Every human is a carrier of one of the three genotypes AA, Aa, or aa. The genotypes are occurring with the probabilities (1 – p) ^ 2, 2 \* p \* (1 – p) and p ^ 2 whereas 0 < p < 1 and testing of n persons yielded

* x persons had the genotype AA
* y persons had the genotype Aa
* z persons had the genotype aa

Describe the corresponding statistical model and determine the Maximum Likelihood Estimator for p.

Solution:

The likelihood function is given by [1]:

*P (x, y, z | p) = \* \* \* \* \**  …….…. (1)

Taking log likelihood of (1) we get,

*ln (P (x, y, z | p) =*

*ln (P (x, y, z | p) = ln + ln () + ln () + ln () + ln () + ln ()*

*ln (P (x, y, z | p) = constant1 + 2 \* x \* ln () + constant2 + y \* ln (p) + y \* ln (1 – p) + constant3 + 2 \* z \* ln (p)* ...……………………………………...……. (2)

We set the derivative equal to zero:

Solving equation (3) we got the value of *p*.

* *p =*

The corresponding statistical model is “multinomial distribution model”. An extension of the binomial distribution is the multinomial distribution. The multinomial distribution is used to simulate the results of *n* experiments, where each trial's outcome has a categorical distribution [3].

Exactly one of the fixed finite number *k* of possible results with probabilities *p1, p2, …, pk* (here *pi* ≥ 0 for *i = 1, …, k* and ), and there are *n* independent trials. Next, the random variable *Xi* indicates the number of times outcome number *i* was observed over the n experiments. Then *X* = (*X1, X2, …, Xk*) follows a multinomial distribution with the parameters *n* and *p*. Where *p = (p1, p2, …, pk)* [2].

The PMF of the multinomial distribution is given by

with, , and

Reference:

1. <https://math.mit.edu/~dav/05.dir/class10-prep.pdf>
2. Sinharay, Sandip. "Discrete Probability Distributions." (2010): 132-134.
3. Multinomial distribution, <https://en.wikipedia.org/wiki/Multinomial_distribution>