BU-287-M

A Minimax Estimator of the Bernoulli Variance p(1-p)

D. S. Robson, Cornell University*

ABSTRACT

An estimator of $p(1-p)$ which is minimax with respect to mean squared error is a linear function of the minimum variance unbiased estimators $s^2 = n\bar{X}(1-\bar{X})/(n-1)$, where $\bar{X}$ is the proportion of "successes" in n Bernoulli trials. This minimax estimator, given by

$$\left(1 + \sqrt{\frac{4n-6}{n(n-1)}}\right)^{-1} \left[\sqrt{\frac{n-1}{4n(4n-6)}} + \frac{n\bar{X}(1-\bar{X})}{n-1}\right] = a + bs^2$$

has a constant mean squared error equal to a^2 and is admissible.

* Paper No. BU-113 in the Biometrics Unit, and No. 491 in the Plant Breeding Department, Cornell University, Ithaca, New York.

A Minimax Estimator of the Bernoulli Variance p(1-p)

D. S. Robson, Cornell University*

The standard unbiased estimator of the Bernoulli variance p(1-p) based upon n independent and identically distributed Bernoulli variables $X_1,\cdots,X_n$,

$$s^2 = \frac{\sum_1^n (X_i-\bar{X})^2}{n(n-1)} = \frac{n}{n-1}\bar{X}(1-\bar{X})$$

is itself subject to a sampling variance of

$$E[s^2-p(1-p)]^2 = \frac{p(1-p)}{n} - \frac{p^2(1-p)^2(4n-6)}{n(n-1)} \quad .$$

In this note we derive a minimax estimator of the form $a+bs^2$ with

$$E[a + bs^2 - p(1-p)]^2 \underset{0 \le p \le 1}{\equiv} a^2 \quad .$$

The mean squared error of a linear function of s^2,

$$b^2E[s^2 - p(1-p)]^2 + [E(a+bs^2) - p(1-p)]^2$$

$$= a^2 + \left[\frac{b^2}{n} - 2a(1-b)\right]p(1-p) - \left[\frac{b^2(4n-6)}{n(n-1)} - (1-b)^2\right]p^2(1-p)^2$$

is identically a^2 when $a = b^2/[2n(1-b)]$ and b is chosen as either of the two

---

* Paper No. BU-113 in the Biometrics Unit, and No. 491 in the Plant Breeding Department, Cornell University, Ithaca, New York.

roots

$$\frac{1}{b_+} = 1 + \sqrt{\frac{4n-6}{n(n-1)}} \qquad \frac{1}{b_-} = 1 - \sqrt{\frac{4n-6}{n(n-1)}}$$

Since the values a_+ and a_- corresponding to these two roots satisfy the relation $|a_+/a_-| = b_+/b_- < 1$ then only the estimator $a_+ + b_+ s^2$ is a candidate for the minimax strategy.

The minimax property and also admissibility are then confirmed by noting that this estimator is a Bayes solution with respect to a (symmetric) Beta prior distribution

$$I_p(\theta,\theta) = \frac{\Gamma(2\theta)}{\Gamma(\theta)\Gamma(\theta)} \int_0^p x^{\theta-1} (1-x)^{\theta-1} \, dx \quad .$$

The a posteriori distribution is then $I_p(\theta+n\bar{x}, \theta+n(1-\bar{x}))$, and the Bayes solution

$$E[p(1-p)|\bar{x}] = \frac{(\theta+n\bar{x})(\theta+n(1-\bar{x}))}{(n+2\theta)(n+2\theta+1)}$$

$$= \frac{\theta(n+\theta)}{(n+2\theta)(n+2\theta+1)} + \frac{n(n-1)s^2}{(n+2\theta)(n+2\theta+1)}$$

is then equal to $a_+ + b_+ s^2$ for

$$\theta = \frac{n}{2}\left[-1 + \sqrt{1 + \frac{4(n-1)a_+}{nb_+}}\right] > 0 \quad .$$