MAXIMUM LIKELIHOOD METHOD
CONFIDENCE INTERVALS

MAXIMUM LIKELIHOOD METHOD

1. Eight trials are conducted of a given system with the following results: **S, F, S, F, S, S, S, S** (**S** – success; **F** – failure). What is the maximum likelihood estimate of p, the probability of successful operation?

   We have $L(p) = p^6(1-p)^2$. Differentiating with respect to p and setting derivative equal to zero: $2p^5[4p^2 - 7p + 3] = 0; \to p = 1$ *or* $3/4$. The $3/4$ maximizes the L.

2. Consider the geometric distribution: probability that an event A, having the probability $\mathcal{P} = p$ will occur after $x-1$ trials in which A does not occur:

   $$f(x;p) = p(1-p)^{x-1}; \qquad x = 1, 2, \ldots$$

   Suppose a series of n such experiments are carried out and let $x_1, x_2, \ldots, x_n$ denote the number of trials before A occurs in each experiment. Find an estimate for p using the method of maximum likelihood.

   $$L = \prod_{i=1}^{n} f(x_i;p) = p(1-p)^{x_1-1}p(1-p)^{x_2-1}\ldots p(1-p)^{x_n-1} = (1-p)^{\left[\sum_{i=1}^{n} x_i - n\right]} \cdot p^n.$$

   $$\ln L = \left[\sum_{i=1}^{n} x_i - n\right] \times \ln(1-p) + n \ln p.$$

   Taking the partial derivative with respect to p and setting the result equal to zero:

   $$\frac{\partial L(x_1, \ldots, x_n; p)}{\partial p} = -\frac{\sum_{i=1}^{n} x_i - n}{1-p} + \frac{n}{p} = \frac{-p\left[\sum_{i=1}^{n} x_i - n\right] + n(1-p)}{p(1-p)} = 0,$$

   which gives

   $$p = \frac{n}{\sum_{i=1}^{n} x_i}.$$

   Thus the maximum likelihood estimate of p is the reciprocal of the arithmetic mean of the # of trials in each experiment.

CONFIDENCE INTERVALS

1. The average zinc concentrations recovered from a sample of zinc measurements in 36 different locations on a river was found to be 2.6 grams per millilitre. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation σ is 0.3 grams per millilitre.

   Hint: The sample is big enough to use the appropriate quantiles of the standard normal distribution. Anyway, we have no other choice – we have no data to estimate the population standard deviation, so we consider σ as being known for good.

   The z-value leaving an area of 0.025 to the right and therefore an area of 0.975 to the left is $z_{0.025} = 1.96$. The 95% confidence interval is

   $$2.6 - (1.96)\left(\frac{0.3}{\sqrt{36}}\right) < \mu < 2.6 + (1.96)\left(\frac{0.3}{\sqrt{36}}\right)$$

   which reduces to $2.50 < \mu < 2.70$. For $1 - \alpha = 0.99$ we obtain $2.47 < \mu < 2.73$.

2. **Adequate sample size.** If $\bar{x}$ is used as an estimate of μ one can (roughly) say that: I am (1-α)100% confident that the error will not exceed $E = z_{1-\alpha/2}\sigma/\sqrt{n}$. So, reshaping this statement: If $\bar{x}$ is used as

an estimate of μ I am (1-α)100% confident that the error will not exceed a specified amount E when the sample size is:

$$n = \left(\frac{z_{1-\alpha/2}\sigma}{E}\right)^2.$$

Find how large a sample is required in the preceding example if we want to be 95% confident that our estimate of μ is off by less than 0.05.

Solving the above equation we have $n = 138.3$. So the sample size has to be ≥ 139.

3. The contents of 7 similar containers of sulphuric acid are: 9.8, 10.2, 10.4, 9.8, 10.0, 10.2, and 9.6 litres. Find the 95% confidence interval for the mean of all such containers, assuming an approximate normal distribution.

   The sample mean (arithmetic average) ix $\bar{x} = 10$, and the unbiased variance estimator

   $$S^{*2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

   gives the value: $S^* = 0.283$. We find the appropriate quantile of Student's distribution: $t_{0.975}(\nu = 7-1 = 6) = 2.447$. So we have

   $$10 - (2.477)\left(\frac{0.283}{\sqrt{7}}\right) < \mu < 10 + (2.477)\left(\frac{0.283}{\sqrt{7}}\right)$$

   which reduces to $9.74 < \mu < 10.26$.

4. In a random sample of $n = 500$ families owning a TV set in the city of Hamilton (Canada) it was found that $x = 340$ subscribed to the HBO (a cable TV company). Find a 95% confidence interval for the actual proportion of families in this city who subscribe to HBO.

   Hint: This is a Bernoulli-type problem, but the Bernoulli distribution tends to normal if the size is adequate (it is). The point estimate of "probability of success" (if one considers it a success to be an HBO subscriber) is $p = x/n = 340/500 = 0.68$. The variance of this average value is the variance of x, $VAR(x) = npq$ divided by n^2 (why?), i.e. pq/n where $q = 1 - p$. Using our favourite normal distribution quantiles (1.96) we get the interval (0.64,0.72).

5. An electrical firm manufactures light bulbs that have a length of life approximately normally distributed with a standard deviation of 40 hours. If a sample of 30 bulbs has an average life of 780 hours, find a 95% confidence interval for the population mean of all bulbs produced by this firm. (765,795).

   How large a sample is needed if we wish to be 95 % confident that our sample mean will be within 10 hours of the true mean? (68)