The Neyman-Pearson lemma for confidence distributions

Tore Schweder
University of Oslo
Box 1095 Blindern
0317 Oslo, Norway
tore.schweder@econ.uio.no

Nils Lid Hjort
University of Oslo
Box 1053 Blindern
0316 Oslo, Norway
nils@math.uio.no

Confidence distributions

Confidence intervals and p-values are the main formats of statistical reporting in the frequentist tradition. R.A. Fisher's fiducial distribution is less often used. Jerzy Neyman did not like fiducial distributions since in his view they are not proper probability distributions. They do not refer to a sample space and a stochastic experiment. Fisher, on the other hand, did not like Neyman's interpretation of the confidence interval as a stochastic interval with a controlled probability of covering the true value of the parameter.

A confidence distribution is a stochastic distribution (it depends on the stochastic data) with the property that its quantiles span confidence intervals. If y is the one-dimensional parameter and $C(y)$ is the cumulative distribution function, then $\left(C^{-1}(a), C^{-1}(b)\right)$ is a confidence interval with degree $b-a$. In other words, if y_0 is the true value of the parameter, $C\left(y_0\right)$ is a stochastic variable with a uniform distribution. Confidence distributions are usually constructed via pivots. For example, when $\hat{y}$ and s are the mean and standard deviation in a normal sample of size n, $\left(y-\hat{y}\right)/\left(s/\sqrt{n}\right)$ is a t-distributed pivot. Thus, with T_{n-1} being the cumulative t-distribution, $C(y)=T_{n-1}\left(\sqrt{n}\left(y-\hat{y}\right)/s\right)$ is a cumulative confidence distribution for y. Before data are obtained, C is stochastic. After the data are obtained, observed values of $\hat{y}$ and s are inserted, and C is a cumulative distribution function for y. Confidence intervals and p-values are related, and they are both obtained from the confidence distribution. The p-value of $H_0: y \le y_0$ *vs.* $H_1: y > y_0$ is simply $C\left(y_0\right)$. Schweder and Hjort (2001) discuss confidence distributions in more detail.

Efron (1993) does also discuss confidence distribution. In simple models, the confidence distribution is identical to the fiducial distribution. Since Neyman objected to the fiducial distribution, he would probably not have liked the concept of confidence distribution. And Fisher, on the other hand, disliked Neyman's concept of confidence, and would probably also have hated the confidence distribution.

Nevertheless, the confidence distribution is a very useful tool in statistical reporting, and should be a competitive frequentist analogue of the Bayesian posterior distribution. After the data have been observed, the confidence distribution is a distribution of confidence and not of probability in the frequentist interpretation. Before data are observed, it is a stochastic element with a probability distribution.

The Neyman-Pearson lemma

The power function is the primary performance characteristic in hypothesis testing. In confidence interval inference, one correspondingly uses the power function defined as the probability of the confidence interval not to include an alternative value of the parameter. What should the performance characteristic be for a confidence distribution?

Since the confidence distribution by design is unbiased (the confidence median is a median-unbiased point estimator), it is reasonable to judge the confidence distribution by its spread. The smaller it's spread the better. Most measures of spread in the distribution C around y_0 can be represented by a functional

$$g(C) = \int_{-\infty}^{0} C(y + y_0) g(dy) + \int_{0}^{\infty} \left(1 - C(y + y_0)\right) g(dy),$$

where g is a measure on the right hand side of the equation. If this measure is chosen as the Lebegue measure, $g(C)$ is the mean absolute variation around y_0 with respect to the distribution C. Other natural choices are: $g(dy) = |y|\, dy$ making $g(C)$ the mean square variation; and point measure at $y_1 < 0$ making $g(C)$ the p-value when testing $H_0 : y \le y_0 + y_1$.

As in other frequentist inference, the performance should be measured *ex ante* data. Before data are observed, the confidence distribution is stochastic, and any measure of spread, $g(C)$, is thus *ex ante* a stochastic variable.

The Neyman-Pearson lemma for confidence distributions (Schweder and Hjort 2001)

If the confidence distribution C is a function of the data through a one-dimensional statistic that is sufficient and in which the likelihood ratio is increasing, then C is uniformly most powerful in the sense that for any other confidence distribution C' and for any choice of spread functional, $g(C)$ is *ex ante* stochastically less than $g(C')$ when the true parameter of the probability distribution is y_0.

The proof is quite simple, and is in the reference. An extension to multi-parameter exponential models is also given there.

REFERENCES

Efron, B. 1998. R.A. Fisher in the 21st century (with discussion). *Statistical Science*, 13:95-122.

Schweder, T and Hjort, N.L. 2001. Confidence and likelihood. *Scandinavian Journal of Statistics*. Accepted.