Chapter 12
Conditionality

12.1 Ancillarity

Conditioning arguments are at the center of many disputes regarding the foundations of statistical inference. We present here only some simple arguments and examples.

Definition 12.1.1. A statistic A is **ancillary** (for θ) if its distribution does not depend on θ.

The basic idea in conditioning is that, since A provides no information on θ, we should use the conditional distribution, given A, for inference on θ. The idea originated with R.A. Fisher and has been discussed and disputed for decades. In some problems most statisticians condition on A, in other problems they do not.

In most problems the sample size is considered fixed, i.e., ancillary even though it may be determined by availability of funds or other considerations not related to the problem (θ) of interest. Similarly in regression type problems (linear models, generalized linear models, etc.) most statisticians condition on the covariates (design matrix). There seems to be no definite guidelines for when to condition and when not to condition.

Example. Cox introduced the following example. Consider two measuring devices. Device P produces measurements which are normal with mean θ and variance σ^2 and device I produces measurements which are normal with variance $k^2\sigma^2$ where k is much larger than 1. Which instrument is used is decided by the flip of a fair coin so that the precision of the measurement (i.e., what instrument is used) is ancillary.

Thus we would report the value of the measurement and the associated value of precision σ^2 or $k^2\sigma^2$ depending on the instrument **actually** used. However, if we do not condition, the true variance of X is

$$\mathbb{V}(X) = \mathbb{E}\left[\mathbb{V}(X|F)\right] + \mathbb{V}\left[\mathbb{E}(X|F)\right]$$
$$= \frac{\sigma^2}{2} + \frac{k^2\sigma^2}{2}$$

Note that

$$\sigma^2 < \sigma^2 \left(\frac{1}{2} + \frac{k^2}{2}\right) < k^2\sigma^2$$

so that the reported standard error will be either too small or too large.

Example (Valliant, Dorfman and Royall). There is a population of size 1,000 from which we have selected a random sample of size 100 without replacement. The population mean is estimated by the sample mean which has variance estimated by

$$\mathbb{V}(\overline{Y}_s) = \left(1 - \frac{100}{1000}\right)\frac{s^2}{100} \quad \text{where} \quad s^2 = \frac{\sum_{i \in s}(y_i - \overline{y}_s)^2}{99}$$

and s denotes the set of items selected.

Before we drew the sample, we considered doing a complete census of all 1,000 objects, but we had another study of interest. To decide whether to do the complete census or a sample of size 100 and the other study we flipped a coin. If the result was a head we did the complete census; if the result was a tail we took the sample of size 100.

The variance of the sample mean is

$$\mathbb{V}(\overline{Y}_s) = \frac{1}{2}\mathbb{V}(\overline{Y}_s|n = 100) + \frac{1}{2}\mathbb{V}(\overline{Y}_s|n = 1000) = \frac{1}{2}\mathbb{V}(\overline{Y}_s|n = 100)$$

Using this an estimate of variability is clearly wrong, yet it is correct from a frequentist point of view. Note that the same variance would be required if we had done the complete census. In this case any confidence interval would consist of a set of points whereas we know the population mean exactly! Clearly there is need for conditioning in situations like this.

12.2   Problems with Conditioning

Examples in the previous section indicate that we should condition whenever there is an ancillary statistic. Unfortunately this is not always so easy. An excellent review article by Ghosh et al. [19] provides many examples and extensions. In particular there are examples given where there is no unique ancillary statistic.

Some authors have suggested that there are really two major types of ancillarity:

1. Experimental
2. Mathematical

Experimental ancillaries are those such as sample size, covariates, etc., i.e., situations where most statisticians routinely condition. Mathematical ancillaries are those that arise because of the specific nature of the statistical model.

Example (Continuous uniform). Let $X_1, X_2, \ldots, X_n$ be iid with pdf

$$f(x; \theta_1, \theta_2) = \begin{cases} \frac{1}{\Delta} & \theta_1 \leq x \leq \theta_2 \\ 0 & \text{elsewhere} \end{cases} \tag{12.1}$$

where $\Delta = \theta_2 - \theta_1$.

The joint density is given by

$$f(x_1, x_2, \ldots, x_n \; ; \; \theta_1, \theta_2) = \begin{cases} \frac{1}{\Delta^n} & \text{all } x_i \in [\theta_1, \theta_2] \\ 0 & \text{elsewhere} \end{cases} \tag{12.2}$$

It follows that the minimum and maximum of $X_1, X_2, \ldots, X_n$ are minimal sufficient statistics for θ_1 and θ_2.

The joint distribution of the minimum and maximum from a random sample with distribution function F and density function f is easily shown to be

$$f(y_1, y_n) = n(n-1)[F(y_n) - F(y_1)]^{n-2} f(y_1) f(y_n) \tag{12.3}$$

where Y_1 is the minimum of the X_i's and Y_n is the maximum

For the uniform distribution, we have that

$$F(y; \theta_1, \theta_2) = \frac{1}{\Delta} \int_{\theta_1}^{y} dx = \frac{y - \theta_1}{\Delta}$$

so that the joint pdf of Y_1 and Y_n is given by

$$f(y_1, y_n; \theta_1, \theta_2) = \frac{1}{\Delta^n} n(n-1)(y_n - y_1)^{n-2} \quad \theta_1 \leq y_1 \leq y_n \leq \theta_2$$

Let $\theta_1 = \theta - \rho$ and $\theta_2 = \theta + \rho$, then we have that $\Delta = 2\rho$ and hence the joint density is

$$f(y_1, y_n; \theta) = \frac{n(n-1)(y_n - y_1)^{n-2}}{\rho^n} \quad ; \quad \theta - \rho \leq y_1 \leq y_n \leq \theta + \rho$$

If we assume that ρ is known, then the likelihood function for θ is

$$\mathcal{L}(\theta) = 1 \;\; ; \;\; y_n - \rho \le \theta \le y_1 + \rho$$

For the special case where $\rho = 1/2$ it is easy to show that $[Y_1, Y_n]$ is a $100\left(1 - \frac{1}{2^{n-1}}\right)$ confidence interval for θ.

Suppose now that

$$n = 5 \quad \text{and} \quad y_1 = 0.01, \; y_n = 0.99$$

Then the $100(1 - \frac{1}{16})\% = 93.75\%$ confidence interval for θ is .01 to .99. But since

$$y_1 \ge \theta - \frac{1}{2} \; ; \; y_n \le \theta + \frac{1}{2}$$

if and only if

$$0.51 = 0.01 + 0.5 \ge \theta \;\; \text{and} \;\; 0.49 = 0.99 - 0.5 \le \theta$$

with certainty.

Thus with these observed values of y_1 and y_n we are certain that

$$0.49 \le \theta \le 0.51$$

and yet our 93.75 % confidence interval is

$$0.01 \le \theta \le 0.99$$

This is silly.

As Cox points out it is imperative to condition on the ancillary statistic in this example which is the range $R = Y_n - Y_1$.