

Chapter 12; Why Is Inference Important?

Steve Elston

9/23/2020

DRAFT ONLY - sorry

“Critical tests of this kind may be called tests of significance, and when such tests are available we may discover whether a second sample is or is not significantly different from the first.”

Ronald A. Fisher (1925)

Inference In Rational Decison Process

Statistical inference can be thought of as the process of making rational decisions based on data. The inference process is based on analysis of probability distributions. Statistical inference can produce one or more of the following:

- A **point estimate** of a most likely value. The most likely value can be computed by several methods. The **maximum likelihood estimate (MLE)** is often used in frequentist statistics. Where as the **maximum (MAP)** estimate is used in Bayesian analysis.
- An interval as a measure of uncertainty. A **confidence interval** is used in frequentist statistics. Whereas, a **credible interval** is used in Bayesian analysis.

Confusingly, the term statistical inference is applied in many contexts. Some examples of applications of statistical inference include the following:

Inference on differences in distributions

In many practical cases it is important to know if samples are drawn from the same distribution or not. An **hypothesis test** is applied to determine if differences in distribution are **statistical** . A **null hypothesis** is the hypothesis that the distributions are the same. The test determines if we can **reject** the null hypothesis with a certain confidence. We will have much more to say about hypothesis testing in the remainder of this section of the book.

Inference for model parameters

Many statistical models, including machine learning models, have unknown parameters for which the values must be estimated. Computing estimates of these parameters is a form of statistical inference. Further, uncertainty of these parameter estimates are computed. A further inference is to determine which, of possibly many, parameters are significant in the model. We will return to this topic in the next section of this book.

Inference for prediction

Traditionally, the goals and methods used for inference and prediction were considered separately. In recent decades this distinction has blurred to the point of being irrelevant. Particularly in the machine learning era we speak of inference and prediction interchangeably.

Consider a common machine learning example, classification. Inference is used to determine the most likely class for each case of input values. Further, inference tools can be used to determine the confidence one should have in the result.

Another example from machine learning is prediction of numeric values; e.g. regression. In this case, the inference is for a numeric value given input values, or independent variables. In addition, the uncertainty of the prediction can be computed.

Inference Cannot Determine Importance

While doing inference we often speak of **statistical significance** one must never confuse this situation with actual importance. Statistical significance is a statement concerning probability. But, the importance of a relationship to the problem at hand can be an entirely different matter. Early workers in statistical inference recognized this fact. For example, Ronald A. Fisher (1932) warned:

“I believe that no one who is familiar, either with mathematical advances in other fields, or with the range of special biological conditions to be considered, would ever conceive that everything could be summed up in a single mathematical formula, however complex.”

So what use is statistical significance then? Many statisticians would say that a statistically significant result indicates a relationship is worthy of further investigation. Further investigation can take many forms. A few examples include:

- Gather more data, either observational or experimental.
- Find other variables which might illuminate the relationship under investigation.
- Consider the theoretical basis of the relationship. Can known science add understanding.

In summary, do not confuse statistical significance with actual importance. Statistical significance can help identify important relationships, but is not evidence of actual importance.

Confidence Intervals; the key to inference

In frequentist statistics uncertainty of an inference is expressed in terms of a **confidence interval**. A confidence interval is defined as the expected range of a parameter estimate. Freest, we will work with **two-sided confidence intervals**. In later sections, we will explore **one-sided confidence intervals**.

One way to understand confidence intervals is to look at the α and $1 - \alpha$ quantiles of a distribution. The confidence interval corresponds to the span of the distribution between these quantiles. In other words, α is the probability of that the value of a random variable, \mathbf{x} , will in the upper α quantile, or in the lower α quantile.

We can express a confidence interval for a random variable, \mathbf{x} , in terms of the probability as:

$$1 - 2\alpha = P(\text{Lower CI} \leq \mathbf{x} \leq \text{Upper CI})$$

An alternative view is to consider the probability of the observed values being outside of the confidence interval. These probabilities can be expressed in terms of the upper and lower quantiles of the confidence interval:

$$\alpha = P(x \geq \text{Upper CI}) \text{ and } 1 - \alpha = P(x \leq \text{Lower CI})$$

Example; confidence intervals of the Normal distribution

We will illustrate the concept of confidence intervals with an example. The **cumulative distribution function (CDF)** of a standard Normal distribution, $N(0,1)$, is shown in the plot generated by the code below. This code executes these steps:

- The cumulates of the CDF are computed. An anonymous function (a lambda) using the *cdf* method of *scipy.stats.norm* is used to compute the cumulates.
- The CDF is plotted along with horizontal lines showing the distribution within the range $\{\alpha, 1.0 - \alpha\}$.
- The roots of the functions $CDF(x) - \alpha$ and $CDF(x) - 1 + \alpha$ are found. These roots are the lower and upper confidence intervals respectively. Vertical lines are plotted at these values of the variable x . The aforementioned function for the CDF is used in the calculation.
- Double ended arrows with annotation are plotted to illustrate the confidence interval on the CDF.

```
def plot_confidence(distribution, alpha, xmin, xmax, dist_type, step=0.05):
    ## Set the font size and compute the CI for display
    plt.rc('font', size=20)
    percent = str(int((1.0-2*alpha)*100))

    ## first find the
    x = np.arange(start=xmin, stop=xmax, step=0.05)
    cumulates = distribution(x)

    ## plot the figure
    fig, ax = plt.subplots(figsize=(20, 16), )
    ax.plot(x, cumulates, linewidth=6)
    ax.hlines(y=0.0, xmin=xmax, xmax=xmin, linewidth=2, color='black')
    ax.set_title('Cumulative distribution of ' + dist_type + ' with ' + percent + ' confidence interval')
    ax.set_ylabel('Probability')
    ax.set_xlabel('Value')

    ## Plot horizontal lines at the quantiles
    ax.hlines(y=alpha, xmin=xmax, xmax=xmin, linewidth=6, color='r', linestyle='dotted')
    ax.hlines(y=1.0-alpha, xmin=xmax, xmax=xmin, linewidth=6, color='r', linestyle='dotted')

    ## Fine the probabilities at the quantiles and plot vertical lines
    ## To do so, find the root of a function of the distribution
    lower_ci = brentq(lambda x: distribution(x) - alpha, xmin, xmax)
    ax.vlines(x=lower_ci, ymin=-0.2, ymax=1.0, linewidth=4, color='r', linestyle='dashed')
    upper_ci = brentq(lambda x: distribution(x) - 1 + alpha, xmin, xmax)
    ax.vlines(x=upper_ci, ymin=-0.2, ymax=1.0, linewidth=4, color='r', linestyle='dashed')

    ## Display the numeric results
    print("Confidence Interval, lower: {0:5.2f}, upper: {1:5.2f}".format(lower_ci, upper_ci))

    ## Place double headed arrows for the range and confidence interval
    ax.arrow(xmin, alpha, 0.0, 1-2*alpha-0.05, head_width=0.2, head_length=0.05, linewidth=3)
```

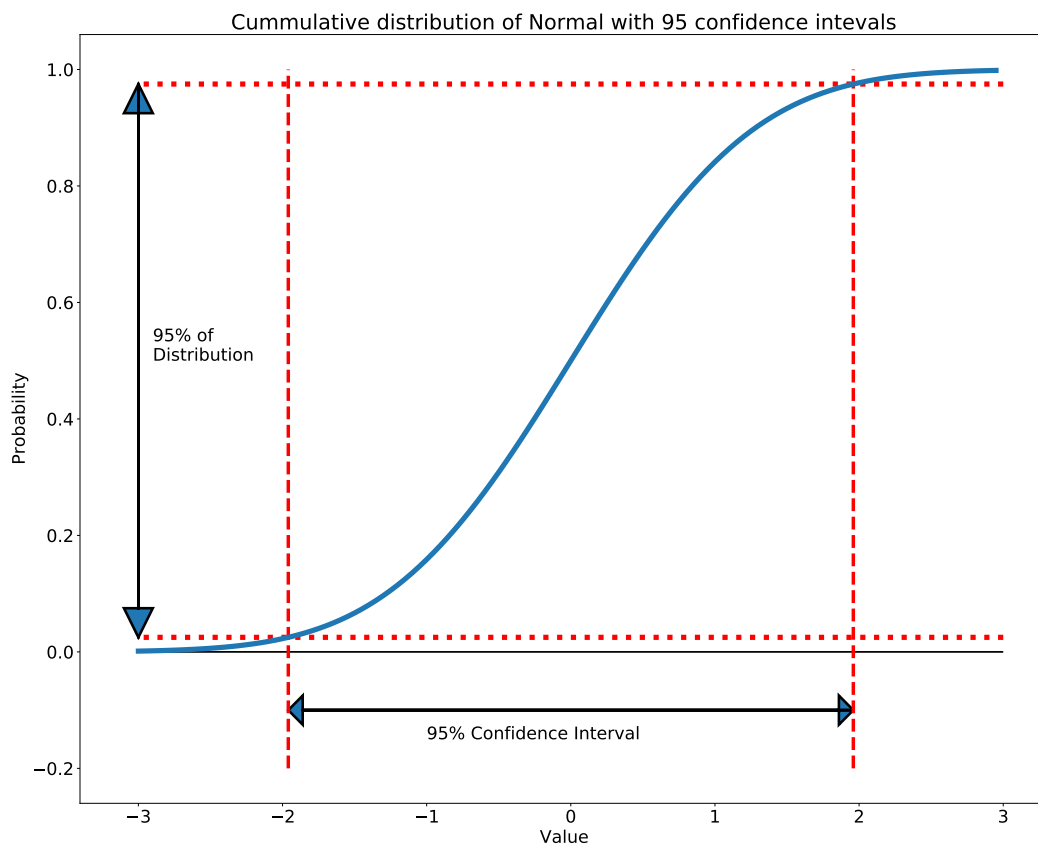
```

ax.arrow(xmin,1.0-alpha,0.0,-1+2*alpha+0.05, head_width=0.2, head_length=0.05, linewidth=3)
ax.text(xmin+0.1, 0.5, percent + '% of\nDistribution')
ax.arrow(lower_ci,-0.1,upper_ci-lower_ci-0.1, 0, head_width=0.05, head_length=0.1, linewidth=3)
ax.arrow(upper_ci,-0.1,-upper_ci+lower_ci+0.1, 0, head_width=0.05, head_length=0.1, linewidth=3)
ax.text((upper_ci + lower_ci)/2-1.0, -0.15, percent + '% Confidence Interval')
plt.show()

distribution = norm.cdf
plot_confidence(distribution, 0.025, -3.0, 3.0, 'Normal')

## Confidence Interval, lower: -1.96, upper: 1.96

```



The chart shows the range of the 95% confidence interval for the CDF of the standard Normal distribution, $N(0, 1)$. This means that for every 20 draws we perform from this distribution we expect 19 of them to be in the range of this confidence interval. On the other hand, we must keep in mind that we expect that 1 time in 20 a draw will be outside this range. Do not be fooled by this randomness!

The 95% confidence interval is the most widely used, but there are many other possible choices. The confidence interval chosen should fit the problem at hand. In other words, what measure of uncertainty is appropriate? If one needs to have greater certainty that a value falls within a confidence interval a large value should be chosen; for example 99% or 0.99. But perhaps, the need for certainty is lower, so a smaller value can be chosen; for example 90% or 0.9.

The foregoing example is specific to a Normal distribution, However, the concept of confidence interval is commonly other distributions. For example, recall that the sample variance of a Normal distribution is χ^2 distributed. It is generally useful to have an estimate of the confidence interval for this estimate.

Do confidence intervals always exist as useful values? Quite simply, no. Some distributions have infinite variance. In this case the confidence intervals are not bounded. However, in practice, this problem is not as bad as it sounds. In data analysis and computational statistics we are always dealing with finite samples. Therefore the confidence intervals we compute are based on empirical estimates of statistics. The confidence interval will exist for nearly all empirical statistical estimates.

Note: There is not general agreement on how α is specified. In this book, we use the more common convention of specifying *alpha* as the quantile defining the limit of the confidence interval. This is why we say the probability of a random variable being within a two-sided confidence interval is $1 - 2\alpha$. You will see some texts and even statistical software packages that use a different convention. In this case the probability of a random variable being within the confidence interval is $1 - \alpha$. The upper and lower quantiles are then at $\pm\alpha/2$. Make sure you are certain which convention the software you are using conforms with.

Exercise: From the foregoing discussion is is clear that the value of alpha can be in the range $0 < \alpha < 1$. In the example, we explored the 95% confidence interval for the standard Normal distribution with *alpha* = $0.05/2 = 0.025$. Now, make similar plots for the 90% and 99% confidence intervals. Does it make sense that the confidence interval becomes wider as we reduce the probability that the value of the random variable will fall outside the confidence interval?

Exercise: As has been stated above, confidence intervals can be computed for most any distribution, not just the Normal distribution. Now, make plots similar to the example, but for the χ^2 distribution. Use 3 and 10 degrees of freedom to find and display the 90%, 95% and 99% confidence intervals. Create an function (a lambda) using the `scipy.stats.chisquared.cdf` function to compute values of the CDF. Use an x-axis range of $0 \leq x \leq 30$. Does it make sense that the confidence interval becomes wider as we increase the degrees of freedom?

One-Sided Confidence Intervals

So far, we have considered two sided confidence intervals. But in many cases, we are interested in one side of the distribution or the other. We then work with a **one-sided confidence interval**.

There are many situations where a one-sided confidence interval is useful. Often we want to know the maximum or minimum value of a random variable with a certain probability. For example, a financial portfolio manager will want to know the maximum losses 99% of the time. This statistic is known as **value at risk VAR**.

When defining one sided confidence intervals the bounds of the random variable must be considered. Take the example of a distribution with possible range of values, $-\infty \leq x \leq \infty$, such as the Normal distribution. If the upper tail of the distribution is of interest, the confidence interval can be defined from the lower limit:

$$1 - \alpha = P(-\infty \leq x \leq \alpha)$$

A one-sided confidence can also be defined for the lower tail of a random variable. In this case, the confidence interval is defined:

$$\alpha = P(-\alpha \leq x \leq \infty)$$

Now consider a distribution with possible range of values $0 \leq x \leq \infty$. Examples include the one-side Normal distribution and the χ^2 distribution. In this case we can define the confidence interval as:

$$1 - \alpha = P(0 \leq x \leq \alpha)$$

Example, Inference for the mean

Let's consider an example of confidence intervals for the mean. We have previously seen how the maximum likelihood estimate of the mean is Normally distributed. This fact makes computing the confidence intervals straight forward using the following relationship:

$$CI(mean) = MLR(\theta|\mathbf{X}) \pm \frac{s}{\sqrt{n}}Z_\alpha$$

Here, Z_α is the standard Normal distribution with evaluated at the confidence level, α . The standard error is given by s/\sqrt{n} , where s is the standard deviation and n is the number of samples.

Keep in mind that the Normal distribution of the maximum likelihood estimate is an asymptotic; $n \rightarrow \infty$. For finite samples, particularly small samples, this relationship may be approximate. Further, the estimate of s will have considerable uncertainty.

Exercise: To help develop understanding of the behavior of confidence intervals for a mean estimate with different sample sizes you will do the following. For Bernoulli sample sizes of 10, 20, 50, 100, 300, 1000, 2000, 5000, 10000 of the log of King County housing prices you will compute and plot the mean estimate with the confidence intervals. Follow these steps:

- 1.

Example, Inference for the variance

As another example we will consider the confidence intervals of the variance estimator. From the log-likelihood, we have seen that variance is more difficult to estimate precisely than the mean. The confidence intervals of the variance depend on the χ^2 distribution.

Copyright 2020, 2021, Stephen F. Elston. All rights reserved.

Fisher, Ronald A. 1925. *Statistical Methods for Research Workers*. First Edition. Oliver; Boyd.

"The Evolutionary Modification of Genetic Phenomena." 1932.