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]:

$$P\,(x, y, z \mid p) = \binom{x+y+z}{x} * (1-p)^{2*x} * \binom{y+z}{y} * (2*p*(1-p))^{y} * \binom{z}{z} * p^{2*z} \quad \ldots\ldots\ldots\ (1)$$

Taking log likelihood of (1) we get,

$$\ln\,(P\,(x, y, z \mid p) = \ln\left(\binom{x+y+z}{x} * (1-p)^{2*x} * \binom{y+z}{y} * (2*p*(1-p))^{y} * \binom{z}{z} * p^{2*z}\right)$$

$$\ln\,(P\,(x, y, z \mid p) = \ln\binom{x+y+z}{x} + \ln\,((1-p)^{2*x}) + \ln\,(\binom{y+z}{y}) + \ln\,((2*p*(1-p))^{y}) + \ln\,(\binom{z}{z}) + \ln\,(\binom{z}{z} * p^{2*z})$$

$$\ln\,(P\,(x, y, z \mid p) = constant_1 + 2*x*\ln\,(1-p) + constant_2 + y*\ln\,(p) + y*\ln\,(1-p) + constant_3 + 2*z*\ln\,(p) \quad \ldots\ldots\ldots\ (2)$$

We set the derivative equal to zero:

$$\frac{2*z+y}{p} - \frac{y+2*x}{1-p} = 0 \quad \ldots\ldots\ldots\ (3)$$

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

$$\frac{2*z+y}{p} - \frac{y+2*x}{1-p} = 0$$

$$\rightarrow \frac{(1-p)*(2*z+y)-p*(y+2*x)}{p*(1-p)} = 0$$

$$\rightarrow 2*z - 2*z*p + y - y*p - y*p - 2*x*p = 0$$

$$\rightarrow 2*z + y - 2*z*p - 2*y*p - 2*p*x = 0$$

$$\rightarrow (2*z+y) - p*(2*z + 2*y + 2*x) = 0$$

$$\rightarrow p = \frac{2*z+y}{2*x+2*y+2*z}$$

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 $p_1, p_2, \ldots, p_k$ (here $p_i \geq 0$ for $i = 1, \ldots, k$ and $\sum_{i=1}^{k} pi = 1$), and there are *n* independent trials. Next, the random variable X_i indicates the number of times outcome number *i* was observed over the n experiments. Then $X = (X_1, X_2, \ldots, X_k)$ follows a multinomial distribution with the parameters *n* and *p*. Where $p = (p_1, p_2, \ldots, p_k)$ [2].

The PMF of the multinomial distribution is given by

$$P\ (X1\ =\ x1, X2\ =\ x2, \ldots, Xk\ =\ xk)\ =\ \frac{n!}{x1!\ x2! \ldots xk!} *\ P1^{x1}\ *\ P2^{x2} \ldots Pk^{xk}$$

with, $\sum_{i=1}^{k} xi\ =\ n$, and $\sum_{i=1}^{k} pi\ =\ 1$

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