Model Based Statistics in Biology.

Part II. Quantifying Uncertainty.
Chapter 7.5   Confidence Limits

ReCap.   Part I (Chapters 1,2,3,4)
ReCap   Part II (Ch 5, 6)
7.0 Inferential Statistics
7.1 The Logic of Hypothesis Testing
7.2 Hypothesis Testing with an Empirical Distribution
7.3 Hypothesis Testing with Cumulative Distribution Functions
7.4 Parameter Estimates
7.5 Confidence Limits

The truth is out there.
We're going to surround it.

Anonymous student. 2011.

<div align="center">on chalk board</div>

ReCap Part I (Chapters 1,2,3,4)
Quantitative reasoning: Example of scallops, which combined
models (what is the relation of scallop density to substrate?)
with statistics (how certain can we be?)
ReCap (Ch5)
Data equations summarize pattern in data as a series of parameters (means, slopes).
ReCap (Ch 6)
Frequency distributions are a key concept in statistics.
They are used to quantify the strength of evidence (likelihood ratios),
They are used to quantify uncertainty (p-values_
Empirical distributions are constructed from data.
Theoretical distributions are models of data.
ReCap (Ch 7)
Inferential statistics are a logical procedure for making decisions when there is
uncertainty due to variable outcomes.
<u>Hypothesis testing</u> is concerned with making a decision about an unknown population
parameter.
<u>Estimation</u> is concerned with the specific value of an unknown population parameter.

Today: Cumulative distributions to compute confidence limits on estimates

Wrap-up
We used cumulative distributions to compute confidence limits, a measure of the
reliability of an estimate.

Confidence Intervals and Conifidence Limits

We will use confidence limits to evaluate the uncertainty on an estimate made from data.

Definition. There are several definitions of a confidence interval. A Confidence Interval consists of two values that bracket the true value of a statistic, at some specified level of confidence (say, 95%). The idea of using an interval to quantify uncertainty is due to Thomas Bayes (1702 – 1761) an Nonconformist minister and mathematician. Bayes proposed and then solved the following problem.

<div align="center">PROBLEM</div>

Given the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.
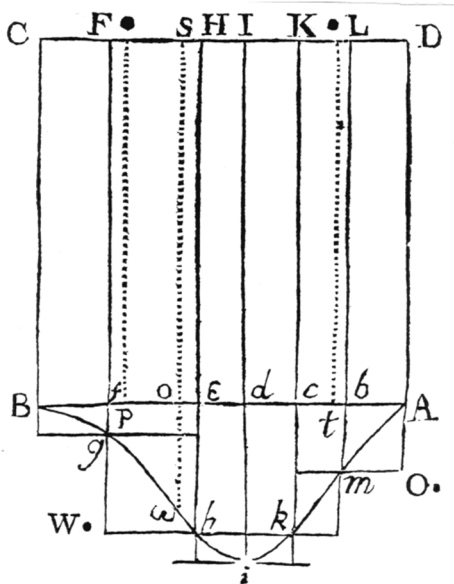
To solve the problem Bayes described a "square table made level" on which a ball W is first thrown, then a ball O is thrown repeatedly. The proportion of times that the second ball passes to one side of ball W is then calculated. A binomial distribution is used to calculate Bayes' Rule 1, that "my guess that the probability of its happening in a single trial lies somewhere between any two degrees of probability" is right. Bayes' displayed a diagram to construct a geometric proof for the cumulative distribution between any two degrees of probability.

While this is a confidence interval by intent, it is not exactly the Neyman type probability interval in use today. The Neyman 95% CI is the interval that will contain the true value on 95% of occasions if a study were repeated many times using samples from the same population. Bayes' publication makes no mention of an unknown 'true' value or of samples from an infinite number of repeats (the frequentist definition).

Bayes' rule for calculating the limit appears to be what is now called a credibility limit (fixed limits for a parameter with a random component). However, Bayes' rule when applied to a fixed object, such as the location of a star or the target of a measurement, might also be interpreted as a confidence limit (fixed parameter between random limits). Bayes' broad wording "between any two limits" does not specify how limits are chosen. Because most people interpret probability limits as being fixed, we might venture a guess that Bayes thought the same way. Binomial confidence limits were developed in the early 20th century (Wilson 1927, Clopper and Pearson 1934). This was then generalized to confidence limits for any distribution (Neyman 1937).

Having given Thomas Bayes due credit for the idea of probability interval, we move on to an example of the Neyman confidence interval, which is the interval that is presented in textbooks and appears in the published research.

Example -- Brook trout lengths

We measure a sample of 16 0-group trout from Cat Arm lake on the Great Northern Peninsula, in 1982, prior to flooding to create a reservoir.
0-group are fish less than 1 year old. First year fish were of interest in 1982 because one potential impact of flooding of Cat Arm lake was reduction in numbers or size of fish recruiting to the population. If there was an effect on size or numbers, then Newfoundland Hydro was committed to build a hatchery to mitigate the effects of flooding on this and other fish in the lake. The hatchery would be built by Newfoundland Hydro, at the expense of those who pay Hydro for electricity. Size was measured prior to flooding to establish a baseline for comparison to first year fish after flooding.

Quantity is fork length $Y =$ mm

Sample size is 16
Total population in the lake is *ca* 700 trout
Sampling was haphazard. Randomized sampling of the population was not possible.
The sampling fraction was 16/700, or approximately 2% of a finite population.

The sample mean is *mean*$(Y) = 53.8$ mm
This is an estimate of the true mean E(Y), which is unknown.

How reliable is our estimate of the mean? We'd like to know whether our estimate is close to the true value, the average length in the population of approximately 700 fish. We can't know the true mean, but we can make a statement about the reliability of our estimate, relative to the true value of the mean.

To make a statement about the reliability of our estimate, we compute a range that includes the true value a high percentage of the time. This is called the confidence limit. Here is a generic recipe for calculating confidence limit on any estimate.

Table 7.5a Generic recipe for calculating a confidence limit.

_____
1. State population; state the statistic of interest.
2. Calculate an estimate of the statistic from data
3. Determine the distribution of the estimate.
4. State tolerance for Type I error.
5. Write a probability statement about the estimate or statistic.
6. Plug values into the statement to obtain confidence limits.
7. Make a statement of degree of certainty about the confidence limit.
   This statement is not about the probability that the limit contains the true value.

Confidence Limits – Computational Flow

To compute the confidence limit, we fix some probability that we can live with, then make a probability statement about a line that includes the true value at a pre-stated level of confidence.
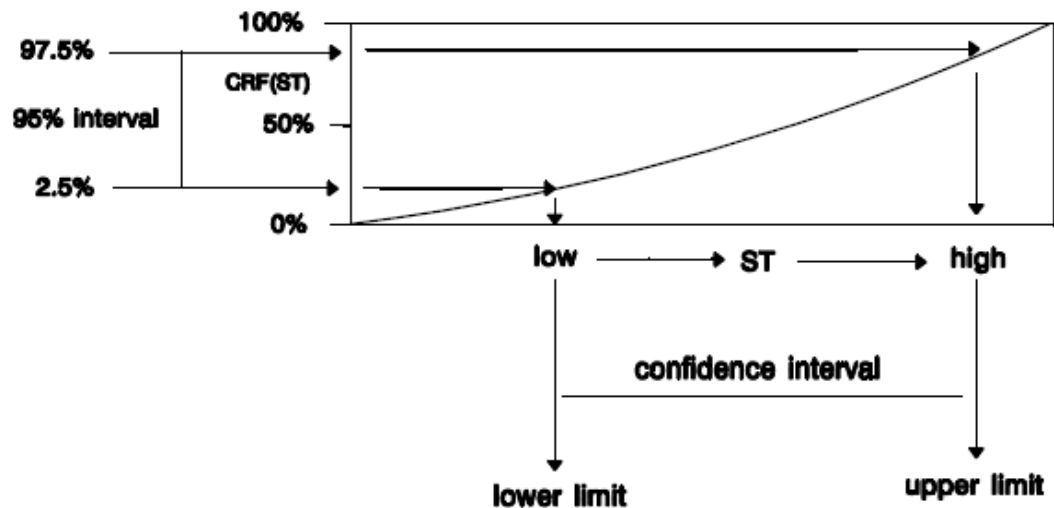
To compute a confidence limit, we will:
    obtain a statistic from data,
    obtain the distribution of this statistic
        (this is not the same as the distribution of the data),
    use the distribution of the statistic to compute confidence limits.

Graph shows this. Start with probability range to compute range of statistic.

```
As example, draw Chisquare distribution,
(labelled df = 4)
The is the distribution for the variance.
(Draw pdf)
Then draw the cdf above or below the pdf.
Here are the  Minitab computations for
text example p 155 (Sokal and Rohlf 1995)

MTB > invcdf 0.025;
SUBC> chisquare 4.
   0.025   0.484
MTB > invcdf 0.975;
SUBC> chisquare 4.
    0.975   11.1433
```

Confidence Limits –Key to choice of distribution for computing limit.

Table 7.2. Key for choosing the frequency distribution of a statistic.

Statistic is the population mean
    If data are normal or cluster around a central value
        If sample is large (n > 300) .... Normal distribution
        If sample is small (n < 300 .... t distribution
    If data are Poisson .... Poisson distribution
    If data are Binomial .... Binomial distribution
    If data do not cluster around central value, examine residuals (deviations
        from the mean)
    If residuals are normal or cluster around a central value
        If sample is large (n > 300) .... Normal distribution
        If sample is small (n < 300) .... t distribution
    If residuals are not normal .... Empirical (bootstrap)

Statistic is the population variance
    If data are normal or cluster around a central value .... Chi-square
    If data do not cluster around central value
        If sample is large (n > 30) .... Chi-square
        If sample is small (n < 30 .... Empirical (bootstrap)

Statistic is the ratio of two variances (ANOVA tables)
    If data are normal or cluster around a central value .... F-distribution
    If data do not cluster around a central value, calculate residuals
    If residuals are normal or cluster around a central value .... F-distribution
    If residuals do not cluster around central values
        If sample is large (n > 30) .... F-distribution
        If sample is small (n < 30) .... Empirical

Statistic is none of the above
    Search statistical literature for appropriate distribution
        or confer with statistician
    If not in literature or cannot be found .... Empirical

Empirical distributions are generated by taking all permutations, by sampling permutations, or by subsampling (bootstrap methods).

Confidence Limits from Generic Recipe -- Brook trout fork lengths

1. The population is all brook trout less than 1 year of age in Cat Arm Lake in 1982
   The statistic of interest is the population mean.

2. The mean for a sample of 16 fish was
   $mean(Y) = 53.8$ mm

3. The distribution of an estimate such as a mean will not be exactly the same as the distribution of data so we use a key (Table 7.2, above).

   The statistic is the population mean,
   the data cluster around a central value,
   and the sample size is small (16),
   so the appropriate distribution is the t-distribution.

4. The tolerance of Type I error will be set at 10%.
   This increases the tolerance for error, compared to the traditional 5% level.
   This increase reduces the risk of Type II error, failing to detect at real difference.

5. Now that we have a distribution, and a stated tolerance for error, we can write a probability statement.

   First in verbal form. "The line from L_1 to L_2 includes the true mean fork length μ_Y of CatArm brook trout with a probability of 90%"

   The probability statement is about $1 - \alpha$
   It is not about the probability that that true value lies between the limits.

   Next in graphical form (refer back to figure, limits on y axis projected to x-axis)
   The frequency distribution is used to go from probability to outcome.
   This is the opposite direction from that used in hypothesis testing,
   the frequency distribution is used to go from outcome to probability.

   Now, the same thing in symbolic form.

   $$P\{L_1 \leq \mu_Y \leq L_2\} = 1 - \alpha$$
   $$P\{\overline{Y} - s_{\overline{Y}} \cdot |t(\alpha/2, n-1)| \leq \mu_Y \leq \overline{Y} + s_{\overline{Y}} \cdot |t(\alpha/2, n-1)|\} = 1 - \alpha$$

   We'll examine each of these components in detail.

   For now, we note $|t(\alpha/2, n-1)|$ is the absolute value of the t statistic from the cdf,

   We subtract this quantity from $\overline{Y}$ to obtain the lower limit L_1.
   We add it to $\overline{Y}$ to obtain the upper limit L_2.

Confidence limits -- Brook trout fork lengths (continued)

5.   Write probability statement (continued)

$$P\left\{\overline{Y} - s_{\overline{Y}} \cdot \left|t(\alpha/2, n-1)\right| \le \mu_Y \le \overline{Y} + s_{\overline{Y}} \cdot t(\alpha/2, n-1)\right\} = 1 - \alpha$$

We use the sample mean $\overline{Y}$ to estimate of the true mean μ_Y

For the t-statistic we use the t-distribution, which is symmetrical around zero.
We use the t-distribution with df $= n-1$
We assign half of our tolerance for uncertainty ($\alpha/2$) to each tail (5% to each tail)
We use the inverse cumulative distribution function cdf to obtain the lower tail
      probability ($\alpha/2$) and the upper tail probability ($1-\alpha/2$) values for
      the t-distribution.
$s_{\overline{Y}}$ is the standard error of the mean.

   The standard error is estimated from the sample standard deviation of the data s_Y

$$s_{\overline{Y}} = \frac{s_Y}{\sqrt{n}} = \sqrt{\frac{\text{var}(Y)}{n}}$$

6.   Plug values into probability statement

$$\overline{Y} = 53.8 \text{ mm} \quad s_Y = 5.8 \text{ mm} \quad \alpha = 10\% \quad \alpha/2 = 5\% \quad t(0.05,15) = -1.753$$

```
MTB> invcdf 0.05;
   SUBC> t 15.
       0.05    -1.753

   MTB> invcdf 0.95;
   SUBC> t 15.
       0.95    +1.753
```

```
Most packages calculate the
upper and lower tails. Tables
for t distribution were usually
two-tailed, showing both tails
for positive values only.
```

P{ 53.8 − 1.753*(5.8 / sqrt(16)) $\le \mu_y \le$ 53.8 + 1.753*(5.8 / sqrt(16)) } = 90%

P{ 53.8 − 2.54 $\le \mu_y \le$ 53.8 + 2.54} = 90%

P{ 51.26 $\le \mu_y \le$ 56.34} = 90%

7. The limits 51.26 mm to 56.3 mm enclose the true population mean 90% of the time.
    The statement is about limits that enclose the true value of the mean.
    The statement is not about the sample mean $\overline{Y}$.
    The probability that the true mean falls within a range is the credibility interval.

Comments on confidence limits

How do we narrow $L_2 - L_1$?

   increase α This increase tolerance of Type I error

   increase n This reduces the standard error of the estimate

   decrease σ This accomplished by eliminating sources of error or unexplained variance.

Confidence limits use the inverse probability distribution.

   Hypothesis testing: go from outcome to probability

```
MTB> cdf 1.753;
SUBC> t 15.
    1.753      0.95
```

   Confidence limits: go from probability to outcome

```
MTB> invcdf 0.95;
SUBC> t 15.
    0.95 1.753
```

Confidence limits are particularly useful in excluding hypotheses other than just H_o

A confidence limit is not of much value if the sample is not representative.

We use the t distribution for the t-statistic, the F-distribution for F-ratios, and χ^2 distribution for statistics known be distributed as chi-square. For other statistics, we use an empirical distribution generated by randomization (*cf* Table 7.2). Or we can use bootstrap methods (Lab 11).

Extension
The fish measurements at Cat Arm Lake and subsequent reservoir become a monitoring program to evaluate the hypothesized loss of spawning habitat due to flooding of a lake into a reservoir. Once a distribution was established during the early stage of monitoring, credible intervals would have been a logical way to update the estimates of uncertainty on size of fish in the reservoir.

Binomial Confidence Limits -- partridge berry example

> This example illustrates asymmetrical limits for an example where the null hypothesis is of no interest.

> Ask students for examples from their work, where showing reliability (confidence limits) would be useful.

Partridgeberries, *Vaccinium vitis-idaea* L, grow on low plants in treeless areas in Newfoundland. Lingonberry fruitworm *Grapholita libertine* Heinrich is a tortricid moth whose larvae feed within the berry, causing it to go mushy when berries ripen. To estimate worm infestation rate a sample of 250 partridge berries were removed from a bucket of berries picked in September 1985 from the barrens above Memorial University's Logy Bay Marine lab.

One worm was found. How reliable is this estimate of the true rate of infestation ?
(*i.e.*, what are the confidence limits for this estimate?)

1. Population: all partridge berries above the Logy Bay lab in Sept 1989.
    This is a finite population that can be estimated by picking every berry in randomly placed plots within a larger area.
    The statistic is the infection rate.

2. Estimate of infection rate is $x/n = p = 1/250 = 0.004$ (4 per thousand)

3. Frequency distribution is binomial with $n = 250$ trials, $\pi = 0.004$.

4. Tolerance of Type I error: $\alpha = 10\%$.

5. Frequency is one infection, for which Minitab notation is $K = 1$.
    $P\{ L_1 \le K \le L_2 \} = 1 - \alpha$

6. Plug in $\alpha/2$, lower limit

```
MTB> invcdf 0.05;
SUBC> binomial n=250  p=0.004.
  K   P(X LESS OR = K)     K    P(X LESS OR = K)
  0        0.000          0        0.3671
```

There is not enough information to set a lower 90% limit. A larger sample is needed.

6. Plug in $1-\alpha/2$, upper limit

```
MTB> invcdf 0.95;
SUBC> Binomial 250 0.004.
   K   P(X LESS OR = K)      K    P(X LESS OR = K)
   2        0.9201      3          0.9813
```

7. $P\{ K > 2 \} = 1 - 0.9201 = 0.08$
    $P\{ K > 3 \} = 1 - 0.9813 = 0.0287$
    The upper 90% limit is closer to 3 than 2.

References.

Bayes, T. 1763. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society* 53: 370–418.

Clopper, CJ, ES Pearson. 1934. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 26: 404–413, https://doi.org/10.1093/biomet/26.4.404

Neyman, J. 1937. "Outline of a theory of statistical estimation based on the classical theory of probability". *Philosophical Transactions of the Royal Society* A. 236 (767): 333–380.

Wilson E.B. 1927. Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association* 22:158, 209-212, https://doi.org/10.1080/01621459.1927.10502953