The Uncertainty Principle for Screening Tests

Valentin Vankov Iliev

December 18, 2020

Introduction

Let us suppose that a certain property α is present in some members of a finite population S and that there exists a screening test designed to detect this property, if present. The test does not always work perfectly — sometimes it is negative under the condition that the property is present (that is, *false negative*), and sometimes it is positive under the condition that the property is absent (that is, *false positive*). Let us suppose that all members of the population are tested.

Assume that the property α is present in $p100\%$ of the members of this population, $0 < p \leq 1$, and for $q100\%$, $0 < q \leq 1$, of them, the test is positive.

The aim of this presentation is, under the above conditions, to describe the relation between the conditional probabilities of a false positive and false negative test.

Non-formally, we can state the main result of the paper as an *Uncertainty Principle*: In general, if one has better knowledge that the test is really positive (the probability F_+ of false positive test is small), then for one is hard to know that the test is really negative (the probability F_- of false negative test is large). And the better one knows that the test is really negative (F_- is small), the harder it is to know that the test is really positive (F_+ is large).
For any subset $B \subset S$ which contains $b100\%$ of the members of this population $0 \leq b \leq 1$, we set $P(B) = b$. Let us consider the two sets A_+, A_- of members of the population which do (not) possess the property α and the two sets T_+, T_- of members of the same population which have a positive (negative) test.

Examples of Tests

(1) Screening tests: According to Merriam-Webster Dictionary, this is "...a preliminary or abridged test intended to eliminate the less probable members of an experimental series". In this setup M_+ is the set of all most probable members of this experimental series and N_+ is the set of all members of this series which are not eliminated by the test.

(2) Medical Screening tests: In accord with Stedman's Medical Dictionary, this is "...a simple test performed on a large number of people to identify those who have or are likely to develop a specified disease". In this setup M_+ is the set of all people who developed or are likely to develop this disease and N_+ is the set of all people who have a positive test. We suppose that this is an universal screening (all members of the given population are tested).

Note that the partition

$$(A_+ \cap T_+) \cup (A_+ \cap T_-) \cup (A_- \cap T_+) \cup (A_- \cap T_-),$$

of S is an experiment (in Kolmogorov's terminology) and the probabilities $x_1 = P(A_+ \cap T_+)$, $x_2 = P(A_+ \cap T_-)$, $x_3 = P(A_- \cap T_+)$, $x_4 = P(A_- \cap T_-)$, of its results satisfy a linear system. The general solution of this system depends on one free variable, say, $x_4 = u$, whose range is a closed interval U varying with p and q. This fact determines the ranges of the remaining probabilities x_1, x_2, x_3, and of the conditional probabilities F_+, and F_-, because they are linear functions in u. Thus, the sets A_+, A_-, T_+, T_- vary with u: $A_+ = A_+(u)$, $A_- = A_-(u)$, $T_+ = T_+(u)$, $T_- = T_-(u)$.
In particular, the product $F_+(u)F_-(u)$ is a quadratic function in u, whose simple behaviour yields the Uncertainty Principle.

Introduction

In case $p = 1$ (all members of the population have property α) the probability of a false negative test is equal to the probability of a negative test and the probability of a false positive test is not defined. In case $q = 1$ (all members of the population have positive test) we have $u = 0$, the probability of a false negative test is 0, and, when $p < 1$, the probability of a false positive test is 1. In the exposition below we exclude the above two trivial cases and suppose $0 < p, q < 1$. Under these conditions, the range U of u is a closed interval with non-empty interior $\mathring{U}$ and neither the probability of false negative test nor the probability of false positive test is 0 on the interior. Thus, a perfect test does not exist in general!

More precisely, a test is said to be *perfect* if there exists $u \in U$ such that $A_+(u) = T_+(u)$. It turns out that the test is perfect if and only if $p = q$ and $u = 1 - p$ (in this case — the right endpoint of its range), or, if and only if $F_+(u) = F_-(u) = 0$ for some u.

Introduction

It turns out that the trustworthiness of the test increases when the probability u increases on its range U. The results vary from statistically acceptable when u lies in a small neighbourhood of the right endpoint of U, to implausible when u lies in a small neighbourhood of the left endpoint of U.

One has to note that in case property α is a disease, the probability x_1 that a person has the complete set of symptoms, and, in the same time, has positive test, can be approximated more or less precisely, being related linearly with u: $x_1 = p + q - 1 + u$. In this way we can approximate the reliability of the test.

In the Introduction we considered implicitly the population S as a finite sample spaces with equally likely outcomes and denoted its probability function by P. Below we assume this explicitly.

Let us suppose $0 < p, q < 1$. The probabilities x_k, $k = 1, 2, 3, 4$, satisfy the linear system

$$\left| \begin{array}{ccccccccc} x_1 & + & x_2 & & & & & = & p \\ & & & & x_3 & + & x_4 & = & 1-p \\ x_1 & & & + & x_3 & & & = & q \\ & & x_2 & & & + & x_4 & = & 1-q \end{array} \right.$$

with general solution

$$x_1 = p + q - 1 + u, x_2 = 1 - q - u, x_3 = 1 - p - u, x_4 = u$$

and constraint conditions $0 \leq x_k \leq 1$, that is,

$$\rho(p, q) \leq u \leq \mu(p, q),$$

where $\rho(p, q) = \max(0, 1 - p - q)$ and $\mu(p, q) = \min(1 - p, 1 - q)$.

Since the equality $\rho(p,q) = \mu(p,q)$ yields that either one of p and q is 1 or one of them is 0, we obtain $\rho(p,q) < \mu(p,q)$. In particular, the range of x_i, $1 \leq i \leq 3$, also is a closed interval $[\rho_i(p,q), \mu_i(p,q)]$ with $\rho_i(p,q) < \mu_i(p,q)$.

We define the *admissible solutions* of the above linear system as its solutions from the compact cube $\prod_{i=1}^{4}[\rho_i(p,q), \mu_i(p,q)]$ in $\mathbb{R}^4$, where $\rho_4(p,q) = \rho(p,q)$ and $\mu_4(p,q) = \mu(p,q)$. More precisely, the admissible solutions are the points on the segment of the straight line of solutions, which have endpoints

$$(\rho_1(p,q), \rho_2(p,q), \rho_3(p,q), \rho_4(p,q)),$$

$$(\mu_1(p,q), \mu_2(p,q), \mu_3(p,q), \mu_4(p,q)).$$

Note that this is a diagonal of the cube.

The probability table below summarizes briefly this information:

	T_+	T_-
A_+	$p+q-1+u$	$1-q-u$
A_-	$1-p-u$	u

The probability $F_-(u) = P(T_-(u)|A_+(u))$ of a false negative test and the probability $F_+(u) = P(T_+(u)|A_-(u))$ of false positive test are

$$F_-(u) = \frac{1-q-u}{p}, F_+(u) = \frac{1-p-u}{1-p}.$$

THEOREM

Let $0 < p, q < 1$, *let* $u \in (\rho(p,q), \mu(p,q))$, *and let us set*

$$f(u) = \frac{(u-1+p)(u-1+q)}{p(1-p)}.$$

(i) *One has*

$$F_-(u)F_+(u) = f(u).$$

(ii) *If* $\rho(p,q) < r < \mu(p,q)$ *and if* $\rho(p,q) \leq u \leq r$, *then* $f(r) > 0$ *and*

$$F_-(u)F_+(u) \geq f(r).$$

The Uncertainty Principle: Remarks

If $r = \mu(p, q)$, then $f(r) = 0$, hence $F_-(r)$ or $F_+(r)$ equals 0. More precisely, if $r = 1 - p$, then the table is

	T_+	T_-
A_+	q	$p - q$
A_-	0	$1 - p$

In particular, $F_-(r) = \frac{p-q}{p}$ and $F_+(r) = 0$. If $r = 1 - q$, then the table is

	T_+	T_-
A_+	p	0
A_-	$q - p$	$1 - q$

In particular, $F_-(r) = 0$ and $F_+(r) = \frac{q-p}{1-p}$.
Thus, in case $p = q$ and $r = 1 - p$, the test is perfect:
$F_+(r) = F_-(r) = 0$.

The Uncertainty Principle: Remarks

If r is close to $\mu(p, q)$, then $f(r)$ is close to 0, and in this case the probabilities $F_{-}(r)$ or $F_{+}(r)$ can be statistically acceptable.
More precisely, when $p < q$ the probability $r = x_4$ approaches $1 - q$. In accord with the above linear system, x_2 approaches 0 and x_3 approaches $q - p$. In other words, $F_{-}(r)$ tends to 0 and $F_{+}(r)$ tends to $\frac{q-p}{1-p}$.
When $p > q$ the probability $r = x_4$ approaches $1 - p$, hence x_3 approaches 0 and x_2 approaches $p - q$. Equivalently, $F_{+}(r)$ tends to 0 and $F_{-}(r)$ tends to $\frac{p-q}{p}$.
Thus, both probabilities $F_{-}(r)$ or $F_{+}(r)$ are statistically acceptable when r is close to $\mu(p, q)$ and the difference $p - q$ is small. In particular, the test "approaches" the perfect test when $p = q$.

The Uncertainty Principle: Remarks

If $r = \rho(p,q)$, that is, $f(r) = \max_{u\in[\rho(p,q),\mu(p,q)]} f(u)$, then it turns out that one of $F_-(r)$ or $F_+(r)$ is 1. More precisely, if $r = 0$, then the table is

	T_+	T_-
A_+	q	$1-q$
A_-	$1-p$	0

In particular, $F_-(r) = \frac{1-q}{p}$ and $F_+(r) = 1$.
If $r = 1-p-q$, then the table is

	T_+	T_-
A_+	0	p
A_-	q	$1-p-q$

In particular, $F_-(r) = 1$ and $F_+(r) = \frac{q}{1-p}$.
Thus, when $p+q=1$ and $r=0$, we have $F_+(r) = F_-(r) = 1$.

The Uncertainty Principle: Remarks

If r is close to $\rho(p,q)$, then $f(r)$ is close to its maximum, and in this case at least one of the probabilities $F_-(r)$ or $F_+(r)$ is not statistically acceptable.

More precisely, when $p+q \geq 1$ the probability $r = x_4$ approaches 0. In accord with the above linear system, x_2 approaches $1-q$ and x_3 approaches $1-p$. In other words, $F_-(r)$ tends to $\frac{1-q}{p}$ and $F_+(r)$ tends to 1.

When $p+q < 1$ the probability $r = x_4$ approaches $1-p-q$, hence x_2 approaches p and x_3 approaches q. Equivalently, $F_-(r)$ tends to 1 and $F_+(r)$ tends to $\frac{q}{1-p}$.

Thus, both probabilities $F_-(r)$ or $F_+(r)$ are not statistically acceptable when r is close to $\rho(p,q)$ and the sum $p+q$ is close to 1.

Examples

(1) Let $p = q = 0.5$. Then $0 \leq u \leq 0.5$ and the table is

	T_+	T_-
A_+	u	$0.50 - u$
A_-	$0.50 - u$	u

Moreover, $f(u) = \frac{(u-0.5)^2}{0.25}$. If $r = 0.25$, then $f(0.25) = 0.25$. If, in addition, $0 \leq u \leq 0.25$, then

$$F_-(u)F_+(u) \geq 0.25.$$

In particular, $F_-(u) \geq 0.25$ and $F_+(u) \geq 0.25$.
If $r = 0.48$, then $f(0.48) = 0.0016$. If, in addition, $0 \leq u \leq 0.48$, then

$$F_-(u)F_+(u) \geq 0.0016.$$

Examples

In particular, $F_{-}(u) \geq 0.0016$ and $F_{+}(u) \geq 0.0016$. In case we need $F_{-}(u) \leq 0.01$ for some u with $0 \leq u \leq 0.48$, then we necessarily have $F_{+}(u) \geq 0.16$!

(2) Let $p = 0.10$ and $q = 0.05$. Then $0.85 \leq u \leq 0.90$ and the table is

	T_{+}	T_{-}
A_{+}	$u - 0.85$	$0.95 - u$
A_{-}	$0.90 - u$	u

We have $f(u) = \frac{(u-0.90)(u-0.95)}{0.09}$.

If $r = 0.87$, then $f(0.87) > 0.026$. If, in addition, $0.85 \leq u \leq 0.87$, then

$$F_{-}(u)F_{+}(u) > 0.026.$$

EXAMPLES

In particular, $F_{-}(u) > 0.026$, $F_{+}(u) > 0.026$.
If, for example, we need $F_{-}(u) \leq 0.05$ for some u with $0.85 \leq u \leq 0.87$, then we necessarily have $F_{+}(u) > 0.52$!
If $r = 0.89$, then $f(0.89) > 0.006$. If, in addition, $0.85 \leq u \leq 0.89$, then

$$F_{-}(u)F_{+}(u) > 0.006.$$

In particular, $F_{-}(u) > 0.006$, $F_{+}(u) > 0.006$.
If, for example, we need $F_{+}(u) \leq 0.05$ for some u with $0.85 \leq u \leq 0.89$, then we necessarily have $F_{-}(u) > 0.12$!

Examples

Let us consider statistical hypothesis testing and let H_0 be the null hypothesis that should be rejected. We denote by p the probability that the hypothesis H_0 is true and by q the probability that the hypothesis H_0 is not rejected. By analogy with the above tables, in the table below the probabilities x_1, x_2, x_3, x_4 of the corresponding intersections satisfy the above linear system:

	H_0 is not rejected	H_0 is rejected
H_0 is true	x_1	x_2
H_0 is false	x_3	x_4

Here, $F_- = \frac{x_2}{p}$ is so called *type I error*: The probability of rejecting H_0 under the condition that H_0 is true.

Examples

Moreover, $F_+ = \frac{x_3}{1-p}$ is so called *type II error*: The probability of not rejecting H_0 under the condition that H_0 is false. In statistics the probability F_- of false rejection of H_0 is also called *significance level of the test* and is denoted by α. The probability F_+ of false not rejection of H_0 is denoted by β and the probability $1-\beta$ is called *power of the test*. In accord with the main result of this paper, in general, there exists a positive constant C such that $\alpha\beta > C$. In other words, α and β can not be simultaneously as small as one wants.

Acknowledgements

I would like to cordially thank both the administration of the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences, and the administration of the American University in Bulgaria, for creating perfect and safe conditions of work.
I would also like to thank the moderators from ArXiv Support (Cornell University Library) with numbers 332 and 445 because they rejected the publication of this text in ArXiv and stimulated my interest to study and generalize the properties of the screening tests.