Chapter 2. CONDITIONING

The most commonly used measures of accuracy of evidence in statistics are *pre-experimental*. A particular procedure is decided upon for use, and the accuracy of the evidence from an experiment is identified with the long run behavior of the procedure, were the experiment repeatedly performed. This long run behavior is evaluated by averaging the performance of the procedure over the sample space $\mathcal{X}$. In contrast, the LP states that *post-experimental* reasoning should be used, wherein only the actual observation x (and not the other observations in $\mathcal{X}$ that could have occured) is relevant. There are a variety of intermediate positions which call for partial conditioning on x and partial long run frequency interpretations. Partly for historical purposes, and partly to indicate that the case for at least some sort of conditioning is compelling, we discuss in this chapter various conditioning viewpoints.

2.1 SIMPLE EXAMPLES

The following simple examples reveal the necessity of at least sometimes thinking conditionally, and will be important later.

EXAMPLE 1. Suppose X_1 and X_2 are independent and

$$P_\theta(X_i = \theta-1) = P_\theta(X_i = \theta+1) = \frac{1}{2}, \quad i = 1,2.$$

Here $-\infty < \theta < \infty$ is an unknown parameter to be estimated from X_1 and X_2. It is easy to see that a 75% confidence set of smallest size for θ is

$$C(X_1,X_2) = \begin{cases} \text{the point } \frac{1}{2}(X_1+X_2) & \text{if } X_1 \neq X_2 \\ \\ \text{the point } X_1-1 & \text{if } X_1 = X_2. \end{cases}$$

Thus, if repeatedly used in this problem, $C(X_1,X_2)$ would contain θ with probability .75.

Notice, however, that when $x_1 \neq x_2$ it is *absolutely certain* that $\theta = \frac{1}{2}(x_1+x_2)$, while when $x_1 = x_2$ it is equally uncertain whether $\theta = x_1-1$ or $\theta = x_1+1$ (assuming no prior knowledge about θ). Thus, from a post-experimental viewpoint, one would say that $C(x_1,x_2)$ contains θ with "confidence" 100% when $x_1 \neq x_2$, but only with "confidence" 50% when $x_1 = x_2$. Common sense certainly supports the post-experimental view here. It is technically correct to call $C(X_1,X_2)$ a 75% confidence set, but, if after seeing the data we know whether it is really a 100% or 50% set, reporting 75% seems rather silly.

The above example focuses the issue somewhat: does it make sense to report a pre-experimental measure when it is known to be misleading after seeing the data? The next example also seems intuitively clear, yet is the key to all that follows.

EXAMPLE 2. Suppose a substance to be analyzed can be sent either to a laboratory in New York or a laboratory in California. The two labs seem equally good, so a fair coin is flipped to choose between them, with "heads" denoting that the lab in New York will be chosen. The coin is flipped and comes up tails, so the California lab is used. After awhile, the experimental results come back and a conclusion must be reached. Should this conclusion take into account the fact that the coin could have been heads, and hence that the experiment in New York might have been performed instead?

This, of course, is a variant of the famous Cox example (Cox (1958)- see also Cornfield (1969)), which concerns being given (at random) either an accurate or an inaccurate measuring instrument (and knowing which was given). Should the conclusion reached by experimentation depend only on the instrument actually used, or should it take into account that the other instrument might have been obtained?

In symbolic form, we can phrase this example as a "mixed experiment" in which with probabilities $\frac{1}{2}$ (independent of θ) either experiment E_1 or experiment E_2 (both pertaining to θ) will be performed. Should the analysis depend only on the experiment actually performed, or should the possibility of having done the other experiment be taken into account?

The obvious intuitive answer to the questions in the above example is that only the experiment actually performed should matter. But this is counter to pre-experimental frequentist reasoning, which says that one should average over all possible outcomes (here, including the coin flip). One could argue that it is correct to condition on the coin flip, and then use the frequentist measures for the experiment actually performed, but the LP disallows this and is (surprisingly) derivable simply from conditioning on the coin flip plus sufficiency (see Chapter 3).

EXAMPLE 3. For a testing example, suppose it is desired to test H_0: $\theta = -1$ versus H_a: $\theta = 1$, based on $X \sim \mathcal{N}(\theta,.25)$. The rejection region $X \geq 0$ gives a test with error probabilities (type I and type II) of .0228. If $x = 0$ is observed, it is then permissible to state that H_0 is rejected, and that the error probability is $\alpha = .0228$. Common sense, however, indicates that the data $x = 0$ fails to discriminate at all between H_0 and H_a. Any sensible person would be equally uncertain as to the truth of H_0 or H_a (based just on the data $x = 0$). Suppose on the other hand, that $x = 1$ is observed. Then (pre-experimentally) one can still only reject at $\alpha = .0228$, but $x = 1$ is four standard deviations from $\theta = -1$, so the evidence against H_0 seems overwhelming.

Clearly, the actual intuitive evidence conveyed by x can be quite different from the pre-experimental evidence. This has led many frequentists to prefer the use of P-values to fixed error probabilities. The P-value (against H_0) would here be $P_{\theta=-1}(X \geq x)$, a measure of evidence against H_0 with much more dependence on the actual observation, x, than mere rejection at $\alpha = .0228$. (Even P-values can be criticized from a conditional viewpoint, however - see Section 4.4.)

Note that there is nothing logically wrong with reporting error probabilities in Example 3; it just seems to be an inadequate reflection of the evidence conveyed by the data to report $\alpha = .0228$ for *both* $x = 0$ and $x = 1$. Pratt (1977) (perhaps somewhat tongue-in-cheek) thus coins

THE PRINCIPLE OF ADEQUACY. A concept of statistical evidence is (very) inadequate if it does not distinguish evidence of (very) different strengths.

EXAMPLE 4a. Suppose X is 1, 2, or 3 and θ is 1 or 2, with $P_\theta(x)$ given in the following table:

		X	
	1	2	3
P_0	.009	.001	.99
P_1	.001	.989	.01

.

The test, which accepts P_0 when $x = 3$ and accepts P_1 otherwise, is a most powerful test with *both* error probabilities equal to .01. Hence, it would be valid to make the frequentist statement, upon observing $x = 1$, "My test has rejected P_0 and the error probability is .01." This seems very misleading, since the likelihood ratio is actually 9 to 1 in *favor* of P_0, which is being *rejected.*

EXAMPLE 4b. One could object in Example 4a, that the .01 level test is inappropriate, and that one should use the .001 level test, which rejects only when $x = 2$. Consider, however, the following slightly changed version:

		X	
	1	2	3
P_0	.005	.005	.99
P_1	.0051	.9849	.01

.

Again the test which rejects P_0 when $x = 1$ or 2 and accepts otherwise has error probabilities equal to .01, and now it indeed seems sensible to take the indicated actions (if we suppose an action *must* be taken). It still seems unreasonable, however, to report an error probability of .01 upon rejecting P_0 when x = 1, since the data provides very little evidence in favor of P_1.

EXAMPLE 5. For a decision theoretic example, consider the interesting Stein phenomenon, concerned with estimation of a p-variate normal mean ($p \geq 3$) based on $X \sim \mathcal{N}_p(\theta,I)$ and under sum of squares error loss. The usual pre-experimental measure of the performance of an estimator δ is the risk function (or expected loss)

$$R(\theta,\delta) = E_\theta \sum_{i=1}^{p} (\theta_i-\delta_i(X))^2.$$

The classical estimator here is $\delta^0(x) = x$, but James and Stein (1960) showed that

$$\delta^{J-S}(x) = (1 - \frac{p-2}{\Sigma x_i^2})x$$

has $R(\theta,\delta^{J-S}) < R(\theta,\delta^0) = p$ for all θ. One can thus report δ^{J-S} as always being better than δ^0 from a pre-experimental viewpoint. However, if p = 3 and x = (0,.01,.01) is observed, then

$$\delta^{J-S}(x) = (0,-49.99,-49.99),$$

which is an absurd estimate of θ. Hence δ^{J-S} can be terrible for certain x. Of course the positive part version of δ^{J-S},

$$\delta^{J-S+}(x) = (1 - \frac{p-2}{\Sigma x_i^2})^+ x,$$

corrects this glaring problem, but the point is that a procedure which looks great pre-experimentally could be terrible for particular x, and it may not always be so obvious when this is the case.

Confidence sets for θ can also be developed (see Casella and Hwang (1982)) which have larger probabilities of coverage than the classical confidence ellipsoids, are never larger in size, and for small $|x|$ consist of the single point {0}. Indeed, these sets are of the simple form

$$C(x) = \begin{cases} \{\theta: \ |\theta-\delta^{J-S+}(x)|^2 \leq \chi_p^2(1-\alpha)\} & \text{if } |x| > \varepsilon \\ \{0\} & \text{if } |x| < \varepsilon, \end{cases}$$

where $\chi_p^2(1-\alpha)$ is the 1-*αth* percentile of the chi-square distribution with p degrees of freedom, and ε is suitably small. Although this confidence procedure looks great pre-experimentally, one would look rather foolish to conclude when p = 3 and x = (0,.01,.01) that θ is the point {0} with confidence 95%.

The above examples, though simple, indicate most of the intuitive reasons for conditioning. There are a wide variety of other such examples. The Uniform (θ-α,θ+β) distribution (α,β known) provides a host of examples where conditional reasoning differs considerably from pre-experimental reasoning (c.f. Welch (1939) and Pratt (1961)). The Stein 2-stage procedure for obtaining a confidence interval of fixed width for the mean of a $\mathcal{N}(\theta,\sigma^2)$ distribution is another example. A preliminary sample allows estimation of σ^2, from which it is possible to determine the sample size needed for a second sample in order to guarantee an overall probability of coverage for a fixed width interval. But what if the second sample indicates that the preliminary estimate of σ^2 was woefully low? Then one would really have much less *real* confidence in the proposed interval (c.f. Lindley (1958) and Savage et. al. (1962)). Another example is regression on random covariates. It is common practice to perform the analysis conditionally on the observed values of the covariates, rather than giving confidence statements, etc., valid in an average sense over all covariates that could have been observed. Robinson (1975) also gives extremely compelling (though artificial) examples of the need to condition. Piccinato (1981) gives some interesting decision-theoretic examples.

A final important example is that of robust estimation. A convincing case can be made that inference statements should be made conditionally on the residuals; if the data looks completely like normal data, use normal theory. Barnard (1981) says

> "We should recognise that 'robustness' of inference is a conditional property - some inferences from some samples are robust. But other inferences, or the same inferences from other samples, may depend strongly on distributional assumptions."

Dempster (1975) contains very convincing discussion and a host of interesting examples concerning this issue. Related to conditional robustness is large sample inference, which should often be done conditionally on shape features of the likelihood function. Thus, in using asymptotic normal theory for the maximum likelihood estimator, $\hat{\theta}$, one should generally use $I(\hat{\theta})^{-1}$, the inverse of *observed* Fisher information, as the covariance matrix, rather than $I(\theta)^{-1}$, the inverse of *expected* Fisher information. For extensive discussion of these and related issues see Jeffreys (1961), Pratt (1965), Andersen (1970), Efron and Hinkley (1978), Barndorff-Nielsen (1980), Cox (1980), and Hinkley (1980a,1982).

2.2 RELEVANT SUBSETS

Fisher (c.f. Fisher(1956a)) long advocated conditioning on what he called *relevant subsets* of $\mathcal{X}$ (also called "recognizable subsets", "reference sets", or "conditional experimental frames of reference"). There is a considerable literature on the subject, which tends to be more formal than the intuitive type of reasoning presented in the examples of Section 2.1. The basic idea is to find subsets of $\mathcal{X}$ (often determined by statistics) which, when conditioned upon, change the pre-experimental measure. In Example 1, for instance,

$$\mathcal{X} = \{x: \ x_1 = x_2\} \cup \{x: \ x_1 \neq x_2\},$$

and the coverage probabilities of $C(X_1,X_2)$ conditioned on observing X in the "relevant" subsets $\{x: \ x_1 = x_2\}$ or $\{x: \ x_1 \neq x_2\}$ are 1 and .5, respectively. In Example 2, the two outcomes of the coin flip determine two relevant subsets. In Examples 3, 4, and 5 it is not clear what subsets should be considered relevant, but many reasonable choices give conditional results quite different from the pre-experimental results.

Formal theories of relevant subsets (c.f. Buehler (1959)) proceed in a fashion analogous to the following. Suppose C(x) is a confidence procedure with confidence coefficient $1-\alpha$ for all θ, i.e.,

(2.2.1) $$P_\theta(C(X) \text{ contains } \theta) = 1-\alpha \quad \text{for all } \theta.$$

Then B is called a relevant subset of $\mathcal{X}$ if, for some $\varepsilon > 0$, either

(2.2.2) $$P_\theta(C(X) \text{ contains } \theta | X \in B) \leq (1-\alpha) - \varepsilon \quad \text{for all } \theta$$

or

(2.2.3) $$P_\theta(C(X) \text{ contains } \theta | X \in B) \geq (1-\alpha) + \varepsilon \quad \text{for all } \theta.$$

When (2.2.2) or (2.2.3) holds and $x \in B$ is observed, it is questionable whether (2.2.1) should be the measure of evidence reported. This formed the basis of Fisher's objection (Fisher (1956b)) to the Aspin-Welch (1949) solution to the Behrens-Fisher problem (see also Yates (1964) and Cornfield (1969)). Another example follows. (For more examples, see Cornfield (1969), Olshen (1977), and Fraser (1977).)

EXAMPLE 6. (Brown (1967), with earlier related examples by Stein (1961) and Buehler and Fedderson (1963)). If $X_1,\ldots,X_n$ is a sample from a $\mathcal{N}(\theta,\sigma^2)$ distribution, both θ and σ^2 unknown, the usual $100(1-\alpha)\%$ confidence interval for θ is

$$C(\bar{x},s) = (\bar{x}-t_{\alpha/2}\frac{s}{\sqrt{n}},\ \bar{x}+t_{\alpha/2}\frac{s}{\sqrt{n}}),$$

where $\bar{x}$ and s are the sample mean and standard deviation, respectively, and $t_{\alpha/2}$ is the appropriate critical value for the t-distribution with n-1 degrees of freedom. For $n = 2$ and $\alpha = .5$ we thus have

$$P_{\theta,\sigma^2}(C(\bar{X},S) \text{ contains } \theta) = .5 \text{ for all } \theta,\sigma^2,$$

but Brown (1967) showed that

$$P_{\theta,\sigma^2}(C(\bar{X},S) \text{ contains } \theta \,\Big|\, |\bar{X}|/S \leq 1 + \sqrt{2}) \geq \frac{2}{3} \quad \text{for all } \theta,\sigma^2,$$

and hence the set

$$B = \{(x_1,\ldots,x_n): \ |\bar{x}|/s \leq 1 + \sqrt{2}\}$$

forms a relevant subset.

There is a considerable literature concerning the establishment of conditions under which relevant subsets do or do not exist (c.f. Buehler (1959), Wallace (1959), Stein (1961), Pierce (1973), Bondar (1977), Robinson (1976, 1979a, 1979b), and Pedersen (1978)). Though interesting, a study of these issues would take us too far afield. (See Section 3.7.3 for some related material, however.) Also, much of the theory is still based on frequentist (though partly conditional) measures, and hence violates the LP. Of course, many researchers in the field study the issue solely to point out inadequacies in the frequentist viewpoint, and not to recommend specific conditional frequentist measures. Indeed, it is fairly clear that the existence of relevant subsets, such as in Example 6, is not necessarily a problem, since when viewed completely conditionally (say from a Bayesian viewpoint conditioned on the data $(\bar{x},s)$), the interval $C(\bar{x},s)$ is very reasonable. Thus the existence of relevant subsets mainly points to a need to think carefully about conditioning.

2.3 ANCILLARITY

The most common type of partial conditioning advocated in statistics is conditioning on an ancillary statistic. An *ancillary statistic*, as introduced by Fisher (see Fisher (1956a) for discussion and earlier references), is a statistic whose distribution is independent of θ. (For a definition when nuisance parameters are present, see Section 3.5.5.) Thus, in Example 1, $S = |X_1-X_2|$ is an ancillary statistic which, when conditioned upon, gives "conditional confidence" for $C(X)$ of 100% or 50% as s is 1 or 0, respectively. And, in Example 2, the outcome of the coin flip is an ancillary statistic. The following is a more interesting example.

EXAMPLE 7. Suppose $X_1,\ldots,X_n$ are i.i.d. Uniform $(\theta - \frac{1}{2}, \theta + \frac{1}{2})$. Then $T = (U,V) = (\min X_i, \max X_i)$ is a sufficient statistic, and $S = V-U$ is an ancillary statistic (having a distribution clearly independent of θ). The conditional distribution of T given S = s is uniform on the set

$$\mathcal{X}_s = \{(u,v): \ v-u = s \text{ and } \theta - \frac{1}{2} < u < \theta + \frac{1}{2} - s\}.$$

Inference with respect to this conditional distribution is straightforward. For instance, a $100(1-\alpha)\%$ (conditional) confidence interval for θ is

$$C(U,V) = \frac{1}{2}(U+V) \pm \frac{1}{2}(1-\alpha)(1-s),$$

one of the solutions proposed by Welch (1939). This conditional interval is considerably more appealing than various "optimal" nonconditional intervals, as discussed in Pratt (1961).

There are a number of difficulties in the definition and use of ancillary statistics (c.f. Basu (1964) and Cox (1971)). Nevertheless, conditioning on ancillaries goes a long way towards providing better conditional procedures. A few references, from which others can be obtained, are Fisher (1956a), Anderson (1973), Barnard (1974), Cox and Hinkley (1974), Cox (1975), Dawid (1975, 1981), Efron and Hinkley (1978), Barndorff-Nielsen (1978, 1980), Hinkley (1978, 1980), Seidenfeld (1979), Grambsch (1980), Amari (1982), Barnett (1982), and Buehler (1982).

2.4 CONDITIONAL FREQUENTIST PROCEDURES

An ambitious attempt to formalize conditioning within a frequentist framework was undertaken by Kiefer (1977). (See also Kiefer (1975, 1976), Brown (1977), Brownie and Kiefer (1977), and Berger (1984c, 1984d).) The formalization was in two distinct directions, which Kiefer called conditional confidence and estimated confidence.

2.4.1 Conditional Confidence

The basic idea of conditional confidence is to define a partition $\{\mathcal{X}_s: \ s \in \mathcal{S}\}$ of $\mathcal{X}$ (the sets in the partition are the relevant subsets of $\mathcal{X}$), and then associate with each set in the partition the appropriate conditional frequency measure for the procedure considered. In Example 1, the partition would be into the sets $\mathcal{X}_1 = \{x: \ x_1 = x_2\}$ and $\mathcal{X}_2 = \{x: \ x_1 \neq x_2\}$. In

Example 2, the partition would be into the sets where heads and tails are observed, respectively.

When dealing with a confidence procedure $\{C(X)\}$, the conditional frequency measure that would be reported, if $x \in \mathcal{X}_s$ were observed, is

$$\Gamma_s(\theta) = P_\theta(C(X) \text{ contains } \theta | X \in \mathcal{X}_s).$$

EXAMPLE 7 (continued). Let the partition be $\{\mathcal{X}_s: 0 \le s \le 1\}$ (see Example 7). Then, for the procedure $\{C(U,V)\}$,

$$\Gamma_s(\theta) = P_\theta(C(U,V) \text{ contains } \theta | (U,V) \in \mathcal{X}_s) \equiv 1-\alpha.$$

In Examples 1, 2, and 7 it is relatively clear what to condition on. In Examples 3, 4, 5, and 6, however, there is no clear choice of a partition. In a situation such as Example 3, the following choice is attractive.

EXAMPLE 3 (continued). Let $\mathcal{X}_s = \{-s,s\}$ (i.e., the two points s and -s) for $s > 0$. (We will ignore $x = 0$, since it has zero probability of occurring.) The "natural" measure of conditional confidence in a testing situation is the conditional error probability function, determined here by

$$(2.4.1) \qquad \Gamma_s(1) = \Gamma_s(-1) = P_{-1}(\text{Rejecting}|\mathcal{X}_s)$$
$$= \frac{P_{-1}(X=s)}{P_{-1}(X=s)+P_{-1}(X=-s)}$$
$$= 1/(1+e^{4s}).$$

One would thus report the test outcome along with the conditional error probability $(1+e^{4|x|})^{-1}$. This conditional error probability has the appealing property of being close to 1/2 if $|x|$ is near zero, while being very small if $|x|$ is large. Thus it satisfies Pratt's "Principle of Adequacy."

The reason (from a frequency viewpoint) for formally introducing a partition is to prevent such "abuses" as conditioning on "favorable" relevant subsets, but ignoring unfavorable ones and presenting the unconditional measure when x is in an unfavorable relevant subset.

2.4.2 Estimated Confidence

An alternative approach to conditioning, which can be justified from a frequentist perspective (c.f. Kiefer (1977) or Berger (1984c)), is to present a data dependent confidence function. If a confidence set procedure $C(x)$ is to be used, for instance, one could report $1-\alpha(x)$ as the "confidence" in $C(x)$ when x is observed. Providing

$$(2.4.2) \qquad E_\theta(1-\alpha(X)) \leq P_\theta(C(X) \text{ contains } \theta) \quad \text{for all } \theta,$$

this "report" has the usual frequentist validity that, in repeated use, $C(X)$ will contain θ with at least the average of the reported confidences. Thus, in Example 2, one could report $1-\alpha(x) = 1$ or $\frac{1}{2}$ as $x_1 \neq x_2$ or $x_1 = x_2$, respectively. Estimated confidence (or, more generally, estimated risk) can be very useful in a number of situations where conditional confidence fails (see Kiefer (1977) or Berger (1984c)).

2.5 CRITICISMS OF PARTIAL CONDITIONING

The need to at least sometimes condition seems to be well recognized, as the brief review in this chapter has indicated. The approaches discussed in Sections 2.2, 2.3, and 2.4.1 consider only partial conditioning, however; one still does a frequency analysis, but with the conditional distribution of X on a subset. There are several major criticisms of such partial conditioning. (The estimated confidence approach in Section 2.4.2 has a quite different basis; criticism of it will be given at the end of this section.)

First, the choice of a relevant subset or an ancillary statistic or a partition $\{\mathcal{X}_s: s \in \mathcal{S}\}$ can be very uncertain. Indeed, it seems fairly clear that it is hard to argue philosophically that one should condition on a certain set or partition, but not on a subset or subpartition. (After all, it seems somewhat strange to observe x, note that it is in, say, $\mathcal{X}_s$, and then forget about x and pretend only that $\mathcal{X}_s$ is known to have obtained.) Researchers working with ancillarity attempt to define "good" ancillary statistics to condition upon, but, as mentioned earlier, there appear to be no completely satisfactory definitions. Also, ancillary statistics do not exist in many situations where it seems important to condition, as the following simple example shows.

EXAMPLE 8. Suppose $\Theta = [0, \frac{1}{2})$, and

$$X = \begin{cases} \theta & \text{with probability } 1-\theta \\ 0 & \text{with probability } \theta. \end{cases}$$

(An instrument measures θ exactly, but will erroneously give a zero reading with probability equal to θ.) Consider the confidence procedure $C(x) = \{x\}$ (the point x). Here $P_\theta(C(X)$ contains $\theta) = 1-\theta$. It is clear, however, that one wants to condition on $\{x:\ x > 0\}$, since $C(x) = \{\theta\}$ for sure if $x > 0$. But there is no ancillary statistic which provides such a conditioning.

In situations such as Examples 3, 4, 5, and 6, the selection of a partition for a conditional confidence analysis seems quite arbitrary. Kiefer (1977) simply says that the choice of a partition must ultimately be left to the user, although he does give certain guidelines. The development of intuition or theory for the choice of a partition seems very hard, however (see also Kiefer (1976), Brown (1977), and Berger (1984c)).

Even more disturbing are examples, such as Example 4(b), where it seems impossible to perform the indicated sensible test and report conditional error probabilities reflecting the true uncertainty when $x = 1$ is observed. (A three point $\mathcal{X}$ cannot be partitioned into two nondegenerate sets, and on a degenerate set the conditional error probability must be zero or one.) Any theory which cannot handle such a simple example is certainly suspect.

The situation for estimated confidence theory is more ambiguous, because it has not been very extensively studied. In particular, the choice of a particular estimated confidence or risk is very difficult, in all but the simplest situations. And, in situations such as Examples 3 and 4(b), estimated confidence functions will have certain undesirable properties. In Example 3, for instance, any estimated error probability, $\alpha(x)$, which is decreasing in $|x|$ and satisfies the frequentist validity criterion (similar to (2.4.2))

$$E_\theta\ \alpha(X) \geq P_\theta(\text{Test is in error}) \qquad \text{for all } \theta,$$

must have $\alpha(0) > \frac{1}{2}$ (since $P_{\frac{1}{2}}$ (Test is in error) $= \frac{1}{2}$). It seems strange, however, to report an error larger than $\frac{1}{2}$ (which could, intuitively, be obtained by simple guessing). For more extensive discussion of estimated confidence, see Berger (1984c).

The final argument against partial conditioning is the already alluded to fact that the most clearcut and "obvious" form of conditioning (Example 2) implies (together with sufficiency) the LP, which states that complete conditioning (down to x itself) is necessary. Since this would eliminate the possible application of frequency measures, new measures of evidence would clearly be called for.

It should be mentioned that certain other forms of statistical inference are very conditional in nature, such as fiducial inference developed by Fisher (see Hacking (1965), Plackett (1966), Wilkinson (1977), Pedersen (1978), and Dawid and Stone (1982) for theory and criticisms), structural inference developed by Fraser (c.f. Fraser (1968, 1972, 1979)), and pivotal inference developed by Barnard (c.f. Barnard (1980, 1981) and Barnard and Sprott (1983)). (Barnett (1982) gives a good introduction to all of these approaches.) The similarities among these methods (and also "objective Bayesian" analysis and frequentist "invariance" analysis) are considerable, but the motivations can be quite different. These methods rarely result in unreasonable conclusions from a conditional viewpoint, and hence do have many useful implications for conditional analysis. Space precludes extensive discussion of these approaches. (Some discussion of structural and pivotal analysis will be given in Sections 3.6 and 3.7, in the course of answering a specific criticism of the LP.) Suffice it to say that they are based on "intuitive" principles which can be at odds with the LP (and Bayesian analysis), and hence leave us doubting their ultimate truth.