Maximum Likelihood Estimators

We use statistics drawn from samples to estimate population parameters. We have learnt principally that we can use the sample mean to estimate the population mean, and the unbiased sample variance to estimate the population variance.

However, this is just one approach to estimation, and others do exist. One such alternative approach goes by the name of "Maximum Likelihood Estimators".

We would use this approach when a single experiment produces two or more sample statistics, each of which could be used as an estimate for a population parameter. The central idea is to base our estimate of the population parameter on the value of the population parameter that would have made the experimental outcome most likely to have occurred.

The most typical situation in which this type of estimator is used is when one is estimating a probability. For example, the likelihood of a success occurring in a trial. We use the sample mean and unbiased sample variance to estimate the population mean and variance, but not a probability of an event occurring, so the context in which a maximum likelihood estimator is used is different, and an alternative approach is required.

Example

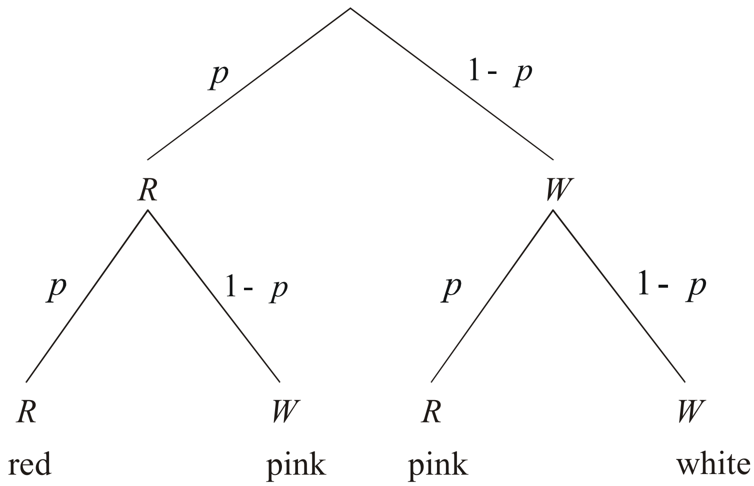
In genetic theory the colour of a flower may be determined by a single gene which may exist in two forms (called alleles). For a particular plant, both forms are co-dominant. One allele codes for red flowers and another for white flowers. The allele for red flowers is represented by R and the allele for white flowers by W. If the flower has one R and one W allele, then the flower will be pink. The probability of having an allele coding for a red flower is p; hence the probability of having an allele code for a white flower is W. Form a probability tree to determine in terms of the population parameter p the probabilities of a flower having red, pink or white flowers. Determine the maximum likelihood estimator for p. Given that in an experiment a flowers were found to be red, b to be pink and c to be white, find an expression for the estimate of p in terms of a, b and c. Find the value of this estimate given that in fact $a = 12$, $b = 6$ and $c = 2$.

We talk you through the solution

Solution

The probability tree and the model

Firstly, we have to draw the probability tree.



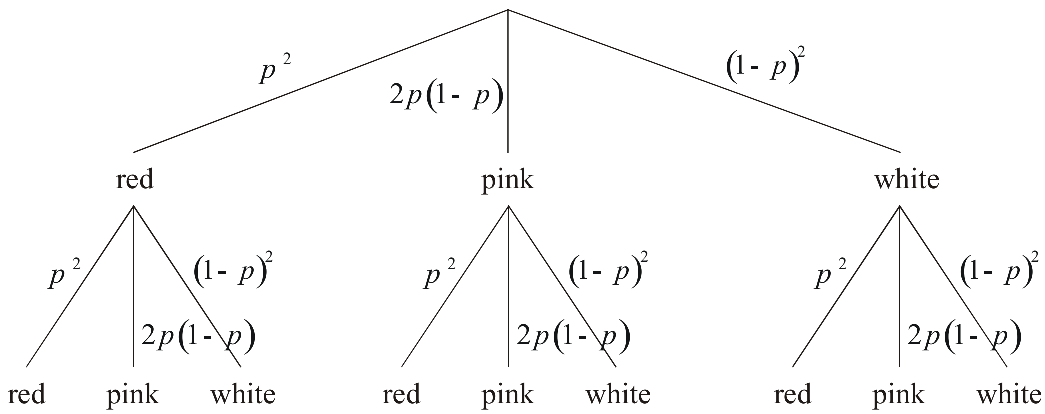
From the tree we can see that the probabilities of the three possible outcomes is

$$P(red) = p^2$$
$$P(pink) = 2p(1-p)$$
$$P(white) = (1-p)^2$$

Finding the maximum likelihood estimator

The experimental outcome was that *a* flowers were red, *b* pink and *c* white. This gives a total of $n = a + b + c$ flowers in the sample. Given that the probability of a red allele is *p*, we want to find an expression for the probability of the sample containing *a* flowers were red, *b* pink and *c* white and then make this a maximum.

If there were just two trials the probability tree would look like this



The probability of obtaining a pink followed by a red flower would be

$$P(\text{pink then red}) = 2p(1-p) \times p^2$$

But for a pink and a red in any order we have to multiply by the number of branches, which is, in two trials, just 3.

For $n = a + b + c$ trials where *a* flowers were red, *b* pink and *c* white in exactly that order the probability would be

$$P(a \text{ red, } b \text{ pink, } c \text{ white}) = \left(p^2\right)^a \times \left(2p(1-p)\right)^b \times \left((1-p)^2\right)^c$$

But if we want to find the probability of *a* red, *b* pink and *c* white in any order we must multiply by the total number of branches that give this combination. That is the total number of permutations of *a* red, *b* pink and *c* white flowers.

This is a problem in permutations were we have *n* objects to permute, but there are *a* identical twins of one kind, *b* identical twins of another and *c* identical twins of a third kind. (You may want to look over the unit on permutations and combinations at this point!) The total number of permutations giving *a* red, *b* pink and *c* white flowers is

$$\frac{(a+b+c)!}{a!b!c!} = \frac{n!}{a!b!c!}$$

Thus the probability of *a* red, *b* pink and *c* white flowers in any order is

$$P(a \text{ red, } b \text{ pink, } c \text{ white}) = \frac{(a+b+c)!}{a!b!c!} \times \left(p^2\right)^a \times \left(2p(1-p)\right)^b \times \left((1-p)^2\right)^c$$

This gives the probability of the particular experimental outcome as a function of the probability of a red allele, *p*. We can express this as a function.

$$L(p) = \frac{(a+b+c)!}{a!b!c!} \times \left(p^2\right)^a \times \left(2p(1-p)\right)^b \times \left((1-p)^2\right)^c$$

This maximum likelihood estimator for the population parameter *p*.

<u>Determining an estimate for *p*</u>

The estimate for *p* will be the value of

$$L(p) = \frac{(a+b+c)!}{a!b!c!} \times \left(p^2\right)^a \times \left(2p(1-p)\right)^b \times \left((1-p)^2\right)^c$$

That makes this function a maximum. That means, we have to differentiate it and set the derivative to zero, and hence find the maximum. We set this problem up as follows

$$\text{Find } \hat{p} \text{ such that } \left.\frac{dL}{dp}\right|_{o=\hat{p}} = 0$$

(We will assume for the present that this value will be a maximum; to prove that it is a maximum we would strictly have to differentiate a second time and use the usual criterion that the second derivative is negative for a maximum. We will not bother with this tedious extra work here!)

This sets up an apparently difficult additional problem, to which there is, fortunately a simple answer – how does one go about differentiating this?

$$L(p) = \frac{(a+b+c)!}{a!b!c!} \times \left(p^2\right)^a \times \left(2p(1-p)\right)^b \times \left((1-p)^2\right)^c$$

However, this problem is simply solved by using logarithmic differentiation instead. Note that we are not really interested in the derivative of $L(p)$ but rather in just when it takes a maximum. But since this function is always positive it always has a logarithm; and the function

$$y = \ln x$$

is an always increasing function. This means that if $L(p)$ has a maximum at $\hat{p}$ then $\ln[L(p)]$ also has a maximum at $\hat{p}$. Differentiating the logarithm of $L(p)$ is much easier than differentiating $L(p)$ itself, as we shall show.

Note also the introduction of the symbol $\hat{p}$ - the *p* with the hat on it. The hat is the standard symbol used to denote an estimate of a population parameter derived from a sample; it is here the value of *p* that makes $L(p)$ a maximum – which is why it is called the "maximum likelihood estimator".

Now to differentiate $\ln[L(p)]$; to do this we use the properties of logarithms that

$$\ln(pq) = \ln p + \ln q$$

and

$$\ln(p)^n = n\ln(p)$$

So firstly,

$$\begin{aligned}
\ln[L(p)] &= \ln\left[\frac{(a+b+c)!}{a!b!c!}\times(p^2)^a\times(2p(1-p))^b\times\left((1-p)^2\right)^c\right] \\
&= \ln\left(\frac{(a+b+c)!}{a!b!c!}\right)+\ln(p^2)^a+\ln(2p(1-p))^b+\ln\left((1-p)^2\right)^c \\
&= \ln\left(\frac{(a+b+c)!}{a!b!c!}\right)+2a\ln p+b\ln 2+b\ln 2p+b\ln(1-p)+2c\ln(1-p) \\
&= \ln\left(\frac{(a+b+c)!}{a!b!c!}\right)+b\ln 2+(2a+b)\ln p+(b+2c)\ln(1-p)
\end{aligned}$$

Now to differentiate it we must remember the chain rule, and the rule that

$$\frac{d}{dx}\ln x = \frac{1}{x}$$

Combinining the chain rule with this

$$\frac{d}{dx}\ln(f(x)) = \frac{1}{f(x)}\times f'(x)$$

Remember the derivative of a constant is also zero; hence

$$\frac{d}{dp}\ln\left[L(p)\right]=\frac{d}{dp}\ln\left(\frac{(a+b+c)!}{a!b!c!}\right)+b\ln 2+(2a+b)\ln p+(b+2c)\ln(1-p)$$

$$=\frac{2a+b}{p}-\frac{(b+2c)}{(1-p)}$$

For turning points, when $p=\hat{p}$ we have

$$\frac{2a+b}{\hat{p}}-\frac{(b+2c)}{(1-\hat{p})}=0$$

$$\frac{2a+b}{\hat{p}}=\frac{(b+2c)}{(1-\hat{p})}$$

$$(2a+b)(1-\hat{p})=\hat{p}(b+2c)$$

$$(2a+b)-(2a+b)\hat{p}=\hat{p}(b+2c)$$

$$\hat{p}(2a+2b+2c)=2a+b$$

$$\hat{p}=\frac{2a+b}{2(a+b+c)}=\frac{2a+b}{2n}$$

<u>Determining the specific estimate</u>

We have shown that

$$\hat{p}=\frac{2a+b}{2(a+b+c)}=\frac{2a+b}{2n}$$

So given that $a=12,\ b=6$ and $c=2$

$$\hat{p}=\frac{2\times 12+6}{2\times 20}=0.75$$