are uniquely determined by second-order ancillarity.

To find the remaining coefficients, it is necessary to compute $\nabla \operatorname{cov}(A_{rs}, A_{tu})$ and to ensure that this difference is $O(n^{-1})$. Calculations similar to those given above show that

$$\nabla \operatorname{cov}(A_{rs}, A_{tu}) = n^{-1/2}\{\nu_{rs,tu,i}\delta^i - \beta_{rs}^{i,jk}\nu_{jk,tu}\nu_{i,l}\delta^l[2]\} + O(n^{-1}).$$

The coefficients $\beta_{rs}^{i,jk}$ are not uniquely determined by the requirement of second-order ancillarity unless the initial first-order statistic is a scalar. Any set of coefficients that satisfies

$$\{\beta_{rs}^{i,jk}\nu_{jk,tu} + \beta_{tu}^{i,jk}\nu_{jk,rs}\}\nu_{i,v} = \nu_{rs,tu,v} \tag{8.8}$$

gives rise to a statistic that is locally ancillary to second order. If, to the order considered, all such ancillaries gave rise to the same sample space conditionally, non-uniqueness would not be a problem. In fact, however, two second-order ancillaries constructed in the above manner need not be jointly ancillary to the same order.

So far, we have not imposed the obvious requirement that the coefficients $\beta_{rs}^{i,jk}$ should satisfy the transformation laws of a tensor. Certainly, (8.8) permits non-tensorial solutions. With the restriction to tensorial solutions, it might appear that the only solution to (8.8) is

$$\beta_{rs}^{i,jk} = \nu_{rs,tu,v}\nu^{tu,jk}\nu^{i,v}/2,$$

where $\nu^{rs,tu}$ is a generalized inverse of $\nu_{rs,tu}$. Without doubt, this is the most obvious and most 'reasonable' solution, but it is readily demonstrated that it is not unique. For instance, if we define the tensor $\epsilon_{rs}^{i,jk}$ by

$$\epsilon_{rs}^{i,jk}\nu_{jk,tu} = \{\nu_{rs,t,u,v} - \nu_{r,s,tu,v}\}\nu^{i,v},$$

it is easily seen that $\beta^{i,jk}_{rs} + \epsilon^{i,jk}_{rs}$ is also a solution to (8.8). In fact, any scalar multiple of ϵ can be used here.

It is certainly possible to devise further conditions that would guarantee uniqueness or, alternatively, to devise criteria in order to select the most 'relevant' of the possible approximate ancillaries. In the remainder of this section, however, our choice is to tolerate the non-uniqueness and to explore its consequences for conditional inference.

The construction used here for improving the order of ancillarity is taken from Cox (1980), who considered the case of one-dimensional ancillary statistics, and from McCullagh (1984a), who dealt with the more general case. Skovgaard (1986c) shows that, under suitable regularity conditions, the order of ancillarity may be improved indefinitely by successive adjustments of decreasing orders. Whether it is desirable in practice to go much beyond second or third order is quite another matter.

8.4 Stable combinations of cumulants

In Section 8.2.2, it was shown that, conditionally on any ancillary however selected, certain combinations of cumulants of log likelihood derivatives are identically zero. Such combinations whose value is unaffected by conditioning may be said to be *stable*. Thus, for instance, ν_i and $\nu_{ij} + \nu_{i,j}$ are identically zero whether we condition on A or not. It is important to recognize stable combinations in order to determine the effect, if any, of the choice of ancillary on the conclusions reached.

In this section, we demonstrate that, for the type of ancillary considered in the previous section, certain cumulants and cumulant combinations, while not exactly stable, are at least stable to first order in n. By way of example, it will be shown that $\nu_{i,j,k,l}$ and $\nu_{ij,kl}$ are both unstable but that the combination

$$\nu_{i,j,k,l} - \nu_{ij,kl}[3] \tag{8.9}$$

is stable to first order.

Suppose then that A, with components A_r, is ancillary, either exactly, or approximately to some suitable order. It is assumed that the joint distribution of (A, V_i, V_{ij}) may be approximated by an Edgeworth series and that all joint cumulants are $O(n)$. The dimension of A need not equal p but must be fixed as $n \to \infty$. The mean and variance of A are taken to be 0 and $n\lambda_{r,s}$ respectively. Thus, $A_r = O_p(n^{1/2})$ and $A^r = n^{-1}\lambda^{r,s}A_s$ is $O_p(n^{-1/2})$.

From the results given in Section 5.6, the conditional covariance of V_i and V_j is

$$\begin{aligned}\kappa_2(V_i, V_j|A) &= n\{\nu_{i,j} + \nu_{r;i,j}A^r + O(n^{-1})\}\\ &= n\nu_{i,j} + O(n^{1/2}),\end{aligned}$$

where $\nu_{r;i,j}$ is the third-order joint cumulant of A_r, V_i, V_j. Note that the ancillarity property

$$\mathrm{cov}(A_r, V_i) = \nu_{r;i} = 0$$

greatly simplifies these calculations. Similarly, in the case of the third cumulant, we have

$$\begin{aligned}\kappa_3(V_i, V_j, V_k|A) &= n\{\nu_{i,j,k} + \nu_{r;i,j,k}A^r + O(n^{-1})\}\\ &= n\nu_{i,j,k} + O(n^{1/2}),\end{aligned}$$

Thus, to first order at least, $\nu_{i,j}$ and $\nu_{i,j,k}$ are unaffected by conditioning. We say that these cumulants are *stable to first order*.

On the other hand, from the final equation in Section 5.6.2, we find that the conditional fourth cumulant of V_i, V_j, V_k, V_l given A is

$$\kappa_4(V_i, V_j, V_k, V_l|A) = n\{\nu_{i,j,k,l} - \nu_{r;i,j}\nu_{s;k,l}\lambda^{r,s}[3] + O(n^{-1/2})\}.$$

In this case, unlike the previous two calculations, conditioning has a substantial effect on the leading term. Thus, $\nu_{i,j,k,l}$ is unstable to first order: its interpretation is heavily dependent on the conditioning event.

Continuing in this way, it may be seen that the conditional covariance matrix of V_{ij} and V_{kl} is

$$\kappa_2(V_{ij}, V_{kl}|A) = n\{\nu_{ij,kl} - \nu_{r;ij}\nu_{s;kl}\lambda^{r,s}[3] + O(n^{-1/2})\}.$$

Again, this is an unstable cumulant. However, from the identity $\nu_{r;ij} = -\nu_{r;i,j}$ it follows that the combination

$$\kappa_4(V_i, V_j, V_k, V_l|A) - \kappa_2(V_{ij}, V_{kl}|A)[3] = n\{\nu_{i,j,k,l} - \nu_{ij,kl}[3] + O(n^{-1/2})\}$$

is stable to first order. Similarly, the conditional third cumulant

$$\kappa_3(V_i, V_j, V_{kl}|A) = n\{\nu_{i,j,kl} - \nu_{r;i,j}\nu_{s;kl}\lambda^{r,s} + O(n^{-1/2})\},$$

is unstable, whereas the combination

$$\nu_{i,j,kl} + \nu_{ij,kl} \tag{8.10}$$

is stable to first order.

These calculations are entirely independent of the choice of ancillary. They do not apply to the random sample size example discussed in Section 8.1 unless $\mu = E(N) \to \infty$, $\text{var}(N) = O(\mu)$ and certain other conditions are satisfied. However, the calculations do apply to the approximate ancillary constructed in the previous section. Note that, conditionally on the ancillary (8.6), the conditional covariance of V_{ij} and V_{kl} is reduced from $O(n)$ to $O(1)$, whereas the third-order joint cumulant of V_i, V_j and V_{kl} remains $O(n)$. This is a consequence of the stability of the combination (8.10).

In this context, it is interesting to note that the Bartlett adjustment factor (7.13) is a combination of the four stable combinations derived here. It follows that, up to and including terms of order $O(n^{-1})$, the likelihood ratio statistic is independent of all ancillary and approximately ancillary statistics.

8.5 Orthogonal statistic

Numerical computation of conditional distributions and conditional tail areas is often a complicated unappealing task. In many simple problems, particularly those involving the normal-theory linear model, the problem can be simplified to a great extent by 'regressing out' the conditioning statistic and forming a 'pivot' that is independent of the conditioning statistic. In normal-theory and other regression problems, the reasons for conditioning are usually connected with the elimination of nuisance parameters, but the same device of constructing a pivotal statistic can also be used to help cope with ancillary statistics. The idea is to start with an arbitrary statistic, V_i say, and by making a suitable minor adjustment, ensure that the adjusted statistic, S_i, is independent of the ancillary to the order required. Inference can then be based on the marginal distribution of S_i: this procedure is sometimes labelled 'conditional inference without tears' or 'conditional inference without conditioning'.

In those special cases where there is a complete sufficient statistic, S, for the parameters, Basu's theorem (Basu, 1955, 1958) tells us that all ancillaries are independent of S. This happy state of affairs means that inferences based on the marginal distribution of S are automatically conditional and are safe from conditionality criticisms of the type levelled in Section 8.1. By the same token, the result given at the end of the previous section shows that the maximized likelihood ratio statistic is similarly independent of all ancillaries and of approximate ancillaries to third order in n. Inferences based on the marginal distribution of the likelihood ratio statistic are, in large measure, protected against criticism on grounds of conditionality.

The score statistic, V_i, is asymptotically independent of all ancillaries, but only to first order in n. In other words, $\mathrm{cov}(V_i, A) = 0$ for all ancillaries. First-order independence is a very weak requirement and is occasionally unsatisfactory if n is not very large or if the ancillary takes on an unusual or extreme value. For this reason we seek an adjustment to V_i to make the adjusted statistic independent of all ancillaries to a higher order in n. For the resulting statistic to be useful, it is helpful to insist that it have a simple null distribution, usually normal, again to the same high order of approximation.

Thus, we seek coefficients $\gamma_r^{i,j}$, $\gamma_r^{i,jk}$ such that the adjusted statistic

$$S_r = Z_r + n^{-1/2}\{\gamma_r^{i,j}(Z_i Z_j - \nu_{i,j}) + \gamma_r^{i,jk} Z_i Z_{jk}\}$$

is independent of A to second order and also normally distributed to second order. Both of these calculations are made under the null density at θ_0.

For the class of ancillaries (8.6) based on the second derivatives of the log likelihood, we find

$$\begin{aligned}\mathrm{cov}(S_r, A_{st}) &= O(n^{-1})\\ \kappa_3(S_r, S_s, A_{tu}) &= n^{-1/2}\{\nu_{r,s,tu} - 2\beta_{tu}^{i,j}\nu_{i,r}\nu_{j,s}\\ &\qquad + \gamma_r^{i,jk}\nu_{jk,tu}\nu_{i,s}[2]\} + O(n^{-1})\end{aligned}$$

On using (8.7), we find

$$(\gamma_r^{i,jk}\nu_{i,s} + \gamma_s^{i,jk}\nu_{i,r})\nu_{jk,tu} = \nu_{rs,tu} \tag{8.11}$$

as a condition for orthogonality to second order. One solution, but by no means the only one, unless $p = 1$, is given by

$$\gamma_r^{i,jk}\nu_{i,s} = \delta_{rs}^{jk}/2. \tag{8.12}$$

The remaining third-order joint cumulant

$$\kappa_3(S_r, A_{st}, A_{uv}) = n^{-1/2}\{\nu_{r,st,uv} - \beta_{st}^{i,jk}\nu_{i,r}\nu_{jk,uv}[2]\} + O(n^{-1})$$

is guaranteed to be $O(n^{-1})$ on account of (8.8). The choice of ancillary among the coefficients satisfying (8.8) is immaterial.

Finally, in order to achieve approximate normality to second order, we require that the third-order cumulant of S_r, S_s and S_t be $O(n^{-1})$. This condition gives

$$\nu_{r,s,t} + \gamma_r^{i,j}\nu_{i,s}\nu_{j,t}[6] = 0.$$

Again, the solution is not unique unless $p = 1$, but it is natural to consider the 'symmetric' solution

$$\gamma_r^{i,j} = -\nu_{r,s,t}\nu^{i,s}\nu^{j,t}/6. \tag{8.13}$$

For the particular choice of coefficients (8.12) and (8.13), comparison with (7.15) shows that

$$S_r = W_r + \nu_{r,s,t}\nu^{s,t}/(6n^{1/2}) + O_p(n^{-1}),$$

where W_r are the components in the tensor decomposition of the likelihood ratio statistic given in Section 7.4.5.

In the case of scalar parameters, and using the coefficients (8.12) and (8.13), $W_r/i_{20}^{1/2}$ is equal to the signed square root of the likelihood ratio statistic. If we denote by ρ_3 the third standardized cumulant of $\partial l/\partial\theta$, then the orthogonal statistic may be written in the form

$$S = \pm\{2l(\hat\theta) - 2l(\theta_0)\}^{1/2} + \rho_3/6. \tag{8.14}$$

This statistic is distributed as $N(0,1) + O(n^{-1})$ independently of all ancillaries. The sign is chosen according to the sign of $\partial l/\partial\theta$.

8.6 Conditional distribution given A

In the previous sections it was shown that ancillary statistics, whether approximate or exact, are, in general, not unique, but yet certain useful formulae can be derived that are valid conditionally on any ancillary, however chosen. Because of the non-uniqueness of ancillaries, the most useful results must apply to as wide a class of *relevant* ancillaries as possible. This section is devoted to finding convenient expressions for the distributions of certain statistics such as the score statistic U_r or the maximum likelihood estimate $\hat{\theta}^r$, given the value of A. At no stage in the development is the ancillary specified. The only condition required is one of relevance, namely that the statistic of interest together with A should be sufficient to high enough order. Minimal sufficiency is not required.

It turns out that there is a particularly simple expression for the conditional distribution of $\hat{\theta}$ given A and that this expression is either exact or, if not exact, accurate to a very high order of asymptotic approximation. This conditional distribution, which may be written in the form

$$p(\hat{\theta};\theta|A) = (2\pi c)^{-p/2}|\hat{\jmath}_{rs}|^{1/2}\exp\{l(\theta) - l(\hat{\theta})\}\{1 + O(n^{-3/2})\}$$

is known as Barndorff-Nielsen's formula (Barndorff-Nielsen, 1980, 1983). One peculiar aspect of the formula is that the ancillary is not specified and the formula appears to be correct for a wide range of ancillary statistics. For this reason the description 'magical mystery formula' is sometimes used. In this expression $\log c = b$, the Bartlett adjustment factor, $\hat{\jmath}$ is the *observed* information matrix regarded as a function of $\hat{\theta}$ and A, and the formula is correct for any relevant ancillary.

It is more convenient at the outset to work with the score statistic with components $U_r = \partial l/\partial\theta$ evaluated at $\theta = 0$, a value chosen here for later convenience of notation. All conditional cumulants of U are assumed to be $O(n)$ as usual. In what follows, it will be assumed that A is locally ancillary to third order and that the pair (U, A) is jointly sufficient to the same order. In other words, for all θ in some neighbourhood of the origin, the conditional log likelihood based on U satisfies

$$l_{U|A}(\theta) - l_{U|A}(0) = l_Y(\theta) - l_Y(0) + O(n^{-3/2}). \tag{8.15}$$

Ancillary statistics satisfying these conditions, at least for θ in an $O(n^{-1/2})$ neighbourhood of the origin, can be constructed along the lines described in Section 8.3, but starting with the second and third derivatives jointly. In the case of location models, or more generally for group transformation models, exactly ancillary statistics exist that satisfy the above property for all θ. Such ancillaries typically have dimension of order $O(n)$. Since no approximation will be used here for the marginal distribution of A, it is not necessary to impose restrictions on its dimension. Such restrictions would be necessary if Edgeworth or saddlepoint approximations were used for the distribution of A.

The first step in the derivation is to find an approximation to the log likelihood function in terms of the conditional cumulants of U given $A = a$. Accordingly, let $K(\xi)$ be the conditional cumulant generating function of U given A at $\theta = 0$. Thus, the conditional cumulants $\kappa_{r,s}$, $\kappa_{r,s,t}, \ldots$ are functions of a. By the saddlepoint approximation, the conditional log density of U given A at $\theta = 0$ is

$$\begin{aligned}-K^*(u) + \log|K^{*rs}(u)|/2 - p\log(2\pi)/2 - (3\rho_4^* - 4\rho_{23}^{*2})/4!& \\ + O(n^{-3/2}),&\end{aligned} \tag{8.16}$$

where $K^*(u)$ is the Legendre transformation of $K(\xi)$. This is the approximate log likelihood at $\theta = 0$. To obtain the value of the log likelihood at an arbitrary point, θ, we require the conditional cumulant generating function $K(\xi;\theta)$ of U given A at θ.

For small values of θ, the conditional cumulants of U have their usual expansions about $\theta = 0$ as follows.

$$\begin{aligned}E(U_r;\theta) &= \kappa_{r,s}\theta^s + \kappa_{r,[st]}\theta^s\theta^t/2! + \kappa_{r,[stu]}\theta^s\theta^t\theta^u/3! + \cdots \\ \text{cov}(U_r, U_s;\theta) &= \kappa_{r,s} + \kappa_{r,s,t}\theta^t + \kappa_{r,s,[tu]}\theta^t\theta^u/2! + \cdots \\ \kappa_3(U_r, U_s, U_t;\theta) &= \kappa_{r,s,t} + \kappa_{r,s,t,u}\theta^u + \cdots .\end{aligned}$$

These expansions can be simplified by suitable choice of coordinate system. For any given value of A, we may choose a coordinate system in the neighbourhood of the origin satisfying $\kappa_{r,st} = 0$, $\kappa_{r,stu} = 0$, so that all higher-order derivatives are conditionally uncorrelated with U_r. This property is achieved using the transformation (7.5). Denoting the cumulants in this coordinate system by ν with appropriate indices, we find after collecting certain terms that

$$\begin{aligned} E(U_r;\theta) &= K_r(\theta) + \nu_{r,s,tu}\theta^s\theta^t\theta^u[3]/3! + \cdots \\ \operatorname{cov}(U_r, U_s;\theta) &= K_{rs}(\theta) + \nu_{r,s,tu}\theta^t\theta^u/2! + \cdots \\ \kappa_3(U_r, U_s, U_t;\theta) &= K_{rst}(\theta) + \cdots . \end{aligned} \tag{8.17}$$

Thus, the conditional cumulant generating function of U under θ is

$$\begin{aligned} K(\xi;\theta) = K(\theta+\xi) - K(\theta) &+ \nu_{r,s,tu}\xi^r\theta^s\theta^t\theta^u[3]/3! \\ &+ \nu_{r,s,tu}\xi^r\xi^s\theta^t\theta^u/4 + \cdots . \end{aligned}$$

The final two terms above measure a kind of departure from simple exponential family form even after conditioning. To this order of approximation, an arbitrary model cannot be reduced to a full exponential family model by conditioning.

To find the Legendre transformation of $K(\xi;\theta)$, we note first that

$$K^*(u) - \theta^r u_r + K(\theta)$$

is the Legendre transformation of $K(\theta+\xi) - K(\theta)$. To this must be added a correction term of order $O(n^{-1})$ involving $\nu_{r,s,tu}$. Note that $\nu_{rs,tu}$, the so-called 'curvature' tensor or covariance matrix of the residual second derivatives, does not enter into these calculations. A straightforward calculation using Taylor expansion shows that the Legendre transformation of $K(\xi;\theta)$ is

$$\begin{aligned} K^*(u;\theta) = K^*(u) - \theta^r u_r + K(\theta) &- \nu_{r,s,tu}u^r u^s\theta^t\theta^u/4 \\ &+ \nu_{r,s,tu}\theta^r\theta^s\theta^t\theta^u/4 + O(n^{-3/2}). \end{aligned}$$

It may be checked that $K^*(u;\theta)$, evaluated at the mean of U under θ given by (8.17), is zero as it ought to be.

We are now in a position to use the saddlepoint approximation for a second time, but first we require the log determinant of the array of second derivatives. In subsequent calculations, terms that have an effect of order $O(n^{-3/2})$ on probability calculations are ignored without comment. Hence, the required log determinant is

$$\log\det K^{*rs}(u;\theta) = \log\det K^{*rs}(u;0) - \nu_{r,s,tu}\nu^{r,s}\theta^t\theta^u/2.$$

Exercise 1.16 is useful for calculating log determinants.

The log likelihood function or the log density function may be written in terms of the Legendre transformation as

$$\begin{aligned} l(\theta) - l(0) &= -K^*(u;\theta) + K^*(u) \\ &\qquad + \tfrac{1}{2}\log\det K^{*rs}(u;\theta) - \tfrac{1}{2}\log\det K^{*rs}(u) \\ &= \theta^r u_r - K(\theta) + \nu_{r,s,tu}u^r u^s\theta^t\theta^u/4 \\ &\qquad - \nu_{r,s,tu}\theta^r\theta^s\theta^t\theta^u/4 - \nu_{r,s,tu}\nu^{r,s}\theta^t\theta^u/4 \end{aligned} \tag{8.18}$$

from which it can be seen that the conditional third cumulant $\nu_{r,s,tu}$ governs the departure from simple exponential form. This completes the first step in our derivation.

It is of interest here to note that the change in the derivative of $l(\theta)$, at least in its stochastic aspects, is governed primarily by the third term on the right of (8.18). It may be possible to interpret $\nu_{r,s,tu}$ as a curvature or torsion tensor, though its effect is qualitatively quite different from Efron's (1975) curvature, which is concerned mainly with the variance of the residual second derivatives. The latter notion of curvature is sensitive to conditioning.

In many instances, the likelihood function and the maximized likelihood ratio statistic can readily be computed either analytically or numerically whereas the conditional cumulant generating function, $K(\xi)$ and the conditional Legendre transformation, $K^*(u)$ cannot, for the simple reason that A is not specified explicitly. Thus, the second step in our derivation is to express $K^*(u)$ and related quantities directly in terms of the likelihood function and its derivatives.

The likelihood function given above has its maximum at the point

$$\hat{\theta}^r = K^{*r}(u) + \epsilon^r(u) + O(n^{-2}) \tag{8.19}$$

where ϵ^r is $O(n^{-3/2})$ given by

$$\epsilon^r = -\nu_{st,u,v}\nu^{r,s}\nu^{u,v}u^t/2 - \nu_{s,t,uv}\nu^{r,s}u^t u^u u^v/2.$$

On substituting $\hat{\theta}$ into (8.18), further simplification shows that the log likelihood ratio statistic is

$$l(\hat{\theta}) - l(0) = K^*(u) - \nu_{r,s,tu}\nu^{r,s}u^t u^u/4 + O(n^{-3/2}),$$

which differs from the conditional Legendre transformation of $K(\xi)$ by a term of order $O(n^{-1})$. Each of these calculations involves a small amount of elementary but tedious algebra that is hardly worth reproducing.

We now observe that

$$-\{l(\hat{\theta}) - l(0)\} + \tfrac{1}{2}\log\det K^{*rs}(u;\hat{\theta}) = -K^*(u) + \tfrac{1}{2}\log\det K^{*rs}(u), \tag{8.20}$$

which is the dominant term in the saddlepoint approximation for the conditional log density of U given A. It now remains to express the left member of the above equation solely in terms of the log likelihood function and its derivatives.

On differentiating (8.18), we find that

$$\begin{aligned} u_{rs}(\theta) = -K_{rs}(\theta) - \nu_{rs,t,u}\nu^{t,u}/2 - \nu_{r,s,tu}\theta^t\theta^u/2 - \nu_{r,t,su}[2]\theta^t\theta^u. \\ + \nu_{rs,t,u}u^t u^u/2 - \nu_{rs,t,u}\theta^t\theta^u/2. \end{aligned}$$

Hence, the observed Fisher information at $\hat{\theta}$ is

$$\hat{\jmath}_{rs} = -u_{rs}(\hat{\theta}) = K_{rs}(\hat{\theta}) + \nu_{r,s,tu}\hat{\theta}^t\hat{\theta}^u/2 + \nu_{rs,t,u}\nu^{t,u}/2 + \nu_{r,t,su}[2]\hat{\theta}^t\hat{\theta}^u$$

and the observed information determinant is given by

$$\log\det\hat{\jmath}_{rs} = -\log\det K^{*rs}(u;\hat{\theta}) + 2\nu_{r,t,su}\nu^{r,s}\hat{\theta}^t\hat{\theta}^u + \nu_{r,s,tu}\nu^{r,s}\nu^{t,u}/2.$$

On substituting into (8.20), it is seen that the saddlepoint approximation with one correction term for the conditional log density of U given A is

$$-\{l(\hat{\theta}) - l(\theta)\} - \tfrac{1}{2}\log\det\hat{\jmath}_{rs} - \tfrac{1}{2}p\log(2\pi) + \nu_{r,t,su}\nu^{r,s}\hat{\theta}^t\hat{\theta}^u - \tfrac{1}{2}pb(\theta), \tag{8.21}$$

where $b(\theta)$ is the Bartlett adjustment, given in this instance by

$$pb(\theta) = (3\rho_{13}^2 + 2\rho_{23}^2 - 3\rho_4)/12 - \nu_{r,s,tu}\nu^{r,s}\nu^{t,u}/2.$$

As pointed out in Section 8.4, $b(\theta)$ can be computed from the unconditional cumulants using the formula (7.13). Similarly, $\nu_{r,t,su}$ in (8.21) can be computed from the unconditional cumulants using (8.10).

Expression (8.21) can be simplified even further. The derivative of the transformation (8.19) from u to $\hat{\theta}$ is

$$\begin{aligned}\hat{\theta}^{rs} &= K^{*rs}(u) - \nu^{r,t}\nu^{s,u}\nu^{v,w}\nu_{tu,v,w}/2 - \nu^{r,t}\nu^{s,u}u^v u^w \nu_{t,u,vw}/2 \\ &\qquad - \nu^{r,t}\nu^{v,s}u^u u^w \nu_{tu,v,w}[2]/2 \\ &= K^{*rs}(u;\hat{\theta}) - \nu^{r,t}\nu^{s,u}\nu^{v,w}\nu_{tu,v,w}/2 - \nu^{r,t}\nu^{v,s}u^u u^w \nu_{tu,v,w}.\end{aligned}$$

Hence, the log determinant of the transformation is

$$\begin{aligned}\log\det\hat{\theta}^{rs} &= \log\det K^{*rs}(u;\hat{\theta}) - \nu^{r,s}\nu^{t,u}\nu_{rs,t,u}/2 - \nu^{r,s}u^t u^u \nu_{rt,s,u} \\ &= -\log\det\hat{\jmath}_{rs} + \nu_{r,t,su}\nu^{r,s}u^t u^u. \end{aligned} \tag{8.22}$$

Hence, under the assumption that $\theta = 0$ is the true parameter point, the conditional log density of $\hat{\theta}$ given A is

$$-\{l(\hat{\theta}) - l(\theta)\} + \tfrac{1}{2}\log\det\hat{\jmath}_{rs} - p\log(2\pi) - pb(\hat{\theta})/2.$$

More generally, if the true parameter point is θ, the conditional density of $\hat{\theta}$ as a function of θ and the conditioning variable may be written

$$p(\hat{\theta};\theta|A) = (2\pi\hat{c})^{-p/2}\exp\{l(\theta) - l(\hat{\theta})\}|\hat{\jmath}|^{1/2} \tag{8.23}$$

where $\hat{c} = \log b(\hat{\theta})$. The log likelihood function $l(\theta)$, in its dependence on the data, is to be regarded as a function of $(\hat{\theta}, a)$. Similarly for the observed information determinant. Thus, for example, $l(\theta) = l(\theta;\hat{\theta},a)$ is the log likelihood function, $u_{rs}(\theta) = u_{rs}(\theta;\hat{\theta},a)$ and $\hat{\jmath}_{rs} = u_{rs}(\hat{\theta};\hat{\theta},a)$ is the observed information matrix.

Approximation (8.23) gives the conditional density of the maximum likelihood estimate of the canonical parameter corresponding to the transformation (7.5). On transforming to any other parameter, the form of the approximation remains the same. In fact, the approximation is an invariant of *weight* 1 under re-parameterization in the sense of Section 6.1. Thus, (8.23) is equally accurate or inaccurate in all parameterizations and the functional form of the approximation is the same whatever parameterization is chosen.

8.7 Bibliographic notes

It would be impossible in the limited space available here to discuss in detail the various articles that have been written on the subject of ancillary statistics and conditional inference. What follows is a minimal set of standard references.

Fisher (1925, 1934) seems to have been first to recognize the need for conditioning to ensure that probability calculations are relevant to the data observed. His criticism of Welch's test (Fisher, 1956) was based on its unsatisfactory conditional properties. Other important papers that discuss the need for conditioning and the difficulties that ensue are Cox (1958, 1971), Basu (1964), Pierce (1973), Robinson (1975, 1978), Kiefer (1977), Lehmann (1981) and Buehler (1982). The book by Fraser (1968) is exceptional for the emphasis placed on group structure as an integral part of the model specification.

The question of the existence or otherwise of ancillary statistics (or similar regions) was first posed by R.A. Fisher as the 'problem of the Nile' in his 1936 Harvard Tercentenary lecture. It is now known that, in the continuous case, exactly ancillary statistics always exist if the parameter space is finite or if attention is restricted to any finite set of parameter values, however numerous.

This conclusion follows from Liapounoff's theorem (Halmos, 1948), which states that the range of a vector measure is closed and, in the non-atomic case, convex. On the other hand, it is also known that no such regions, satisfying reasonable continuity conditions, exist for the Behrens-Fisher problem (Linnik, 1968). It seems, then, that as the number of parameter points under consideration increases, regions whose probability content is exactly the same for all parameter values are liable to become increasingly 'irregular' in some sense. However, acceptable regions whose probability content is approximately the same for all parameter values do appear to exist in most instances.

Kalbfleisch (1975) makes a distinction, similar to that made in Section 8.1, between experimental ancillaries and mathematical ancillaries. Lloyd (1985ab), on the other hand, distinguishes between internal and external ancillaries. The former are functions of the minimal sufficient statistic.

Much of the recent work has concentrated on the notion of approximate ancillarity or asymptotic ancillarity. See, for example, Efron & Hinkley (1978), Cox (1980), Hinkley (1980), Barndorff-Nielsen (1980) for further details. The results given in Section 8.5 are taken from McCullagh (1984a).

Formula (8.23) was first given by Barndorff-Nielsen (1980) synthesizing known exact conditional results for translation models due to Fisher (1934), and approximate results for full exponential family models based on the saddlepoint approximation. In subsequent papers, (Barndorff-Nielsen, 1983, 1984, 1985), the formula has been developed and used to obtain conditional confidence intervals for one-dimensional parameters. More recent applications of the formula have been to problems involving nuisance parameters. The formulae in Section 6.4 are special cases of what is called the 'modified profile likelihood'.

The derivation of Barndorff-Nielsen's formula in Section 8.6 appears to be new.

8.8 Further results and exercises 8

8.1 Show that if (X_1, X_2) has the bivariate normal distribution with zero mean, variances σ_1^2, σ_2^2 and covariance $\rho\sigma_1\sigma_2$, then the ratio $U = X_1/X_2$ has the Cauchy distribution with median $\theta = \rho\sigma_1/\sigma_2$ and dispersion parameter $\tau^2 = \sigma_1^2(1-\rho^2)/\sigma_2^2$. Explicitly,

$$f_U(u;\theta,\tau) = \tau^{-1}\pi^{-1}\{1 + (u-\theta)^2/\tau^2\}^{-1},$$

where $-\infty < \theta < \infty$ and $\tau > 0$. Deduce that $1/U$ also has the Cauchy distribution with median $\theta/(\tau^2+\theta^2)$ and dispersion parameter $\tau^2/(\tau^2+\theta^2)^2$. Interpret the conclusion that θ/τ is invariant.

8.2 Let $X_1, \ldots, X_n$ be independent and identically distributed Cauchy random variables with unknown parameters (θ, τ). Let $\bar{X}$ and s_X^2 be the sample mean and sample variance respectively. By writing $X_i = \theta + \tau\epsilon_i$, show that the joint distribution of the *configuration statistic* A with components $A_i = (X_i - \bar{X})/s_X$ is independent of the parameters and hence that A is ancillary. [This result applies equally to any location-scale family where the ϵ_i are *i.i.d.* with known distribution.]

8.3 Using the notation of the previous exercise, show that for any constants a,b,c,d satisfying $ad - bc \neq 0$,

$$Y_i = (a + bX_i)/(c + dX_i) \qquad i = 1, \ldots, n$$

are independent and identically distributed Cauchy random variables. Deduce that the derived statistic A^* with components $A_i^* = (Y_i - \bar{Y})/s_Y$ has a distribution not depending on (θ, τ). Hence conclude that the maximal ancillary for the problem described in Exercise 8.2 is not unique. Demonstrate explicitly that two such ancillaries are not jointly ancillary. [This construction is specific to the two-parameter Cauchy problem.]

8.4 Suppose, in the notation previously established, that $n = 3$. Write the ancillary in the form $\{\text{sign}(X_3 - X_2), \text{sign}(X_2 - X_1)\}$, together with an additional component

$$A_X = (X_{(3)} - X_{(2)})/(X_{(2)} - X_{(1)}),$$

where $X_{(j)}$ are the ordered values of X. Show that A_X is a function of the sufficient statistic, whereas the first two components are not. Let $Y_i(t) = 1/(X_i - t)$ and denote by $A(t)$ the corresponding ancillary computed as a function of the transformed values. Show that the function $A(t)$ is continuous except at the three points $t = X_i$ and hence deduce that the data values may be recovered from the set of ancillaries $\{A(t), -\infty < t < \infty\}$. [In fact, it is enough to know the values of $A(t)$ at three distinct points interlacing the observed values. However, these cannot be specified in advance.]

8.5 Suppose that (X_1, X_2) are bivariate normal variables with zero mean, unit variance and unknown correlation ρ. Show that $A_1 = X_1$ and $A_2 = X_2$ are each ancillary, though not jointly ancillary, and that neither is a component of the sufficient statistic. Let $T = X_1 X_2$. Show that

$$\begin{aligned} T|A_1 = a_1 &\sim N\{\rho a_1^2, (1-\rho^2)a_1^2\} \\ T|A_2 = a_2 &\sim N\{\rho a_2^2, (1-\rho^2)a_2^2\}. \end{aligned}$$

8.6 In the notation of the previous exercise, suppose that it is required to test the hypothesis $H_0 : \rho = 0$, and that the observed values are $x_1 = 2$, $x_2 = 1$. Compute the conditional tail areas $\text{pr}(T \geq t|A_1 = a_1)$ and $\text{pr}(T \geq t|A_2 = a_2)$. Comment on the appropriateness of these tail areas as measures of evidence against H_0.

8.7 Suppose that (X_1, X_2, X_3) have the trivariate normal distribution with zero mean and intraclass covariance matrix with variances σ^2 and correlations ρ. Show that $-\frac{1}{2} \leq \rho \leq 1$. Prove that the moments of $X_1 + \omega X_2 + \omega^2 X_3$ and $X_1 + \omega X_3 + \omega^2 X_2$ are independent of both parameters, but that neither statistic is ancillary $[\omega = \exp(2\pi i/3)]$.

8.8 Suppose in the previous exercise that $\sigma^2 = 1$. Show that this information has no effect on the sufficient statistic but gives rise to ancillaries, namely X_1, X_2, X_3, no two of which are jointly ancillary.

8.9 Show that the tensorial decomposition of the likelihood ratio statistic in (7.15) is not unique but that all such decompositions are orthogonal statistics in the sense used in Section 8.5 above.

8.10 Show that the Legendre transformation of

$$\begin{aligned} K(\xi; \theta) = K(\theta + \xi) - K(\theta) &+ \nu_{r,s,tu}\xi^r\theta^s\theta^t\theta^u[3]/3! \\ &+ \nu_{r,s,tu}\xi^r\xi^s\theta^t\theta^u/4 + \cdots \end{aligned}$$

with respect to the first argument is approximately

$$\begin{aligned} K^*(u; \theta) = K^*(u) - \theta^r u_r + K(\theta) &- \nu_{r,s,tu}u^r u^s\theta^t\theta^u/4 \\ &+ \nu_{r,s,tu}\theta^r\theta^s\theta^t\theta^u/4. \end{aligned}$$

8.11 Show that the Legendre transformation, $K^*(u; \theta)$, evaluated at

$$u_r = K_r(\theta) + \nu_{r,s,tu}\theta^s\theta^t\theta^u[3]/3!,$$

is zero to the same order of approximation.

8.12 Show that the maximum of the log likelihood function (8.18) is given by (8.19).

8.13 Beginning with the canonical coordinate system introduced at (8.17), transform from θ to ϕ with components

$$\phi_r = K_r(\theta) + \nu_{r,s,tu}\theta^s\theta^t\theta^u/2 + \nu_{r,s,tu}\theta^s\nu^{t,u}/2.$$

Show that, although $E(U_r;\theta) \neq \phi_r$, nevertheless $\hat{\phi}_r = U_r$. Show also that the observed information determinant with respect to the components of ϕ satisfies

$$\tfrac{1}{2}\log\det K^{*rs}(u;\hat{\theta}) = \tfrac{1}{2}\log\det \hat{\jmath}_{\phi}^{rs} + \nu_{r,s,tu}\nu^{r,s}\nu^{t,u}$$

at the maximum likelihood estimate. Hence deduce (8.23) directly from (8.20).

8.14 Suppose that $Y_1, \ldots, Y_n$ are independent and identically distributed on the interval $\theta,\ \theta{+}1$. Show that the likelihood function is constant in the interval $(y_{(n)} - 1,\ y_{(1)})$ and is zero otherwise. Hence, interpret $r = y_{(n)} - y_{(1)}$ as an indicator of the shape of the likelihood function.