MODELING AND STATISTICAL ANALYSIS OF A LABORATORY EXPERIMENT
TO MEASURE CANNIBALISM IN HELIOTHIS VIRESCENS LARVAE

by

GOLDE IVAN HOLTZMAN

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

DEPARTMENT OF STATISTICS, BIOMATHEMATICS PROGRAM

RALEIGH
1980

APPROVED BY:

[Signatures]

Chairman of Advisory Committee
HOLTZMAN, GOLDE IVAN. Modeling and Statistical Analysis of a Laboratory Experiment to Measure Cannibalism in Heliothis virescens Larvae. (Under the direction of H. R. VAN DER VAART.)

Statistical and biomathematical techniques are employed to analyze a simple laboratory experiment investigating genetic variation in the propensity for intraspecific predation (cannibalism) among eleven geographic strains of Heliothis virescens (F.) larvae.

The analysis is based on a statistical model of the replicated experiment and a complementary stochastic model of the relevant biological phenomena. The statistical model, based on the trinomial distribution, obtains from the usual assumption of independent replications. The biomathematical model defines a parameter which represents the propensity for cannibalism, and thereby quantifies the propensity for cannibalism in the sense that the statistical estimate of that parameter is an estimate of the propensity for cannibalism exhibited by each strain. The same model determines which statistical hypotheses ought to be tested in order to decide whether or not cannibalism occurs. As a result of the modeling process, the biological assumptions are explicitly recognized and exposed to critical scrutiny. The scope of inferences based on the experiment is delineated, and additional experiments that might be performed in the future are suggested.

A unique, uniformly most powerful size .05, Neyman structure randomized test of the hypothesis that cannibalism does not occur is
derived from the Neyman-Pearson theory, and then implemented using an efficient computer program. (A more complex program is used to compute the power against 100 alternatives.) This test is compared with a supposedly size .05 chi-squared goodness-of-fit test that was used by previous investigators, and with two other tests based on normal approximations of the distribution of the relevant statistic.

The difference between the optimal Neyman structure test and each of the comparable approximate tests is excessive unless the sample size is extremely large. Since the actual size of each of the three approximate tests is substantially greater than the supposed size, and since the exact Neyman structure test can be computed inexpensively, the exact Neyman structure test is preferable.

Comparison of the tests is facilitated by the introduction of the concept of an "expected critical region" for a randomized test (based on a discrete distribution) that is analogous to the critical region of a nonrandomized test (based on a continuous distribution). The nature of the inaccuracy of the approximate tests is investigated by computing and comparing the exact and assumed approximate distributions of the various test statistics.

The dissertation exemplifies the usefulness and importance of biomathematical modeling and the nonroutine practical application of statistical theory.
Golde Ivan Holtzman was born on February 10, 1946 in Pittsburgh, Pennsylvania, where he attended Taylor Alderdice High School. After graduation from high school Mr. Holtzman studied mathematics at the University of California at Los Angeles, where he was awarded the Bachelor of Arts degree on December 8, 1972.

For the next seven months the author was employed as a data coordinator for a public health survey by the Pima County Health Department in Tucson, Arizona. From August 1973 until July, 1975 he worked as a pharmacology research technician in the laboratory of I. Glenn Sipes and Bernard B. Brodie at the University of Arizona College of Medicine in Tucson. During this period he earned a Master of Arts degree in mathematics at the University of Arizona, which was awarded on May 17, 1975.

In July, 1975 the author entered the Biomathematics Program of the Department of Statistics at North Carolina State University in Raleigh. From July, 1975 until June, 1977 he was supported partially by a National Institute of General Medical Sciences fellowship. From January, 1976 until June, 1977 he taught mathematics at NCSU.

In July, 1977, the author became associated with an integrated pest management research program under the direction of Ron Stinner and Robert L. Rabb in the Department of Entomology. In this group Mr. Holtzman functioned as a statistical consultant and ecosystem modeler, and was given wide latitude to pursue his own interests. Interaction with Fred Gould, an ecologist in this group, led to the research presented in this dissertation.
Mr. Holtzman is affiliated with the Institute of Mathematical Statistics, the American Statistical Association, the Biometric Society, the American Association for the Advancement of Science, and Sigma Xi.
ACKNOWLEDGEMENTS

H. R. van der Vaart, chairman of the Advisory Committee, contributed immeasurably to every aspect of this project. He and Ron Stinner gave generously of their guidance, encouragement, and friendship.

Drs. H. R. van der Vaart, Ron E. Stinner, Harvey J. Gold, and David A. Dickey, members of the Advisory Committee, and Dr. Ernest D. Seneca, the Graduate School representative, carefully read the manuscript and suggested many substantial improvements. Fred Gould helped with criticism of the first chapter.

The research was partially supported by National Science Foundation grant number DEB-7822738 to North Carolina State University.
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>2. Statistical Model</td>
<td>11</td>
</tr>
<tr>
<td>3. Biological Model</td>
<td>22</td>
</tr>
<tr>
<td>4. Hypothesis Testing</td>
<td>46</td>
</tr>
<tr>
<td>5. Exact Distribution</td>
<td>79</td>
</tr>
<tr>
<td>6. Approximate Tests</td>
<td>90</td>
</tr>
<tr>
<td>7. Computations</td>
<td>119</td>
</tr>
<tr>
<td>8. References Cited</td>
<td>140</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

Two objectives are pursued in this dissertation. The first is thoroughly to analyze a biological experiment investigating genetic variation in the propensity for cannibalism among different geographic strains of a particular insect species (Gould, et al., 1980). The second objective is to expose the modeling philosophy and techniques, both biomathematical and statistical, that are the basis of this analysis.

The purpose of the laboratory experiment was (1) to determine, for each of eleven geographic strains of Heliothis virescens (F.) larvae, whether intraspecific predation (cannibalism) is a mortality factor or not, and (2) to quantify the propensity for cannibalism exhibited by each strain.

The elementary laboratory experiment consisted of placing two neonate larvae of the same strain in a 25 ml cup containing an excess of diet; rearing them in a chamber controlling temperature, humidity, and light-dark ratio; and, after a period of time long enough for all viable larvae to pupate, observing the number of pupae. Simultaneously, in the same chamber, for each of eleven different geographic strains, the elementary experiment was replicated employing a large number of cups.

The data of the experiment are summarized in table 1.1. For each strain, the number of occurrences (absolute frequency) of each of the three elementary outcomes

\[{\text{no pupae in cup}}, \ {\text{one pupa in cup}}, \ {\text{two pupae in cup}}\] (1.1)

is listed.
Table 1.1: The absolute frequency of each outcome for the laboratory experiment (data from Gould et al., 1980)

<table>
<thead>
<tr>
<th>Strain</th>
<th>Number of pupae</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>College Station, TX (COL)</td>
<td>9</td>
</tr>
<tr>
<td>LD-Brownsville, TX (LDB)</td>
<td>3</td>
</tr>
<tr>
<td>Weslaco, TX (WES)</td>
<td>15</td>
</tr>
<tr>
<td>Brownsville, TX (BRO)</td>
<td>17</td>
</tr>
<tr>
<td>Fargo, ND (FAR)</td>
<td>1</td>
</tr>
<tr>
<td>Riverside, CA (RIV)</td>
<td>0</td>
</tr>
<tr>
<td>Tellulah, LA (TAL)</td>
<td>1</td>
</tr>
<tr>
<td>Arkansas (Tobacco) (ART)</td>
<td>24</td>
</tr>
<tr>
<td>Arkansas (Cotton) (ARC)</td>
<td>15</td>
</tr>
<tr>
<td>Clayton, NC (CLA)</td>
<td>6</td>
</tr>
<tr>
<td>Whiteville, NC (WHI)</td>
<td>6</td>
</tr>
</tbody>
</table>

The analysis includes a stochastic biomathematical model for the cannibalism phenomena (chapter 3), a statistical model for the replicated laboratory experiment (chapter 2), and an efficient Neyman structure test of the null hypothesis that there is no cannibalism (chapter 4).

In exposing the modeling technique, great care has been taken to distinguish two types of modeling—statistical and biomathematical.

The statistical model obtains from the assumption that for each strain, each of the large number of cups employed is an independent replication of the same elementary experiment of raising two larvae in
one cup. The classical theory of statistics is sufficient to model precisely this situation. The particular model used here is old and well known, but as will be shown in the next chapter, the statistical model alone, in the absence of an underlying biological theory of cannibalism, is inadequate for the purpose of analyzing the experiment.

The biomathematical model is a rigorous formulation of the underlying biological theory. As will be shown in chapter 2, the biomathematical model is required to impart meaning to the otherwise ambiguous statistical parameters.

Both models are stochastic. That is, random variables and their probability density functions (pdf's) are used to portray the laboratory experiment at different hierarchical levels. Specifically, the biomathematical model is a pair of random variables \((V, U)\) and their joint pdf representing the biological phenomena that result in the outcome of the elementary experiment (i.e. that result in the development of 0, 1, or 2 pupae in one cup). The statistical model is a hierarchy of random variables, their pdf's, and their interrelationships, representing the experimental procedures that, beginning with the elementary experiment, result in the data reported in table 1.1. The biological and statistical models are connected at the level of the elementary experiment by the mathematical relationship between the biological and statistical parameters. Essentially, this mathematical relationship is the underlying model upon which the entire analysis and all statistical inferences are based.

Afterward, the relevant statistical random variable is transformed by measurable mappings into a random variable which is (in a statistical
sense) sufficient for the purpose of the experiment—namely, to test the hypothesis that there is no cannibalism, and to estimate the propensity for cannibalism. The point is that these forms are determined by the underlying biomathematical model. Without such a model, we would not know which hypotheses to test; nor would we know which function of the statistical parameters to estimate.

In addition to facilitating the analysis of the experiment, the modeling process stimulates further scrutiny of the underlying biological assumptions. Ideas for future experiments become apparent, and a method for testing the validity of the biological model presented here is revealed (chapter 4).

Gould and his co-workers (1980) used a chi-squared goodness-of-fit test to determine from their data whether cannibalism is a mortality factor in each of the eleven geographic strains. This standard statistical procedure is compared with the optimal test which arises from the Neyman structure theory as applied to the biomathematical model (chapter 6).

The explanation of most of the complicated technical computations needed for the analyses are isolated in chapters 5 and 7.

A few years ago in a seminar at the North Carolina State University Department of Statistics, its late founder, Gertrude Cox, apprised the students as prospective consultants of the importance of becoming thoroughly familiar with the underlying subject matter of every statistical problem. To this end we now shall review briefly the bionomics of the species involved and relevant aspects of the phenomenon of cannibalism.
sense) sufficient for the purpose of the experiment—namely, to test the hypothesis that there is no cannibalism, and to estimate the propensity for cannibalism. The point is that these forms are determined by the underlying biomathematical model. Without such a model, we would not know which hypotheses to test; nor would we know which function of the statistical parameters to estimate.

In addition to facilitating the analysis of the experiment, the modeling process stimulates further scrutiny of the underlying biological assumptions. Ideas for future experiments become apparent, and a method for testing the validity of the biological model presented here is revealed (chapter 4).

Gould and his co-workers (1980) used a chi-squared goodness-of-fit test to determine from their data whether cannibalism is a mortality factor in each of the eleven geographic strains. This standard statistical procedure is compared with the optimal test which arises from the Neyman structure theory as applied to the biomathematical model (chapter 6).

The explanation of most of the complicated technical computations needed for the analyses are isolated in chapters 5 and 7.

A few years ago in a seminar at the North Carolina State University Department of Statistics, its late founder, Gertrude Cox, apprised the students as prospective consultants of the importance of becoming thoroughly familiar with the underlying subject matter of every statistical problem. To this end we now shall review briefly the bionomics of the species involved and relevent aspects of the phenomenon of cannibalism.
Three species of the genus *Heliothis* inhabit the southern United States. The tobacco budworm, *H. virescens* (F.), and the corn earworm, *H. zea* (Boddie), are economically serious crop pests throughout much of the western hemisphere. A third species, *H. subflexa* (Guenee), is indigenous to the southern U. S., but feeds only on noncultivated wild plants, and is therefore not economically important. A fourth species *H. armigera* (Hubner), which is almost identical to *H. zea*, is common in Europe and Africa.

Like all insects of the order Lepidoptera, *Heliothis* spp. are holometabolous (Borrer and Delong, 1954, p. 454). According to Peter Price (1975, p. 274), developmental polymorphism is a manifestation within species of Darwin's principle of divergence. As a result of complete metamorphosis, the immature caterpillars and adult moths are different both morphologically and functionally, and therefore occupy distinct ecological niches. The evolutionary advantage of this particular life history strategy is that it eliminates intraspecific competition between the adults and their offspring, and automatically results in higher population levels than would be realized otherwise (Price, 1975, pp. 272-274). The adult stage of *Heliothis* is predominantly reproductive and dispersive, while the larval stage is exclusively vegetative and relatively immobile.

The host plants for *Heliothis* species include tobacco, corn, cotton, soybeans, sorghum, alfalfa, tomatoes, strawberries, clover, and other grains and weeds (Anon., 1972). Host crops, and consequently food for the larvae, are "selected" by the ovipositing adult according to the characteristic preferences of each species, which in turn are influenced
by many exogenous factors: fitness, variety, and phenological state of the host; the diversity, distribution, and biomass of the available hosts (Johnson, et al., 1975); and, perhaps, the phenology and variety of neighboring nonhost plants.

The eggs are laid individually rather than in clusters, but adults do not appear to discriminate between hosts that are already occupied and those that are not. It is therefore common for several eggs to be deposited in close proximity on a single host plant. For example, on corn, _H. armigera_ and _H. zea_ generally oviposit on the silk growing out of the husk. On tobacco, _H. zea_ and _H. virescens_ females oviposit on the younger leaves and floral buds.

Under typical summer conditions, the eruciform (caterpillar-like) neonates hatch three or four days after oviposition. The larval period usually includes five or six instars. The eggs are approximately 0.57 mm in diameter, first instar larvae average 1.5 mm in length, and sixth instar larvae reach a length of 42.3 mm on the average, depending on environmental conditions (Deitz, et al., 1976).

It is during the larval period that cannibalism occurs, if at all. Immediately after they hatch, the neonates crawl to a nearby feeding site. _H. armigera_ and _H. zea_ on corn, for example, crawl from the silk into the husk (Kirkpatrick, 1957). _H. zea_ and _H. virescens_ on tobacco and cotton prefer tender leaves and new reproductive parts. The larvae feed rather indiscriminately. If they happen to encounter another larva of their own species (or of other lepidopterous species), they are liable to eat it as well. Thus, to a certain extent, which we intend to quantify, the larvae are intraspecific predators, or cannibals.
This high degree of adaptability to a wide spectrum of hosts, the "adaptive host plant shift" (Freeman, et al., 1967), contributes significantly to the agricultural problems associated with the boll worm complex, H. zea and H. virescens (Anon., 1972). A second factor is that the adults are strong flyers, have a considerable capacity for local dispersal, and migrate great distances in large numbers borne on high altitude wind currents (Callahan, et al., 1972). The combination of their wide range of hosts and their ability to move great distances makes Heliothis species an ecologically successful opportunist, and therefore quite difficult to control locally with insecticides. Millions of dollars are spent annually attempting to reduce the damage caused by these pests (Anon., 1972).

Two more factors have diminished the effectiveness of chemical control in recent years. First, many arthropod predators of Heliothis are more susceptible to broad spectrum insecticides than Heliothis spp. Second, Heliothis spp. have developed resistance to most organic insecticides since their use became widespread in the late 1940's (Adkisson, 1965).

It's not surprising, therefore, that growers are turning to biological control. Methods of natural control which have been or are being considered include conservation of natural enemies, augmentation of predators and parasites, introduction of pathogens, and mass release of sexually sterilized adults (Anon., 1972).

The United States Department of Agriculture is currently planning a biological control program aimed at H. virescens. A H. virescens-H. subflexa hybrid cross of sterile males and fertile females has been
produced in the laboratory (Laster, 1972). Repeated back crosses of hybrid females with *H. virescens* males have been produced in excess of 25 generations. The hybrids resulting from a large number of backcrosses are in many ways phenotypically indistinguishable from *H. virescens*, but the males remain sterile in each successive generation. It is reasonable to expect, therefore, that a mass release of hybrids would significantly lower the population densities of *H. virescens*.

There are two mechanisms which could effect this control. First, if a hybrid male were to mate with an indigenous *H. virescens* female, no offspring would be produced. The presence of hybrid males, therefore, would lower the reproductive potential of the indigenous populations, all other factors being equal. Second, if the hybrids were to have a relatively high propensity for cannibalism, and if they were phenotypically similar to the local population of *H. virescens* with respect to vigor and host selection, the hybrids might be significant predators of *H. virescens*, and thereby cause a diminution of the number and severity of pest outbreaks (Gould, et al., 1980). Moreover, the offspring of hybrid female x indigenous male matings would also be hybrids, presumably with a higher propensity for cannibalism inherited from the hybrid female.

It is therefore of practical importance to determine the intensity of cannibalism exhibited by different geographic strains. If significant differences among geographic strains were determined, then those strains with the greater cannibalism intensities presumably would be the better candidates for the cross with *H. subflexa* (Dawkins, 1976). Such differences have been found by Gould and his co-investigators.
Their experimental technique, together with the method of analysis presented here, could also be applied to measure the intensity of cannibalism in successive back-crosses. The same technique could be used to quantify interspecific predation between closely related species as well (e.g., Gould et al., 1980).

The development of statistical procedures for accurately measuring the propensity for cannibalism is therefore of some practical value. However, the study of Gould and his co-workers (1980), and the results presented here, apply only to laboratory populations in a strictly controlled artificial environment. To be sure, physical factors were controlled at typical summertime levels, and the size of the container simulates the natural confinement of a contact site (Stinner, et al., 1977) in the field. But the fact that the age structure was fixed, and, most important, that the initial population size was fixed at two, severely limits the range of inferences concerning field populations. Clearly, the propensity for cannibalism determined in these studies is merely an index of some strain characteristic.

This index, on the other hand, is ideally suited for another more theoretical biological application. The phenomenon of cannibalism itself, is the subject of avid research by population ecologists and geneticists who are interested in the evolution of behavioral, rather than physical, traits. Cannibalism is prevalent in a wide variety of animals, from protazoa to mammals (Fox, 1975). In insects (such as Lepidoptera) that do not possess a capacity for learning, the potential for a specific behavior is a heritable phenotypic characteristic (Park, et al., 1965).
Beyond being a behavioral phenomenon, cannibalism is clearly a group, rather than an individual, phenomenon. It can occur only if two or more individuals are present. Cannibalism, therefore, is relevant to the study of interdeme selection (Write, 1960). The fact that cannibalism is both a behavioral and a group phenomenon is the primary reason it has received so much attention, particularly in the studies of cannibalism in laboratory populations of flour beetles (Tribolium, eg. Wade, 1976). In this respect, Gould, et al. (1980) used *H. virescens* as an experimental animal. They established that the propensity for cannibalism is indeed a heritable trait in *H. virescens*, and that it is significantly different among some geographic populations.
2. STATISTICAL MODEL

The data of the experiment are summarized in table 1.1. For each strain, the numbers listed there are the absolute frequencies in \( n \) replications of the three elementary events (1.1). Since the \( n \) replications are assumed to be stochastically independent and identical, the \( n \) replications (or cups) constitute a random sample of size \( n \). Each outcome must fall into one of three mutually exclusive and exhaustive categories. Clearly, the trinomial distribution is the natural model to use for this situation.

Although the formulation of this model is well known and old, dating back over 250 years to the time of J. Bernoulli (Johnson and Kotz, 1969, p. 281), it is worthwhile formally to derive it here for this particular application. To do so will reveal that the trinomial model alone, in the absence of an underlying mathematical model for the biological phenomena, is inadequate for our purpose. A formal derivation also will clarify precisely how and where the biomathematical model complements the statistical model.

We shall begin with the elementary experiment of observing the development of two larvae in one cup, then model the replicated experiment of \( n \) cups, and finally derive the model for the trinomial absolute frequencies of table 1.1, the data of the experiment.

The stochastic model for the elementary experiment is a discrete random variable \( U \) and its pdf \( P_U \). The random variable \( U \) represents the number of larvae that survive to pupate in any particular cup. Thus
U-space is the set of three integers \{0, 1, 2\}, and the three elementary events (1.1) are represented equivalently by

\[
\{U = 0\}, \quad \{U = 1\}, \quad \{U = 2\}. \quad (2.1)
\]

The pdf of \(U\) is denoted by

\[
p^U(i) = a_i, \quad i = 0, 1, 2 \quad (2.2)
\]

where

\[
0 < a_0, a_1, a_2, \quad \text{and} \quad a_0 + a_1 + a_2 = 1. \quad (2.3)
\]

The assumption implicit in (2.3), that none of the outcomes (1.1) or (2.1) is impossible, is realistic biologically and convenient mathematically, in that it eliminates technically bothersome degenerate cases.

It is important to recognize that (2.2) tells us nothing about the nature of the events (2.1). The parameters \(a_i\) are merely symbols for \(p^U(i)\), and nothing more.

The notation (2.2) and the restriction (2.3) establish a parameter space

\[
A = \{(a_0, a_1, a_2) \in \mathbb{R}^3 : 0 < a_0, a_1, a_2; a_0 + a_1 + a_2 = 1\}. \quad (2.4)
\]

Parameter space \(A\) is a two-dimensional subset of Euclidean 3-space \(\mathbb{R}^3\). A point of \(A\) is determined by specification of any two of its coordinates, the third coordinate always being determined by (2.3). Here we will concentrate on the projection of \((a_0, a_1, a_2)\)-space onto the \((a_0, a_2)\) coordinate plane. Parameter space \(A\) is thusly illustrated in figure 2.1.
Figure 2.1: Parameter space $A$ (2.4), projected onto the $(a_0, a_2)$ coordinate plane.
We will refer to the pair \((U, p^U)\) as the elementary experiment. \((U, p^U)\) represents generically the elementary experiment of observing the development of two larvae in one cup. In order to model the replicated experiment of \(n\) cups, we define \(n\) random variables \(U_1, U_2, \ldots, U_n\) representing the number of pupae per cup in each of the \(n\) cups. Thus, for the \(j^{th}\) cup there are three mutually exclusive and exhaustive events which can occur:

\[\{U_j = 0\}, \{U_j = 1\}, \{U_j = 2\}. \quad (2.5)\]

Equivalently, we can represent each of the elementary outcomes (2.5) by a three dimensional Euclidean vector \(v_j\) that has \(i^{th}\) component equal to 1 if \(i\) larvae pupate in cup \(j\), and \(i^{th}\) component equal to 0 otherwise. Thus, for the \(j^{th}\) cup there are three mutually exclusive and exhaustive events which can occur:

\[\{v_j = (1, 0, 0)\}, \{v_j = (0, 1, 0)\}, \{v_j = (0, 0, 1)\}. \quad (2.6)\]

The two representations are equivalent by the identity

\[v_j = [1\{0\}(U_j), 1\{1\}(U_j), 1\{2\}(U_j)],\]

where the indicator function \(1_A(x)\) is defined, as usual, by \(1_A(x) = 1\) if \(x\) is an element of the set \(A\), and \(1_A(x) = 0\), otherwise.

Both representations (2.5) (2.6) are displayed to re-emphasize that it is the occurrence or nonoccurrence of each of the three possible elementary events (1.1) which is important, rather than either representation of the events as the value of a random variable.
Now the parameters $a_0, a_1, a_2$ denote the probability of each of the elementary events as listed in (1.1), (2.5), (2.6). That is,

\[ a_i = P^U(i) = P[U_j = i] = P[\{i\}(U_j) = 1], \quad i = 0, 1, 2; \ j = 1, 2, \ldots, n. \tag{2.7} \]

For each strain the replicated experiment is modeled by the vector valued random variable $U = (U_1, U_2, \ldots, U_n)$ and its pdf $P_U$ defined by

\[ P^U(u) = P[U_1 = u_1, \ldots, U_n = u_n] = \prod_{j=1}^{n} P[U_j = u_j] = \prod_{j=1}^{n} P^U(u_j). \]

It is implicit in the second equality that $U_1, \ldots, U_n$ are independent random variables, and in the third that they and $U$ are identically distributed, reflecting the assumption that the $n$ random variables constitute a random sample. We will refer to the pair $(U, P^U)$ as the replicated experiment.

Note that by the definition of $P^U$ we can compute, for a particular order of values $u_1, \ldots, u_n$ (i.e., for a particular labeling of the cups),

\[ P[U_1 = u_1, \ldots, U_n = u_n] = \prod_{j=1}^{n} P^U(u_j) = a_0^x a_1^y a_2^z, \tag{2.8} \]

where

\[ x = \sum_{j=1}^{n} 1\{0\}(u_j), \ y = \sum_{j=1}^{n} 1\{1\}(u_j), \ z = \sum_{j=1}^{n} 1\{2\}(u_j). \tag{2.9} \]

The computation (2.8) leads us to define a vector valued random variable $X = (X_0, X_1, X_2)$ on $U$-space by

\[ X_i = \sum_{j=1}^{n} 1\{i\}(U_j), \ i = 0, 1, 2 \tag{2.10} \]

or by
\[ X = \sum_{j=1}^{n} v_j. \]  

Thus, the symbols \( x, y, z \) (2.9) denote the feasible values of \( X_0, X_1, X_2 \) (2.10). \( X_0, X_1, X_2 \), are the frequencies in \( n \) replications of the elementary events as listed in (1.1), (2.1), (2.5), and (2.6). In other words, \( X \) represents the data of table 1.1. Moreover, according to (2.8), \( X \) is a sufficient statistic for the family of distributions \( \{P_{\theta} : \theta \in A\} \) (\( X \) is sufficient for \( \theta \) in \( A \), i.e., \( P_{\theta}(U = u \mid X = x) \) is a constant function of \( \theta \) for every \( \theta \) in \( A \)). This means that table 1.1 summarizes all the information in the replicated experiment \( (U, P_{\theta}) \).

Hence, we consider the sufficient statistic \( X \) and the induced pdf \( P^X \) defined by

\[ P^X(x, y, z) = P\{U : \sum_{j=1}^{n} v_j = (x, y, z)\}. \]

\( X \)-space, denoted \( W^X \), is the set of all points \( x = (x, y, z) \) with nonnegative integer components satisfying \( x + y + z = n \). That is, letting \( Z \) represent the set of all integers,

\[ W^X = \{(x, y, z) \in Z^3 : 0 \leq x, y, z; x + y + z = n\}. \]  

(2.12)

Each triple \( (x, y, z) \) in \( W^X \) corresponds to many different orderings of 0, 1, 2 among \( u_1, \ldots, u_n \), each with \( x \) zeros, \( y \) ones, and \( z \) twos, as counted in (2.9). Precisely, to each point \( x \) in \( W^X \) there corresponds \( n! / (x! y! z!) \) points \( u \) in \( U \)-space. According to (2.8), for each value of \( (x, y, z) \) in \( W^X \), every ordering has equal probability \( (a_0^x a_1^y a_2^z) \) of occurring. Therefore \( P^X \) is the trinomial pdf

\[ P^X(x, y, z) = a_0^x a_1^y a_2^z n! / (x! y! z!), \]  

\( (x, y, z) \) in \( W^X \).  

(2.13)
Thus we have a formal probabilistic framework for the statistical model of the laboratory experiment. This model is composed of submodels at three hierarchical levels. At the lowest level there are the generic elementary experiment \((U, p^U)\) and the \(n\) identical and independent elementary experiments \((U_j, p^U_j)\), \(j = 1, 2, \ldots, n\). At the next level there is the replicated experiment \((U, p^U)\). Finally, there is the sufficient statistic \(X\) with trinomial pdf \(p^X\) as given in (2.13).

Each \(X_i\) has expectation \(na_i\) and variance \(na_i(1 - a_i)\), \(i = 0, 1, 2\). Of course the trinomial random variables \(X_0, X_1, X_2\) are not stochastically independent. The covariance of \(X_i, X_j\) is

\[
\text{Cov}(X_i, X_j) = -n a_i a_j, \quad i \neq j,
\]

reflecting the obvious fact that the more outcomes fall into one class \((1,1)\), the fewer can fall into another.

Now we consider the random variables \(X_i/n, i = 0, 1, 2\), which are the relative frequencies of each of the three elementary events \((1,1)\). In agreement with intuition, the expected relative frequency of each event is given by

\[
E(X_i/n) = a_i = P\{i \text{ pupae per cup}\}, \quad i = 0, 1, 2.
\]

That is, the average relative frequency of the event \(\{i \text{ pupae}\}\) is the probability of that event, \(i = 0, 1, 2\). Moreover, according to (2.15), \(X_i/n\) is an unbiased estimator for \(a_i\), \(i = 0, 1, 2\). We shall therefore denote \(\hat{a} = (\hat{a}_0, \hat{a}_1, \hat{a}_2)\), where

\[
\hat{a}_i = X_i/n, \quad i = 0, 1, 2.
\]
The relative frequencies $\hat{f}$ for each of the eleven strains are listed in table 2.1 and graphed in figure 2.2 (cf. table 1.1). For each strain, the overall relative frequency of pupation

$$\hat{f} = \frac{(X_1 + 2 X_2)}{(2 n)} = \frac{(\hat{a}_1 + 2 \hat{a}_2)}{2}$$

(2.17)

is tabulated also. For each strain, the statistic $\hat{f}$ is an unbiased, consistent, maximum likelihood estimator for the average relative frequency of pupation of larvae raised two per cup.

$$r = E U/2 = E(X_1 + 2 X_2) / (2 n) = (a_1 + 2 a_2) / 2.$$  

(2.18)

Table 2.1: The relative frequency of each outcome for the laboratory experiment (data from Gould et al., 1980)

<table>
<thead>
<tr>
<th>Strain</th>
<th>n</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>College station, TX</td>
<td>166</td>
<td>0.054217</td>
<td>0.783133</td>
<td>0.162651</td>
</tr>
<tr>
<td>LD-Brownsville, TX</td>
<td>337</td>
<td>0.008902</td>
<td>0.670623</td>
<td>0.320475</td>
</tr>
<tr>
<td>Weslaco, TX</td>
<td>128</td>
<td>0.117188</td>
<td>0.656250</td>
<td>0.226563</td>
</tr>
<tr>
<td>Brownsville, TX</td>
<td>512</td>
<td>0.033203</td>
<td>0.939453</td>
<td>0.027344</td>
</tr>
<tr>
<td>Fargo, ND</td>
<td>138</td>
<td>0.007246</td>
<td>0.818841</td>
<td>0.173913</td>
</tr>
<tr>
<td>Riverside, CA</td>
<td>55</td>
<td>0.000000</td>
<td>0.890909</td>
<td>0.109091</td>
</tr>
<tr>
<td>Tellulah, LA</td>
<td>74</td>
<td>0.013514</td>
<td>0.702703</td>
<td>0.283784</td>
</tr>
<tr>
<td>Arkansas (Tobacco)</td>
<td>225</td>
<td>0.106667</td>
<td>0.764444</td>
<td>0.128889</td>
</tr>
<tr>
<td>Arkansas (Cotton)</td>
<td>228</td>
<td>0.065789</td>
<td>0.653509</td>
<td>0.280702</td>
</tr>
<tr>
<td>Clayton, NC</td>
<td>281</td>
<td>0.021352</td>
<td>0.291815</td>
<td>0.686833</td>
</tr>
<tr>
<td>Whiteville, NC</td>
<td>1073</td>
<td>0.005592</td>
<td>0.295433</td>
<td>0.698975</td>
</tr>
</tbody>
</table>
Figure 2.2: \((a_2, a_0)\), the data of table 2.1.
The parameters of the statistical model are \( n \) and \( \alpha \). The number of replications \( n \), which represents the number of cups employed by the experimentalist, is fixed and known. The parameter \( \alpha \) is unknown but estimable. Indeed, the observable \( \hat{\alpha} = (X_0/n, X_1/n, X_2/n) \) is an unbiased, maximum likelihood estimator of \( \alpha = (\alpha_0, \alpha_1, \alpha_2) \) for \( \alpha \) in \( A \) (2.4). This estimator (\( \hat{\alpha} \)) presumably should tell us all we need to know at the elementary level [i.e., using the elementary experiment model \((U, P^U)\)] where the biological phenomena occur, but this is not the case.

Suppose the value of \( \alpha \) were known. Then the probability of zero, one, or two pupae per cup would be known. That is, the average relative frequency of each of the elementary outcomes (1.1) would be known. Moreover, by (2.18), the average relative frequency of pupation \( r \) would be known. But what would this tell us about the existence or nonexistence of cannibalism as a significant mortality factor? What would this tell us about the intensity of (i.e., propensity for) cannibalism when it is significant? Obviously, a higher propensity for cannibalism would result in greater mortality, which would result, on the average, in a lower observed frequency of pupation \( \bar{r} \). But even if there were no cannibalism at all, there still would be some mortality, which also would result, on the average, in a lower observed frequency of pupation \( \bar{r} \). The average relative frequency of pupation \( r \) is indeed a biologically meaningful parameter. It measures the frequency with which a biological phenomenon—pupation—occurs. What is sought, however, is a biologically meaningful parameter that measures the frequency with which a different phenomenon—cannibalism—occurs. Clearly, the statistical model alone is inadequate for determining the intensity of cannibalism.
The parameters of the statistical model are \( n \) and \( \alpha \). The number of replications \( n \), which represents the number of cups employed by the experimentalist, is fixed and known. The parameter \( \alpha \) is unknown but estimable. Indeed, the observable \( \hat{\alpha} = (X_0/n, X_1/n, X_2/n) \) is an unbiased, maximum likelihood estimator of \( \alpha = (\alpha_0, \alpha_1, \alpha_2) \) for \( \alpha \) in \( A \) (2.4). This estimator \( \hat{\alpha} \) presumably should tell us all we need to know at the elementary level [i.e., using the elementary experiment model \( (U, p^U) \)] where the biological phenomena occur, but this is not the case.

Suppose the value of \( \alpha \) were known. Then the probability of zero, one, or two pupae per cup would be known. That is, the average relative frequency of each of the elementary outcomes \( (1,1) \) would be known. Moreover, by (2.18), the average relative frequency of pupation \( r \) would be known. But what would this tell us about the existence or nonexistence of cannibalism as a significant mortality factor? What would this tell us about the intensity of (i.e., propensity for) cannibalism when it is significant? Obviously, a higher propensity for cannibalism would result in greater mortality, which would result, on the average, in a lower observed frequency of pupation \( p \). But even if there were no cannibalism at all, there still would be some mortality, which also would result, on the average, in a lower observed frequency of pupation \( p \). The average relative frequency of pupation \( r \) is indeed a biologically meaningful parameter. It measures the frequency with which a biological phenomenon--pupation--occurs. What is sought, however, is a biologically meaningful parameter that measures the frequency with which a different phenomenon--cannibalism--occurs. Clearly, the statistical model alone is inadequate for determining the intensity of cannibalism.
A mathematical model for the cannibalism phenomenon is required. We need a biological model, or theory, for cannibalism, and a mathematical statement of that theory. The biomathematical model is the subject of the next chapter.
3. BIOLOGICAL MODEL

What inferences can be drawn from the data of table 1.1 regarding the propensity for cannibalism in a given strain? And exactly what is meant by "the propensity for cannibalism" in this experimental situation? It is well known that for each strain, the data $X_0$, $X_1$, $X_2$, and $\bar{r}$ (table 2.1) are consistent, unbiased estimators for the probabilities $a_0$, $a_1$, $a_2$ of each of the elementary outcomes (1.1) and the relative frequency of pupation $r$ (2.18). What is needed is a theory that precisely defines "the propensity for cannibalism" and that specifies how the propensity for cannibalism is related to the probabilities $a_0$, $a_1$, $a_2$ of each of the elementary outcomes (1.1) in the same way that (2.18) defines $r$ in terms of $U$ and thereby specifies the relationship between the biological parameter $r$ and the statistical parameter $a$.

What has been observed is either survival (pupation) or mortality (lack of pupation). The problem is that (in this experimental situation) cannibalism is only one of two biological phenomena that result in larval mortality and thereby decrease the number $U$ of larvae which survive to pupate in any particular cup. As mentioned at the conclusion of chapter 2, even when larvae are raised one per cup, not all survive to pupate. Therefore the phenomenon of noninteractive, noncannibalistic, independent, or "random" mortality must be considered in addition to the phenomenon of interactive mortality, cannibalism, or intraspecific predation.
Unfortunately, when it is observed that exactly one pupa has developed, the experimentalist cannot discern which of the two phenomena has occurred—random mortality or cannibalism.

Suppose {1 pupa in cup} is the elementary outcome. If an intact carcass is present, that larva may have died either randomly or as the result of an invisible, perhaps internal, wound. Even if a visible wound is found on the carcass, that still does not imply that cannibalism was the cause of mortality; the animal could have died randomly, and subsequently have been encountered and bitten. On the other hand, if no carcass is found, it likewise is not known whether mortality was interactive, or whether death was noninteractive and the dead larva was subsequently consumed. Clearly, the two causes of mortality are indistinguishable in the present laboratory experiment.

For this reason, the two biological phenomena that must be modeled are the two potential causes of mortality—random death and cannibalism.

The mathematical model for the two biological phenomena is based on two biological assumptions:

1. during a short interval of time immediately following hatch—the neonatal period—cannibalistic mortality is insignificant and, therefore, may be ignored; and

* Justification, validity, and limitations of the assumptions will be discussed after the model is formulated.
(2) during the remainder of the larval period—the post-neonatal period—when cannibalism is possible, noncannibalistic mortality is insignificant and, therefore, may be ignored, as well.

As a result of assumptions 1 and 2, the elementary experiment \((U, p_U)\) of observing the development of two larvae in one cup is modeled as a compound experiment (Ash, 1972, p. 96). A compound experiment is the combination of two experiments which are performed consecutively such that the execution, and thereby the outcome, of the second is influenced by the outcome of the first. Here, the former experiment, the compounding experiment, consists of observing the number of larvae \(V\) that survive the neonatal period. The latter experiment, the compounded experiment, consists of observing the number of larvae \(U\) which survive the post-neonatal period (i.e., pupate), given that \(V\) have survived the neonatal period. This terminology for experiments is analogous to that which is commonly employed to classify compound or mixture distributions (Johnson and Kotz, 1969, pp. 27-28, 183-215). The pdf \(p_{U|V}\), which gives the conditional probability that \(U\) larvae pupate given that \(V\) survive the neonatal period, is said to be compounded by the compounding pdf \(p_{V}(V)\) because \(p_{U|V}\) contains the parameter \(V\) which is a random variable. The compound distribution is \(p_{U}\), the expectation of \(p_{U|V}\) with respect to \(p_{V}\).

In this situation, \((V, p_{V})\) and \((U, p_{U|V})\) represent "thought" experiments (van der Vaart, 1962) which cannot be performed directly because \(V\) is not an observable. Their utility is the resulting formulation of the elementary experiment \((U, p_{U})\) (which, because \(U\) is an...
observable, represents an actual experiment) in terms of the underlying biological phenomena—random neonate mortality and post-neonatal cannibalism.

To model the former experiment, which conceptually portrays the number of larvae $V$ that survive the neonatal period, let $1_a$ and $1_b$ be indicator random variables for the fate of the two cohabiting neonates, respectively. That is, let $1_a = 1$ indicate that larva 'a' survives the neonatal period, and let $1_a = 0$ indicate that larva 'a' does not survive, and likewise for larva 'b'. Letting $V$ represent the number of viable larvae which survive the neonatal period, we have $V = 1_a + 1_b$.

The probability measure $P^V$ is determined by the following two biological assumptions, which delineate the meaning of random mortality:

(3) during the initial neonatal period, the fate of each larva is stochastically independent of the fate of the other larva; and,

(4) during the initial neonatal period, the fates of the individual larvae are stochastically identical.

Assumptions 3 and 4 have the mathematical form

(3') $1_a$ and $1_b$ are stochastically independent random variables:

$$P\{1_a = 1, 1_b = 1\} = P\{1_a = 1\} P\{1_b = 1\}. \quad (3.1a)$$

(4') $1_a$ and $1_b$ are identically distributed random variables:

$$P\{1_a = 1\} = P\{1_b = 1\}. \quad (3.1b)$$
Thus, if $p$ is the probability that an individual larva survives the neonatal period, independently of the fate of the other larvae during this period, then

$$p = P\{1_a = 1\} = P\{1_b = 1\}$$

$$= P\{1_a = 1 \mid 1_b = 0\} = P\{1_a = 1 \mid 1_b = 1\},$$

(3.2)

and likewise for $1_a$ and $1_b$ interchanged. This is a rigorous mathematical definition of a biological parameter, $p$.

As a consequence of assumptions 3 and 4, $V = 1_a + 1_b$ is the sum of two independent, identically distributed (iid), Bernoulli random variables. Therefore $V$ is binomially distributed with parameters 2 and $p$:

$$P\{V = 0\} = P^V(0) = (1 - p)^2,$$  \hspace{1cm} (3.3a)

$$P\{V = 1\} = P^V(1) = 2p(1 - p),$$  \hspace{1cm} (3.3b)

$$P\{V = 2\} = P^V(2) = p^2.$$  \hspace{1cm} (3.3c)

$(V, P^V)$ models the outcome of the neonatal period under biological assumptions 1 through 4.

The latter half of the compound experiment, represented by $(U, P^{U\mid V})$, models the outcome of the post-neonatal period, during which, by assumption 2, cannibalism is possible, while noncannibalistic mortality is not possible. If less than two neonates survive, that is if $V < 2$, neither cannibalism nor random death can occur during the post-neonatal period; therefore

$$P\{U = u \mid V = v\} = P^{U\mid V}(u) = 1_{\{v\}}(u); \hspace{0.5cm} v = 0, 1; \hspace{0.5cm} u = 0, 1, 2.$$
That is,

\[ P(U = V \mid V < 2) = 1, \quad P(U \neq V \mid V < 2) = 0. \]

Cannibalism can occur if and only if \( V = 2 \), that is, if and only if both larvae survive the initial neonatal period. The propensity for cannibalism \((1 - q)\) is defined precisely by

\[
\begin{align*}
    p_{U|2}(0) &= 0, \quad (3.4a) \\
    p_{U|2}(1) &= 1 - q, \quad (3.4b) \\
    p_{U|2}(2) &= q. \quad (3.4c)
\end{align*}
\]

A fifth biological assumption is implicit in the definition of the propensity for cannibalism (3.4):

(5) Only one larva can die (as the result of cannibalism) during the post-neonatal period.

The compounded distribution \( P_{U|V} \) which represents the law for the latter experiment which is thought of as being executed during the post-neonatal period, is presented in table 3.1.

For each value of \( v \), \( P_{U|v} \) represents the pdf which governs (or describes) the survival of larvae during the post-neonatal period, conditionally upon the number \( v \) that survive the initial neonatal period.

At this point the modeling is essentially completed.
Table 3.1: $p_{U|V}(u)$, the conditional pdf of $U$ given $V = v$, which models the outcome of the post-neonatal period depending on the number $v$ of neonates which survive random mortality.

<table>
<thead>
<tr>
<th>$v$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>(1 - $q$)</td>
<td>$q$</td>
</tr>
</tbody>
</table>

As already explained, since conceptually the two experiments are performed consecutively, it is natural to view the distribution $p_U$ of $U$ as the expectation with respect to $p_V$ of the distribution $p_{U|V}$ where the parameter $V$ is a random variable:

$$p_U = E_{p_V} p_{U|V} = \sum_{v=0}^{2} p_{U|V}(u) p_V(v).$$

(3.5)

That is,

$$p_U(u) = \sum_{v=0}^{2} p_{V|U}(u, v) = \sum_{v=0}^{2} p_{U|V}(u) p_V(v), \quad u = 0, 1, 2,$$

or

$$p(U = u) = \sum_{v=0}^{2} p(V = v, U = u) = \sum_{v=0}^{2} p(U = u | V = v) p(V = v).$$

$p_U$ equivalently can be derived as follows. $p_V$ (3.3) and $p_{U|V}$ (table 3.1) determine the joint pdf of $V$ and $U$, by

$$p(V = v, U = u) = p_{V|U}(v, u) = p_{U|V}(u) p_V(v),$$

$v = 0, 1, 2, \quad u = 0, 1, 2$

(see table 3.2).
Table 3.2: $p^{VU}(v, u)$ the joint pdf of $V$ and $U$, which models the outcome of both developmental periods.

<table>
<thead>
<tr>
<th>$u$</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$(1 - p)^2$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>$2p(1 - p)$</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$p^2(1 - q)$</td>
<td>$p^2q$</td>
</tr>
</tbody>
</table>

The biological formulation of $p^U$ follows directly from $p^{VU}$ as the marginal pdf of $U$:

$$P\{U = u\} = p^U(u) = \sum_{v=0}^{2} p^{VU}(v, u).$$

The result is a biomathematical model for the elementary experiment $(U, p^U)$, the outcome of the entire larval period,

$$P\{U = 0\} = p^U(0) = (1 - p)^2, \quad (3.6a)$$

$$P\{U = 1\} = p^U(1) = 2p(1 - p) + p^2(1 - q), \quad (3.6b)$$

$$P\{U = 2\} = p^U(2) = p^2q. \quad (3.6c)$$

This is the model sought all along. $p^U$, the probability of each of the elementary events (1.1), has finally been defined in terms of biological parameters—namely, the probability $p$ (3.2) that an individual larva survives noninteractive (random) mortality during the neonatal period (3.2), and the probability $q$ (3.4) that both larvae survive cannibalism during the post-neonatal period, given the event that both larvae have survived random mortality during the initial neonatal period (3.4).
A rigorous definition for "the propensity for cannibalism" has also been obtained:

\[ P(U = 1 \mid v = 2) = P(U|2)(1) = (1 - q). \]  

(3.7)

The relationship between the statistical parameter \( a \) and the biological parameter \( (p, q) \) follows from (2.7) and (3.6):

\[
\begin{align*}
    a_0 &= (1 - p)^2, \\
    a_1 &= 2p(1 - p) + p^2(1 - q), \\
    a_2 &= p^2q.
\end{align*}
\]

(3.8a) \hspace{1cm} (3.8b) \hspace{1cm} (3.8c)

The inverse relationship is

\[
\begin{align*}
    p &= 1 - a_0^{1/2} \\
    q &= a_2 / p^2 = a_2 / (1 - a_0^{1/2})^2.
\end{align*}
\]

(3.9) \hspace{1cm} (3.10)

These equations constitute the underlying biological model. They will be shown to reveal which statistical hypotheses should be tested in order to determine whether or not cannibalism is a mortality factor, as well as which function of the statistical parameters must be estimated in order to quantify cannibalism.

To determine whether or not cannibalism is significant, we shall test the biological hypothesis that there is no cannibalism

\[ H_0: \ q = 1, \quad \]  

(3.11)

against the biological alternative that cannibalism does occur,

\[ H_1: \ q < 1. \quad \]  

(3.12)
That these are the correct biological hypotheses is clear from the definition of $q$ (3.4). The probability $p$ that an individual survives random mortality during the neonatal period is an auxiliary parameter in this context.

The statistical hypotheses follow from the relationship between the biological parameters $p$, $q$, and the statistical parameters $a_0$, $a_1$, $a_2$. The correct statistical hypotheses are

$H_0$: $a_0 a_2 / a_1^2 = 1/4$, \hspace{1cm} (3.13)

$H_1$: $a_0 a_2 / a_1^2 < 1/4$. \hspace{1cm} (3.14)

To prove that these expressions are the correct hypotheses it will suffice to show that (3.11) is equivalent to (3.13), and that (3.12) is equivalent to (3.14).

To see that (3.11) is equivalent to (3.13), (3.10) is substituted into (3.11) to obtain

$$
a_2 / (1 - a_0^{1/2})^2 = 1.
$$

Multiplying by $(1 - a_0^{1/2})^2$ we get $a_2 = (1 - a_0^{1/2})^2$. Taking the positive square root yields $a_2^{1/2} = (1 - a_0^{1/2})$, subtracting yields $a_0^{1/2} + a_2^{1/2} = 1$, squaring both sides yields $a_0 + 2(a_0 a_2)^{1/2} + a_2 = 1$, and subtracting yields $2(a_0 a_2)^{1/2} = 1 - a_0 - a_2$. By (2.3) we have $2(a_0 a_2)^{1/2} = a_1$. Squaring both sides and dividing by 4$a_1^2$ (which is always positive for a in A) shows that (3.11) implies (3.13).

The computations work in reverse order; therefore (3.13) implies (3.11). Moreover, starting with (3.12) and substituting as above, the same computations yield (3.14), because for each $i$, $a_i$ and $(1 - a_i)$ are
positive, and the direction of the inequality is not reversed by division, or multiplication, or the taking of a positive square root.

For the estimation of the propensity for cannibalism \((1 - q)\), an estimate of the biological parameter \(q\) is required. Recall that \(\hat{a}\) as given in (2.16) is an unbiased, maximum likelihood estimator for the statistical parameter \(a\), for \(a\) in \(A\) (2.4). Since \(q\) is a continuous function of \(a\) (3.10), it follows that

\[
\hat{q} = \frac{\hat{a}_2}{(1 - \hat{a}_0^{1/2})^2}
\]

is a maximum likelihood estimator for \(q\). Likewise, it follows from (3.9) that

\[
\hat{p} = (1 - \hat{a}_0^{1/2})
\]

is a maximum likelihood estimator for the biological parameter \(p\).

These point estimates of \(p\) and \(q\) are listed in table 3.1 and graphed in figure 3.1.

Figure 3.2, where lines of constant \(p\) and \(q\) are graphed in \((a_2, a_0)\)-space, illustrates how the biological model confers meaning to the statistical model.

Thus a biomathematical model has been constructed which reflects biological conjectures regarding the nature of the phenomena. In addition to determining the analysis of the laboratory experiment, the detailed biomathematical model serves a broader purpose. As a result of the modeling process, each biological assumption has been recognized explicitly and is thereby exposed to critical scrutiny. The scope of inferences based on the experiment is delineated, and additional experiments which might be performed in the future are suggested.
Figure 3.1: Point estimates \( \hat{p} \) and \( \hat{q} \) of the biological parameters \( p \) and \( q \), the data of table 3.1. Apparently \( \hat{p} \) and \( \hat{q} \) are uncorrelated in the sample of eleven strains being investigated here (corr. coef. -0.0534; explanation in text).
Figure 3.2: Lines of constant $p$ and $q$ in parameter space $A$. Horizontal lines are $\beta = 1/4$ (upper line) and $\beta = 1/2$ (lower line). Curves are, from top to bottom, $q = 1, 3/4, 1/2, 1/4$. 
Table 3.1: Point estimates of the biological parameters.

<table>
<thead>
<tr>
<th>Strain</th>
<th>n</th>
<th>$\hat{\beta}$</th>
<th>$\hat{q}$</th>
<th>$\hat{p}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>COL</td>
<td>166</td>
<td>0.76715</td>
<td>0.276369</td>
<td>0.554215</td>
</tr>
<tr>
<td>LDB</td>
<td>337</td>
<td>0.90565</td>
<td>0.390727</td>
<td>0.655787</td>
</tr>
<tr>
<td>WES</td>
<td>128</td>
<td>0.65767</td>
<td>0.523802</td>
<td>0.554651</td>
</tr>
<tr>
<td>BRO</td>
<td>512</td>
<td>0.81778</td>
<td>0.040887</td>
<td>0.497070</td>
</tr>
<tr>
<td>FAR</td>
<td>138</td>
<td>0.91487</td>
<td>0.207783</td>
<td>0.583333</td>
</tr>
<tr>
<td>RIV</td>
<td>55</td>
<td>1.00000</td>
<td>0.109091</td>
<td>0.554545</td>
</tr>
<tr>
<td>TAL</td>
<td>74</td>
<td>0.88375</td>
<td>0.363351</td>
<td>0.635134</td>
</tr>
<tr>
<td>ART</td>
<td>225</td>
<td>0.67340</td>
<td>0.284228</td>
<td>0.511110</td>
</tr>
<tr>
<td>ARC</td>
<td>228</td>
<td>0.74351</td>
<td>0.507781</td>
<td>0.607459</td>
</tr>
<tr>
<td>CLA</td>
<td>281</td>
<td>0.85388</td>
<td>0.942023</td>
<td>0.832744</td>
</tr>
<tr>
<td>WHI</td>
<td>1073</td>
<td>0.92522</td>
<td>0.816526</td>
<td>0.846690</td>
</tr>
</tbody>
</table>

Assumptions 1 and 2 eliminate confounding between random and cannibalistic mortality. They also are reasonable biologically.

The first assumption is easy to rationalize. In order for cannibalism to occur, the larvae must first encounter one another. This is unlikely for neonates, as they are very small relative to the experimental cup and the initial food supply. Unless they are placed on top of each other initially, it is unlikely that they will come into contact.

The second assumption is rationalized by assuming that under laboratory conditions, those larvae that would fail to pupate even in the absence of cannibalism as a significant mortality factor, do not develop beyond the initial neonatal period, and that such larvae do not
develop sufficiently to become intraspecific predators during the post-neonatal period.

The mathematical form 3' and 4' of the biological assumptions 3 and 4 reveal a striking similarity to the one explicit assumption of the statistical model: that the n replications $U_1, ..., U_n$ constitute a random sample. Both the biological and the statistical assumptions are that certain random variables are stochastically independent and identically distributed. In this chapter the assumption is a conjecture regarding the nature of a biological phenomenon (neonate mortality), whereas in chapter 2 it was a conjecture regarding an experimental procedure (replication). It is in this sense that one model is biological and the other is statistical, albeit both are expressed in mathematical and stochastic terms.

Assumption 5 may seem trivial at first glance, but to believe so is anthropocentric. Anyone who has vented his sadistic tendencies on insects knows that they do not necessarily cease to function when dismembered or otherwise mortally wounded. It is possible that two larvae encountering one another head on would mortally wound one another, thereby causing two deaths by intraspecific predation. It is assumed here that such encounters are rare enough to be discounted.

Biological parameters are defined rigorously in the model. Parameter $p$ is the probability that any single larva survives the neonatal period, and $q$ is the probability that both larvae pupate in the event that both have survived the neonatal period.

Quite logically, cannibalism is not deemed possible in the event that only one viable larva is present. This underlines the statement
made in chapter 1 that cannibalism is a group phenomenon. Cannibalism can occur only if two or more larvae are present on the same contact site.

Also, \((1 - q)\) is the probability that at least one larva fails to pupate, given that both have survived the initial neonatal period. Since cannibalism is the only cause of mortality during the post-neonatal period (assumption 2) and since cannibalism occurs only during the post-neonatal period (assumption 1), \((1 - q)\) is the propensity for cannibalism.

Thus, the propensity for cannibalism has been defined mathematically, that is precisely, under the five assumptions of the biological model by (3.7) which follows from (3.4). We have arrived at this particular definition because we are interested in the propensity for cannibalism primarily as a phenotypic trait exclusive of other mortality factors (here, random death). Were we interested in the propensity for cannibalism primarily as a mortality factor affecting population dynamics, we might have needed an entirely different definition of the term "propensity for cannibalism". This point exemplifies the fact that the selection of a model is necessarily influenced by the purpose of the investigator (Spedding, 1980).

It is also clear that we could not have used a more complex model. For example, assuming that there actually is a distinct neonatal period as described in assumption 1, it would be nice to estimate the length of that period. But to estimate a third parameter would require more information than is present in the domain of \(P^u\) (or \(P^x\)). Parameter space \(A\) (2.4) has only two "degrees of freedom" (Neyman and Pearson,
in the sense that if any two components (e.g., \(a_0\) and \(a_2\)) are fixed, then the remaining component \(a_1\) is fixed as well (\(a_1 = 1 - a_0 - a_2\)). Thus, the biomathematical model can have only two independent parameters, here \(p\) and \(q\).

Even though, as was explained in chapter 1, the laboratory experiment is too simple to produce any far reaching inferences regarding population dynamics on a contact site in the field (because populations aren't limited to two there), it is relevant to ask whether this model of two mortality factors reasonably describes population dynamics under the experimental conditions now under discussion. For this simple model, it is obvious that this is so, but the technique of analysis is worth practicing because the same method would be used in a more complex model for a more elaborate experiment.

To this end, the expected individual survivorship (relative frequency of pupation) \(r\) as defined in (2.18) is contemplated as a function of \(p\) and \(q\):

\[
r = \frac{a_1 + 2a_2}{2} = \frac{[2p(1-p) + p^2(1-q) + 2p^2q]}{2} = p - p^2(1-q)/2.
\]

The relationship between average survivorship \(r\), average neonatal survivorship \(p\), and the average propensity for cannibalism \((1 - q)\) is easily determined by analysing the surface (3.15), which is graphed in figure 3.3.
Figure 3.3: Survivorship r as a function of p and q.
Elementary methods suffice to show that (1) if there is no cannibalism \((q = 1)\), then the average proportion which pupate is the average number of neonates that survive random mortality \((r = p)\). (2) if there is no neonatal mortality \((p = 1)\) then the average survivorship is \(r = (1 + q)/2\)—on the average at least one larva survives (cf. assumption 5), and one dies with probability \((1 - q)\), the propensity for cannibalism; (3) The maximum average survivorship \((r = 1)\) occurs when there is no mortality \((p = 1, q = 1)\), and otherwise survivorship \(r\) is an increasing function of \(p\) and \(q\).

The model is self-consistent and reasonable in this regard; any other results would be unacceptable.

It is also worthwhile to look at lines of constant survivorship \(r\) in the \((p, q)\) plane (figure 3.4). The point is that a wide range of combinations of \(p\) and \(q\) give the same survivorship. This would be of considerable importance in a more complex model capable of simulating conditions in the field (i.e. populations larger than two). For a mass release program it would be well worth studying the different combinations of \(p\), \(q\), and \(r\) that are available. In this regard it is important to observe in figure 3.1 that \(\hat{p}\) and \(\hat{q}\) are apparently uncorrelated for the eleven strains under investigation here (the correlation coefficient is \(-0.0534\)). The same observation supports the proposition that \(p\) and \(q\) measure distinct genotypic traits rather than different expressions of the same trait (e.g., larval vigor in general).

Ideas for future experiments follow from the compound nature of the model experiment. In the sense of equation (3.5) the experiment \((U, p^U)\) has been viewed as an "average experiment", the weighted average of the
Figure 3.4: Contours of constant survivorship projected onto the p q plane. Values shown are $r = 3/8, 1/2, 5/8, 3/4$ (bottom to top).
three experiments \((U, p^U|V), v = 0, 1, 2\). Unfortunately, the weights \(p^V(v), v = 0, 1, 2\), assigned to each of the experiments \((U, p^U|V), v = 0, 1, 2\), must be estimated from the same experiment \((U, P^U)\).

Obviously, it would be desirable to perform each of the two conceptual experiments separately, and thereby to estimate \(p\) and \(q\) independently of each other.

Actually executing the first thought experiment, observing the outcome \(V\) of the neonatal period, would produce an independent estimate of \(p\), and thereby the weights \(p^V(v), v = 0, 1, 2\). This could be done simply by raising \(n\) larvae one per cup in \(n\) cups and observing the number of larvae \(Y\) that eventually pupate. The relative frequency of pupation \(\hat{p} = Y/n\) would be an unbiased estimator of \(p\). If this experiment were done concurrently with a two per cup experiment then the two independent estimates of \(p\) could be compared. If the two estimates were significantly different, the model would be judged invalid.

In addition, raising the larvae one per cup would provide a test of assumption 1 specifically. Among those larvae that fail to pupate, it could be observed at which instar \(i\) they ceased to develop, \(i = 1, 2, \ldots, 6\). If assumption 1 were correct, most larvae that fail to pupate would cease to develop during the first few instars, that is for a value of \(i\), say \(i^*_1\), toward the lower end of the range. Thus \(i^*_1\) or \(i^*_1 + 1\), would be an estimate of the end of the neonatal period. On the other hand, if it were found that larvae ceased to develop at all stages assumption 2 would be invalidated.

Raising larvae one per cup would thus provide a realization \(V\) of the compounding experiment \((V, P^V)\). To directly estimate the propensity
for cannibalism \((1 - q)\) independently of \(p\) would require a realization of the one compounded experiment \(p^{i_2}\) where cannibalism can occur (i.e., the one which depends on \(q\)). To do this would require that two surviving, equal aged larvae be placed in a cup before the beginning of the post-neonatal period, say instar \(i_2\), and after the end of the neonatal period \((i_1\) above). Since the interval \((i_1, i_2)\) is completely unknown, the following complex experiment would be required.

A large stock of neonates would be raised one per cup under identical conditions. From this stock six groups of \(2n\) larvae, one group for each instar, would be selected and placed by pairs in cups. The result would be six comparable experiments with six different starting times \(i = 1, 2, \ldots, 6\). The average survivorship (rate of pupation) in each group obviously would be an increasing function of the starting time \(i\), provided cannibalism contributes significantly to mortality. If assumption 2 is correct, however, there should be a jump at some time \(i_2\) toward the upper end of the range \([1, 6]\). This time \(i_2\) would be an estimate of the starting time of the post-neonatal period.

If this experiment and the one per cup experiment were performed simultaneously, \(i_1\) and \(i_2\) could be compared. If \(i_2\) were significantly less than \(i_1\), the model proposed would be judged invalid.

The simultaneous performance of these two experiments would also provide a test of assumption 5. If assumption 5 is true, multiple deaths (neither larvae pupates) among pairs joined later than the estimate of \(i_1\), the estimated end of the neonatal period, ought to be extremely rare.
One test of the validity of this model is possible without further experimentation. This derives from the fact that the biomathematical model implies that the domain \( A'' \) of the statistical parameter \( a \) is strictly smaller than the trinomial parameter space \( A \) (2.4). As defined by the model, the parameters \( p \) (3.2) and \( q \) (3.4) must be such that \( 0 < p < 1 \), and \( 0 < q \leq 1 \), and therefore the statistical parameter \( a \) must be such that \( a_0 a_2 / a_1^2 \leq 1/4 \) [cf. (3.13) (3.14)]. Thus, the biomathematical model claims a smaller parameter space

\[
A'' = \{a \in A : a_0 a_2 / a_1^2 \leq 1/4\}
\]

than the trinomial model alone (figure 3.5).

Since the observable statistic \( \hat{a} \) (2.16) is an unbiased estimator for \( a \), the validity of the biological model would be questionable for any strain with a realization of \( \hat{a} \) (computed by (2.16) from the data of table 1.1) very "far" outside of \( A'' \). Precisely how far is a problem in the realm of statistical inference, which need not be considered presently because all eleven values of \( \hat{a} \) lie in \( A'' \), as illustrated in figure 3.5.
Figure 3.5: Parameter space A. The subregion specified by the null hypothesis (3.13) is the parabola. The interior of the region below the parabola is $A_1$, the subregion specified by the alternative hypothesis (3.13). The crosshatched region is the subregion of A which is excluded by the biological model for canibalism. (The symbol '+' marks the observed values of $\hat{\lambda}$.)
4. HYPOTHESIS TESTING

In this chapter we shall construct an optimal test of the null hypothesis (3.13) against the alternative hypothesis (3.14). The test derived will be uniformly most powerful among all unbiased tests of the same or smaller size.

Before deriving the test, let us review some relevant definitions and lemmas of the Neyman Pearson theory of hypothesis testing. Those cited are simplified versions of more general results (Lehmann, 1959) adapted to the discrete problem presented here.

Suppose an experiment has been performed consisting of observing a random variable \( U \), where the unknown pdf of \( U \) is assumed to be a member of the known parametric family \( \{p_\alpha \colon \alpha \in A\} \). Suppose further that we want to test a null hypothesis

\[
H_0: \alpha \in A_0, \tag{4.1}
\]

against an alternative hypothesis

\[
H_1: \alpha \in A_1. \tag{4.2}
\]

Let \( X \) represent a discrete vector-valued random variable and suppose that \( X \) is sufficient for \( \alpha \) in \( A \), that is, \( P_{\alpha}^U[U = u \mid X = x] \) is a constant function of \( \alpha \) for each fixed \( u \) and \( x \). Then, we need consider only tests based on \( X \) or some function of \( X \).

According to the randomized version of the Neyman-Pearson scheme, a test of (4.1) against (4.2) is completely characterized by a critical function \( f \) mapping statistic space (\( X \)-space) into the interval \([0, 1]\).
For each point \( x \) in statistic space, \( f(x) \) connotes the conditional probability of rejecting the null hypothesis given the event \( \{ X = x \} \) is observed.

The quality of the test \( f \) is judged by its "power to discriminate between the false and the true" (Neyman and Pearson, 1933). This is determined by the \([0, 1]\)-valued function \( B_f \) defined on parameter space \( \Gamma \) by

\[
B_f(\theta) = \mathbb{E}_\theta f(X), \quad \theta \in \Gamma,
\]

the power function of the test \( f \). Since \( f(X) \) is the conditional probability of rejecting the null hypothesis given \( X \) is observed, \( B_f(\theta) = \mathbb{E}_\theta f(X) \) represents the probability of rejecting the null hypothesis as a function of \( \theta \) in \( \Gamma \).

The maximum (over all \( \theta \) in \( \Gamma_0 \)) probability of rejecting the null hypothesis \( H_0 \) when it is true, that is,

\[
w = \sup_{\Gamma_0} B_f(\theta) = \sup_{\Gamma_0} \mathbb{E}_\theta f(X),
\]

is called the size \( w \) of the test \( f \). For each \( \theta \) in \( \Gamma_1 \), \( B_f(\theta) \) is called the power of the test \( f \) against the alternative \( \theta \). For each alternative (i.e., for each \( \theta \) in \( \Gamma_1 \)) the ideal test would be that with the greatest power among all those that do not exceed a certain size, that is the test with the greatest probability of rejecting the null hypothesis when it is false, among all those for which the probability of rejecting \( H_0 \) when it is true does not exceed \( w \).

A size \( w \) test \( f \) is said to be uniformly most powerful size \( w \) (UMP-\( w \)) if for every size \( w \) test \( g \), \( B_f(\theta) \geq B_g(\theta) \) for all \( \theta \) in \( \Gamma_1 \).
A size \( w \) test \( f \) is said to be \textbf{unbiased} if \( B_f(a) \geq w \) for all \( a \) in \( A_1 \).

A size \( w \) test \( f \) is said to be \textbf{uniformly most powerful unbiased} size \( w \) (UMPU-\( w \)) if \( f \) is UMP among all unbiased size \( w \) tests.

If \( B_f(a) \) is a constant function of \( a \) on a subset \( A' \) of \( A \), say \( B_f(a) = k \) for all \( a \) in \( A' \), then \( f \) is said to be \textbf{\( k \)-similar} on \( A' \). When \( k \) is the size of the test and \( A' \) is the common boundary between \( A_0 \) and \( A_1 \), we sometimes shall say "\( f \) is similar" omitting the qualifiers.

**Lemma 4.1:** Let \( B_f \) be continuous at every point in parameter space, and suppose \( f \) is an unbiased size \( w \) test. Then \( f \) is necessarily \( w \)-similar on the common boundary of \( A_0 \) and \( A_1 \) (Lehman, 1959, p. 125).

Indeed, \( f \) being of size \( w \) implies \( B_f(a) \leq w \) for all \( a \) in \( A_0 \). If \( f \) is also unbiased, then \( B_f(a) \geq w \) for all \( a \) in \( A_1 \). By continuity, therefore, \( B_f(a) = w \) on the boundary between \( A_0 \) and \( A_1 \), that is, \( f \) is \( w \)-similar on the common boundary of \( A_0 \) and \( A_1 \).

**Lemma 4.2:** If \( B_f \) is continuous and \( f \) is UMP-\( w \) among all tests that are \( w \)-similar on the common boundary between \( A_0 \) and \( A_1 \), then \( f \) is UMPU-\( w \) (Lehmann, 1959, p. 126, Lemma 1).

Let \( A' \) denote the common boundary between \( A_0 \) and \( A_1 \). By lemma 4.1, if a size \( w \) test is unbiased and its power function is continuous, it must be \( w \)-similar on \( A' \). Therefore, since \( f \) is UMP-\( w \) among all tests that are \( w \)-similar on \( A' \), \( f \) must be at least as powerful as every
unbiased size \( w \) test. Also because \( f \) is UMP-\( w \) among all tests \( w \)-similar on \( A' \), \( f \) must be at least as powerful as the constant size \( w \) test \( g \) (i.e., \( g(x) = w \) for all \( x \)) which is \( w \)-similar on \( A' \), that is \( B_f(a) \geq B_g = w \) for all \( a \) in \( A' \). Hence, the test \( f \) is unbiased.

Let \( S \) be a statistic defined on \( X \)-space such that \( S \) is sufficient for \( a \) in \( A' \). A size \( w \) test \( f \) is said to have Neyman structure with respect to \( A' \) and the (discrete) sufficient statistic \( S \) if for all feasible values \( s \) of \( S \),

\[
E_{a}[f(X) \mid S = s] = w, \quad \text{for all } a \text{ in } A', \tag{4.4}
\]

where \( A' \) denotes the common boundary between \( A_0 \) and \( A_1 \).

A discrete statistic \( S \) is said to be complete on a region \( A' \) of a parameter space if for every function \( h \),

\[
E_{a}h(S) = 0 \quad \text{for every } a \text{ in } A'
\]

implies

\[
h(s) = 0, \quad \text{for all } s \text{ such that } P(S = s) > 0.
\]

**Lemma 4.3:** Let \( S \) be sufficient and complete for \( a \) in \( A' \). A test \( f \) is \( w \)-similar on the common boundary \( A' \) between \( A_0 \) and \( A_1 \) if and only if \( f \) has Neyman structure with respect to \( S \) (Lehmann, 1959, p. 134, theorem 2).

That Neyman structure implies similarity is obvious. If a size \( w \) test has Neyman structure with respect to \( S \), then by definition, (4.4) is true. In fact, since

\[
E_{a \in A'} \{ E_{a \in A'} [f(X) \mid S = s] \} = E_{a \in A'} f(X) \tag{4.5}
\]

it follows that

\[
E_{a \in A'} f(X) = w \text{ for all } a \in A', \tag{4.6}
\]

that is \( f \) is \( w \)-similar on \( A' \).

To prove the converse requires completeness. It will suffice to show that (4.6) implies (4.4). Assuming (4.6), it follows from (4.5) that

\[
E_{a \in A'} \{ E_{a \in A'} [f(X) \mid S = s] \} = w, \ a \in A', \text{ and furthermore, that}
\]

\[
E_{a \in A'} \{ E_{a \in A'} [f(X) \mid S = s] - w \} = 0. \tag{4.7}
\]

Since \( S \) is sufficient for \( a \in A' \), \( E_{a \in A'} [f(X) \mid S = s] \) is a constant function of \( a \) on \( A' \). Therefore \( \{ E_{a \in A'} [f(X) \mid S = s] - w \} \) is a function of \( s \) only. Since \( S \) is complete on \( A' \), it follows from (4.7) that

\[
\{ E_{a \in A'} [f(X) \mid S = s] - w \} = 0, \ a \in A', \text{ that is, } E_{a \in A'} [f(X) \mid S = s] = w \text{ for all } a \in A', \text{ the desired result.}
\]

In the present problem parameter space \( A \) is the two dimensional open subset of Euclidean 3-space \( \mathbb{R}^3 \) defined by (2.4). The region of parameter space \( A \) specified by \( H_0 \) (3.13) is

\[
A_0 = \{ a \in A : a_0 a_2 / a_1^2 = 1/4 \},
\]
a one dimensional subregion of $A$, and the region specified by $H_1$ (3.14) is

$$A_1 = \{a \in A : a_0 a_2 / a_1^2 < 1/4\},$$

a two dimensional subregion of $A$ (see figure 3.5).

Also, it has already been shown that the trinomially distributed random variable $X$ defined by (2.11) is sufficient for $A$ (2.8). Consequently, the domain of the test function sought here is $\mathbb{X}$-space $\mathbb{W}^X$, and our mission is to report on the family of trinomial pdf's

$$\{P_a^X : a \in A\}$$

(3.16). Specifically, we are to find an optimal critical function $f$ to decide between the two families $\{P_a^X : a \in A_0\}$ and $\{P_a^X : a \in A_1\}$ (or equivalently, between $A_0$ and $A_1$).

Our first priority is to avoid the mistake of reporting that cannibalism occurs when in fact it does not. For this reason (3.13) and (3.14) have been designated the null and alternative hypotheses, respectively. We now choose a small number $w$ for the size of our test; that is, the probability of making the above mistake shall be no larger than $w$. Our second priority is to maximize the probability of detecting cannibalism when it does, in fact, occur.

In short, we seek a UMPU-$w$ test of (3.13) against (3.14). There is nothing objectionable in the restriction to unbiased tests. It just eliminates from consideration tests which allow the chance of rejecting $H_0$ when $H_0$ is true to be larger than the chance of rejecting $H_0$ when some alternatives belonging to $H_1$ are true.

In our problem, for each fixed $x$, $P_a^X(x)$ is a continuous function of $a$, and therefore so is $B_f(a) = E_a f(X)$ for any critical function $f$. 
Also, the boundary between \( A_0 \) and \( A_1 \) is \( A_0 \) here. This means that any unbiased size \( w \) test is \( w \)-similar on \( A_0 \) (lemma 4.1). Therefore, in this particular problem, every unbiased size \( w \) test has Neyman structure with respect to a statistic that is complete and sufficient for \( a \) in \( A_0 \), provided such a statistic exists (lemma 4.3). If such a statistic exists, the problem will be significantly simplified by finding that statistic. For then we will be able to solve the problem separately for each feasible value of that sufficient statistic, as will be seen forthwith.

The first step is to find a sufficient statistic for the subfamily of pdf's specified by (3.13), that is, for \( a \) in \( A_0 \). Substituting \((4a_0a_2)^{1/2}\) for \( a_1 \) in (2.13) yields

\[
P_{X_\alpha} (x, y, z) = 2^y \frac{a_0^{(2x+y)/2} a_2^{(y+2z)/2}}{n!} \frac{n!}{(x! y! z!)},
\]

\[x + y + z = n, \quad a \text{ in } A_0.
\]

Observing that \( n - (2x + y) = (y + 2z) - n = z - x \) reveals that \( 2X_0 + X_1, X_1 + 2X_2, \) and \( X_2 - X_0 \) are equivalent statistics and that \( S = X_2 - X_0 \) is a minimal sufficient and complete statistic for this subfamily. (These assertions will be proven below.)

It is therefore worthwhile to define a new statistic space by the linear transformation

\[
S = X_2 - X_0, \quad T = X_0 + X_2. \tag{4.8}
\]

The statistic \( T = X_0 + X_2 \) is chosen as the second coordinate because lines of constant \( T \) and lines of constant \( S \) are perpendicular in \( X \)-space \( W^X \) (2.12), as illustrated in figure 4.1. As a consequence, the
Figure 4.1: The transformation from \((X_2, X_0)\)-space to \((S, T)\)-space is a rotation and translation.
transformation from \( X \)-space to ST-space is a simple translation and rotation of the plane \( X_0 + X_1 + X_2 = n \). In addition, the line \( S = 0 \) is a line of symmetry for a certain function of \( S \) and \( T \) which we shall want to compute later.

The inverse transformation (back onto \( X \)-space \( W^X \)) is

\[
X_0 = (T - S) / 2, \quad X_1 = n - T, \quad X_2 = (T + S) / 2.
\]

In short, a new statistic space has been induced: ST-space \( W^{ST} \), defined by

\[
W^{ST} = \{(s, t) \in \mathbb{Z}^2 : [(t-s)/2, n-t, (t+s)/2] \in W^X \}. \tag{4.9}
\]

It follows from (2.12) that the feasible values of \( S \) are \( s = -n, -n+1, \ldots, n-1, n \); and, for each fixed value of \( s \), the corresponding feasible values of \( T \) are \( t = |s|, |s|+2, |s|+4, \ldots, |s|+2[(n-|s|)/2]' \).

Under this transformation, there is a one-to-one correspondence between the points in \( W^X \) and \( W^{ST} \). Therefore, since \( X \) is sufficient for \( a \) in \( A \), so is \( (S, T) \). Finally, the two probability measures \( P^X \) and \( P^{ST} \) enjoy the relationship

\[
P^X(x, y, z) = P^{ST}(z - x, x + z), \quad (x, y, z) \in W^X,
\]

\[
P^{ST}(s, t) = P^X[(t - s) / 2, n-t, (t + s) / 2], \quad (s, t) \in W^{ST}.
\]

This means that reporting on the family \( \{P^{ST}_a : a \in A \} \) is equivalent to reporting on the family \( \{P^X_a : a \in A \} \).

Denoting ST-space by \( W^{ST} \), we see that the induced pdf of \( (S, T) \) is
\[ P_{a}^{ST}(s, t) = P_{a}[S = s, T = t] \]
\[ = a_0^{(t-s)/2} a_1^{n-t} a_2^{(t+s)/2} n! / \{[(t-s)/2]! (n-t)! [(t+s)/2]!\}, \]
\[(s, t) \text{ in } W^{ST}. \] (4.10)

Henceforth \( f \) will denote critical functions defined on \( ST \)-space \( W^{ST} \) rather than on \( X \)-space \( W^{X} \). Of course \( f \) is still based on \( X \), because \( f(S, T) = f(X_2 - X_0, X_0 + X_2) \).

We shall now prove the assertions made above that \( S = X_2 - X_0 \) is sufficient and complete for \( a \) in \( A_0 \).

**Fact 4.1:** \( S \) is sufficient for the family of distributions \( \{P_{a}^{ST} : a \text{ in } A_0\} \) specified by (3.13).

To prove fact 4.1 it will suffice to show that for every fixed \( (s, t) \) in \( W^{ST} \), \( P_{a}[T = t \mid S = s] \), the conditional probability of \( ((S, T) = (s, t)) \) given \( S = s \), is a constant function of \( a \) for all \( a \) in \( A_0 \), or, equivalently, for all \( a \) in \( A \) such that \( a_0 a_2 / a_1^2 = 1/4 \) (i.e., under the null hypothesis).

Let

\[ P_{T \mid S}(t) = P[T = t \mid S = s], \quad (s, t) \text{ in } W^{ST}. \]

By the definition of conditional probability on a discrete sample space (Feller, 1968, p. 115),

\[ P[T = t \mid S = s] = P[S = s, T = t] / P[S = s]. \] (4.11)
We compute $p_{T|S}$ by combining (4.10) and (4.11) in the following way. The numerator (4.10) of (4.11) may be written as

$$p_{ST}(s, t)$$

$$= n! a^n (a_2/a_0)^s (a_0 a_2/a_1^2)^{t/2} / \{ [(t-s)/2]! (n-t)! [(t+s)/2]! \}$$

(4.12)

The denominator of (4.11) is

$$p^S(s) = P[S = s] = \sum_t P[S = s, T = t] = \sum_t p_{ST}(s, t)$$

$$= n! a^n a^{s/2} \sum_t b^{t/2} / \{ [(t-s)/2]! (n-t)! [(t+s)/2]! \}$$

(4.13)

where $a = a_2/a_0$ and $b = (a_0 a_2/a_1^2)$, and where for each fixed $s = -n, \ldots, n$, the sum is over all $t$ in the section of $W_{ST}$ (4.9) at $s$,

$$W_s = \{ t \in Z : (s, t) \in W_{ST} \},$$

$$= |s|, |s| + 2, |s| + 4, \ldots, |s| + 2 \lfloor (n-|s|)/2 \rfloor,'$$

(4.14)

where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to $x$. The quotient, therefore, is

$$p_{T|S}(t) = \frac{(a_0 a_2/a_1^2)^{t/2} / \{ [(t-s)/2]! (n-t)! [(t+s)/2]! \}}{\sum_t (a_0 a_2/a_1^2)^{t/2} / \{ [(t-s)/2]! (n-t)! [(t+s)/2]! \}}$$

$$a_0, a_1, a_2 > 0,$$

(4.15)
\[ p_{b}^{T|s}(t) = \frac{b^{t/2}}{\text{SUM} b^{t/2}} \frac{[(t-s)/2]! (n-t)! [(t+s)/2]!}{[(t-s)/2]! (n-t)! [(t+s)/2]!} \]

\( b > 0. \) \hfill (4.16)

Under \( H_0, b = \frac{a_0 a_2}{a_1^2} = 1/4, \) and clearly, under \( H_0, p_{T|s} \) does not depend on any unknown parameter values.

\textbf{Fact 4.2:} The statistic \( S = X_2 - X_0 \) is complete on \( A_0 \) (i.e., for the subfamily of trinomial distributions \( \{p_{ST}^{a} : a \text{ in } A_0 \} \) specified by (3.13)).

According to the definition of completeness, it must be shown that for any real valued function \( h, \)

\[ \text{SUM} h(s) P(S = s) = 0 \] \hfill (4.17)

implies \( h(s) = 0 \) for \( s = -n, \ldots, n. \) Assuming hypothesis (4.17), it follows from (4.13) that for every \( a \text{ in } A_0, \)

\[ \text{SUM} h(s) a^{s/2} \text{SUM} b^{t/2} / [[(t-s)/2]! (n-t)! [(t+s)/2]!] = 0, \]

because \( a_1 \) is positive for every \( a \text{ in } A_0. \) Since \( b \) is constant in \( A_0 \) we denote

\[ f(s) = \text{SUM} b^{t/2} / [[(t-s)/2]! (n-t)! [(t+s)/2]!] \]

to obtain

\[ a^{-n/2} \text{SUM} h(s) f(s) a^{(s+n)/2} = 0, \quad \text{for all } a \text{ in } A_0. \]
Since 'a' is always positive in $A_0$, it follows that

$$\sum_{s} h(s) f(s) a^{(s+n)/2} = 0, \quad \text{for all } a \text{ in } A_0.$$  

This can be reparametrized as

$$2n \sum_{s=0}^{2n} h(s-n) f(s-n) a^{s/2} = 0, \quad \text{for all } a \text{ in } A_0,$$

to reveal that we are dealing with a $2n$-degree polynomial in $a^{1/2}$.

Since an $m$-degree polynomial has at most $m$ distinct zeros (Henrici, 1967, p. 35), and since 'a' takes on infinitely many values for $a$ in $A_0$, and since $f(s-n) > 0$ for $s = 0, \ldots, 2n$, $h(s)$ must equal zero for $s = -n, \ldots, n$.

In addition to being thought of as a conditional pdf on $W^{ST}$ (4.9), $p_{Ti} s$ may be thought of as an unconditional pdf on $W_s$ (4.14), the normalized contraction of $P^{ST}$ by the event $\{S = s\}$. The fact that $p_{Ti} s$ is at once a conditional pdf on $W^{ST}$ and an unconditional pdf on $W_s$ is fundamental to the Neyman-Pearson theory of statistical inference.

The proof of sufficiency also suggests a reparametrization for the family of $P^{ST}$. Defining a new parameter $b$ by $b = a_0 a_2 / a_1^2$, we see that $a$ in $A_0$ is equivalent to $b = 1/4$ (i.e., $\{b = 1/4\}$ connotes the set $\{a in A : b = 1/4\} = A_0$). To fix a particular point in $A_0$ or $A$, we need specify only one more parameter. As will be seen later, it is most convenient to specify the parameter $a_0$. We have thus created a new parameter space $B = \{(a_0, b) \text{ in } R^2 : 0 < a_0 < 1, b > 0\}$. $B$ is the image of $A$ under the one-to-one nonlinear mapping defined by

$$a_0(a) = a_0, \quad b(a) = a_0 a_2 / a_1^2, \quad a \text{ in } A.$$  

(4.18)
Since 'a' is always positive in $A_0$, it follows that

$$\sum_{s} h(s) f(s) a^{(s+n)/2} = 0, \quad \text{for all } a \in A_0.$$  

This can be reparameterized as

$$\sum_{s=0}^{2n} h(s-n) f(s-n) a^{s/2} = 0, \quad \text{for all } a \in A_0,$$

to reveal that we are dealing with a $2n$-degree polynomial in $a^{1/2}$.

Since an $m$-degree polynomial has at most $m$ distinct zeros (Henrici, 1967, p. 35), and since 'a' takes on infinitely many values for $a \in A_0$, and since $f(s-n) > 0$ for $s = 0, \ldots, 2n$, $h(s)$ must equal zero for $s = -n, \ldots, n$.

In addition to being thought of as a conditional pdf on $W^{ST}$ (4.9), $p_{T|S}$ may be thought of as an unconditional pdf on $W_S$ (4.14), the normalized contraction of $P^{ST}$ by the event \{S = s\}. The fact that $p_{T|S}$ is at once a conditional pdf on $W^{ST}$ and an unconditional pdf on $W_S$ is fundamental to the Neyman-Pearson theory of statistical inference.

The proof of sufficiency also suggests a reparameterization for the family of $P^{ST}$. Defining a new parameter $b$ by $b = a_0 a_2 / a_1^2$, we see that $a$ in $A_0$ is equivalent to $b = 1/4$ (i.e., $\{b = 1/4\}$ connotes the set $\{a \in A : b = 1/4\} = A_0$). To fix a particular point in $A_0$ or $A$, we need specify only one more parameter. As will be seen later, it is most convenient to specify the parameter $a_0$. We have thus created a new parameter space $B = \{(a_0, b) \in \mathbb{R}^2 : 0 < a_0 < 1, b > 0\}$. $B$ is the image of $A$ under the one-to-one nonlinear mapping defined by

$$a_0(a) = a_0, \quad b(a) = a_0 a_2 / a_1^2, \quad a \in A.$$  

(4.18)
Under this reparametrization, the hypotheses (3.13) and (3.14) are equivalent to

\[ H_0: \ b = 1/4, \]  \hspace{1cm} (4.19)

\[ H_1: \ b < 1/4, \]  \hspace{1cm} (4.20)

respectively.

Note also that, for any member of the trinomial family, regardless of the hypothesis (precisely, for any \( a \) in \( A \), or for any \( b > 0 \)), \( p_{T_s}^{T} \) does not depend on the parameter \( a_0 \). This exposes the utility of the Neyman structure test. By conditioning on the sufficient statistic \( S \), the problem of testing the composite hypothesis (4.1) on \( W_{ST} \) (4.9) is reduced to that of testing the simple hypothesis (4.19) on \( W_s \) (4.14). In other words, rather than reporting on the two-dimensionally parametrized family \( \{ p_{ST}: a \text{ in } A \} \), we report on the one-dimensionally parametrized family \( \{ p_{T_s}^{T}: b > 0 \} \). This means that in practice, once \( S \) and \( T \) have been observed, we need only conduct a conditional test \( f_s \) on \( W_s \), rather than the unconditional test \( f \) on \( W_{ST} \), where

\[ f_s(\cdot) = f(s, \cdot) \text{ for all } s = -n, \ldots, n. \]  \hspace{1cm} (4.21)

The entire process, however, is equivalent to conducting the unconditional test.

Finally, it can be discerned from (4.16) that \( \{ p_{T_s}^{T}: b > 0 \} \) constitutes an exponential family of pdf's. In other words, (4.16) can be reparametrized such that it will have the form

\[ p_{v}(t) = c_s(v) \exp(tv) h_s(x). \]  \hspace{1cm} (4.22)
Fact 4.3: For each $s = -n, -n + 1, \ldots, n$, the collection of pdf's 
\[
\{ P^s_b : b > 0 \}
\]
is a one-parameter exponential family.

To verify this fact it will suffice to show that (4.16) can be 
expressed in the form (4.22). This is so if we reparameterize as 
follows:

\[
v = \log b^{1/2} = \log (a_0 a_2 / a_1^{1/2}). \tag{4.23}
\]

\[
C_s(v) = 1 / \sum_t b^{t/2} / (((t-s)/2)! (n-t)! [(t+s)/2]!)
\]

\[
= 1 / \sum_t e^{tv} / (((t-s)/2)! (n-t)! [(t+s)/2]!).
\]

Finally, define $h_s$ of (4.22) by

\[
h_s(t) = 1 / (((t-s)/2)! (n-t)! [(t+s)/2]!).
\]

For families of pdf's of the form (4.22), the following lemma 
prescribes a UMP level $\omega$ conditional test $f_s$ (4.21) on $W_s$.

Lemma 4.4: Let $v$ be a real parameter and let $T$ have probability 
density

\[
P_v(t) = C(v) \exp(tv) h(t).
\]

Then there exists a unique UMP-$\omega$ test $g$ for testing $H_0: v = v_0$ against $H_1: v < v_0$ given by
\[ g(t) = 1 \quad \text{if} \quad t < t', \]
\[ g(t) = R \quad \text{if} \quad t = t', \]
\[ g(t) = 0 \quad \text{if} \quad t > t', \]

where \( t' \) and \( R \) are determined by the restriction
\[ \mathbb{E}_{v_0} g(T) = w \quad (0 < R < 1) \quad (4.24) \]

(adapted from Lehman, 1959, p. 70, Corollary 2).

The basis for lemma 4.4 is that any pdf \( P_v \) of the form (4.22) has monotone likelihood ratio in \( t \). That is, for any \( v' < v'' \)

\[ \frac{P_{v'}}{P_{v''}} = \exp[t (v' - v'')] C(v') / C(v'') \]

is a monotone function of \( t \). Here \( \frac{P_{v'}}{P_{v''}} \) is strictly monotone decreasing for every \( v' \) and \( v'' \) such that \( v' < v'' \); in particular, for any \( v_1 < v_0 \), that is for every \( b < 1/4 \). This implies that if \( t_1 \) and \( t_2 \) are any two fixed points in statistic space such that \( t_1 < t_2 \), then

\[ \frac{P_{v'}(t_1)}{P_{v''}(t_1)} > \frac{P_{v'}(t_2)}{P_{v''}(t_2)}, \quad v' < v''; \quad (4.25) \]

in particular, for \( v' = v_1 \) and \( v'' = v_0 \).

Any weight \( g(t) \) assigned by the test \( g \) to the point \( t \) adds \( g(t) P_{v_1} \) to the power \( B_g(v_1) = \text{SUM}_t g(t) P_{v_1} \) of \( g \) against \( v_1 \), and simultaneously adds \( g(t) P_{v_0}(t) \) to the size \( B_g(v_0) = \text{SUM}_t g(t) P_{v_0} \). The most powerful test against any particular alternative \( v_1 \) is that which assigns the greatest allowable weight (i.e., \( g(t) = 1 \)) to the smallest \( t \)'s in statistic space, because, as a consequence of (4.25), if \( t_1 < t_2 \), then...
any weight \( g(t_1) \) assigned to \( t_1 \) adds more to the power, relative to the size, than would the same weight assigned to \( t_2 \). These are the essential arguments of the Neyman-Pearson fundamental lemma (Lehmann, 1959, p. 65). Moreover, since (4.25) is true for \( v'' = v_0 \) and for any \( v' = v_1 \) as long as \( v_1 < v_0 \), \( g \), thusly constructed, is uniformly most powerful.

The uniqueness of the UMP-w test \( g \) follows from the strict monotonicity of \( P_v'/P_{v''} \), \( v' \neq v'' \).

We shall now define a critical function \( f \) for testing (3.13) against (3.14), using the preceding lemmas in such a way that it has Neyman structure with respect to \( S \). We shall take advantage, notationally, of the fact that \( P_{b|S}^T \) automatically defines a probability measure on the events (i.e., the power set) of \( W_s \):

\[
P_b(T = t \mid t \in W_s) = P_b(T = t \mid S = s) = P_{b|S}^T(t) = P_{b|S}^T(T = t),
\]

and

\[
P_b(T < t \mid t \in W_s) = P_{b|S}^T(T < t).
\]

Define a critical function \( f \) mapping \( W^{ST} \) into [0, 1] by

\[
f(s, t) = 1 \quad \text{if } t < t'(s), \quad (4.26a)
\]

\[
f(s, t) = R(s) \quad \text{if } t = t'(s), \quad (4.26b)
\]

\[
f(s, t) = 0 \quad \text{if } t > t'(s), \quad (4.26c)
\]

where, for \( b = 1/4 \),

\[
\]
\[ t'(s) = \max\{t \text{ in } W_s : P^T_{b,S}(T < t) \leq w\}, \quad (4.26d) \]

and

\[ R(s) = (w - P^T_{b,S}(T < t'(s))) / P^T_{b,S}(T = t'(s)). \quad (4.26e) \]

**Fact 4.4:** The test \( f \) defined by (4.26) has Neyman structure with respect to the sufficient statistic \( S \).

To show that \( f \) has Neyman structure with respect to \( S \), it will suffice to show that \( f \) satisfies (4.4) for \( s = -n, \ldots, n \). But under the reparametrization (4.18), (4.4) is equivalent to

\[ E_b[f_s(T)] = w, \quad b = 1/4 \quad (4.27) \]

(where the expectation is with respect to \( P^T_{b,S} \)). Therefore, it will suffice to show that \( f \) satisfies (4.27). Indeed, setting \( b = 1/4 \) and taking expectation with respect to \( P^T_{b,S} \),

\[ E_b[f_s(T)] = \sum_t f(s, t) P^T_{b,S}(T = t) \]

\[ = P^T_{b,S}(T < t'(s)) + R(s) P^T_{b,S}(T = t'(s)) \]

\[ = P^T_{b,S}(T < t'(s)) + w - P^T_{b,S}(T < t'(s)) = w. \]

**Fact 4.5:** For each feasible \( s \), given \( \{S = s\} \), \( f_s \) as defined by (4.21) and (4.26) is UMP among all (conditional) test of size \( w \).

This fact follows immediately from lemma 4.4.
Theorem 4.1: The critical function $f$ defined in (4.26) is the unique UMP unbiased size $w$ test of (3.13) versus (3.14).

It must be shown that $f$ (4.26) is size $w$, UMPU-$w$, and finally, that no other test is UMPU-$w$.

To see that $f$ (4.26) has size $w$, recall that the given family of distributions is such that the power function of any test is continuous; the statistic $S$ is sufficient (fact 4.1) and complete (fact 4.2) for $\alpha$ in $A_0$; and $f$ has Neyman structure with respect to $S$ and $A_0$ (fact 4.4). Therefore, by lemma 4.3, $f$ (4.26) is $w$-similar on the common boundary between $A_0$ and $A_1$. In this problem, however, the common boundary is $A_0$; so $f$ is $w$-similar on $A_0$, the subregion of parameter space specified by the null hypothesis (4.19). Hence, by the definition of similarity, $f$ (4.26) is a size $w$ test.

To see that $f$ (4.26) is UMPU-$w$ recall that the family of pdf's \{\text{pdf} : b > 0\} constitutes an exponential family of pdf's (fact 4.3), and therefore, by lemma 4.4, for each $s = -n, \ldots, n$, $f_s = f(s, \cdot)$ is UMP among all size $w$ tests on $W_s$. This means that

$$E(f(s, T) \mid S = s) \geq E(g(s, T) \mid S = s)$$

for every size $w$ test $g(s, \cdot)$ on $W_s$. This implies that

$$E[E[f(S, T) \mid S = s]] \geq E[E[g(s, T) \mid S = s]]$$

or equivalently,

$$E[f(S, T)] \geq E[g(S, T)]$$

(4.28)

for every size $w$ test $g(s, \cdot)$ on $W_s$. But $g(s, \cdot)$ has size $w$ on $W_s$ means...
In other words, (4.28) holds for all tests $g$ which have Neyman structure, that is $f$ is UMP-w among all Neyman structure tests. Therefore, by lemma 4.3, $f$ is UMP-w among all similar tests. Finally, by lemma 4.2, $f$ is UMPU-w, uniformly most powerful among all unbiased size $w$ tests.

To see that $f$ (4.26) is the unique UMPU-w test, observe that any critical function $g$ that would be defined differently from $f$ (4.26) at even one point, say $(s', t')$, of $W^S$ (4.9), would imply a difference between the two conditional test functions $g(s', t')$ and $f(s', t')$. As asserted in lemma 4.4, however, the UMP tests for maximum likelihood ration families are unique in the sense that alterations on sets of positive probability lower the power, provided the size is held constant. So the power of $g$ would be strictly less than that of $f$ (4.26).

Implementation of the randomized test $f$ (4.26) requires the computation of one number, $t'(s)$ (4.26) for the observed value $s$ of $S$. The test procedure calls for the rejection of $H_0$ if the observed value of $T$ is less than $t'(s)$, and the acceptance of $H_0$ if $t$ is greater than $t'(s)$.

A minor problem arises if the observed value of $T$ is equal to $t'(s)$. In that case, theoretically, the hypothesis is to be rejected with probability $f(s, t'(s))$ and accepted with probability $1 - f(s, t'(s))$. In other words, theoretically, an auxiliary experiment is to be performed, such as flipping a coin or consulting a table of
random numbers, that has two possible outcomes, one that occurs with probability \( f(s, t'(s)) \), and one with probability \( 1 - f(s, t'(s)) \). Of course this second experiment (intentionally and necessarily) would have nothing to do with the phenomena that are the subject of the experiment; so it is difficult to imagine a scientific experiment for which such a procedure would be employed. For an experiment such as that being considered here, the conservative choice is to accept \( H_0 \) if \( t'(s) \) is observed. Such a procedure would amount to a test which is most powerful among all unbiased nonrandomized tests of significance level \( w \) or smaller. (A test is said to have significance level \( w \) for any \( w \) greater than or equal to the size of the test.)

Such an accommodation is implicit whenever a continuous distribution (e.g., the chi-squared) is used to form an approximate test based on a finite random sample. For example, in a chi-squared test, the procedure is to reject the null hypothesis if the observed value of the test statistic exceeds that value \( \chi^2_{1-w} \) of a chi-squared random variable at which the chi-squared cumulative distribution function (cdf) has value \( 1 - w \). In general, no outcome can have exactly that value. Yet no one ever randomizes such a test by conditionally performing an auxiliary experiment if the least feasible value of the test statistic greater than \( \chi^2_{1-w} \) is realized.

A number comparable to the value \( \chi^2_{1-w} \) of a chi-squared statistic does exist for the Neyman structure test (4.26).

Consider the number \( t''(s) = t'(s) - 2 + 2R(s) \). \( R(s) \) (4.26) is such that \( t'(s) - 2 \leq t''(s) \leq t'(s) \), and \( t''(s) \) is not a feasible value of \( T \) given \( S = s \) (i.e., not a point in \( W_s \)) unless \( R(s) = 0 \).
The number $t''(s)$ connotes the expected value of the greatest $t$ in $W_s$ for which the null hypothesis would be rejected. Indeed, that point is $t'(s) - 2$ with probability $1 - R(s)$, and $t'(s)$ with probability $R(s)$; so its expected value is $(t'(s) - 2)[1 - R(s)] + t'(s)R(s)$. Thus,

$$\{(s, t''(s)) \in \mathbb{R}^2 : s = -n, \ldots, n\} \quad (4.29)$$

is the "expected boundary" of the critical region of the test (4.26).

Likewise, the union of the set (4.29) and the set

$$\{(s, t) \in W_{ST} : t \leq t''(s), s = -n, \ldots, n\}$$

is the "expected critical region" of the test (4.26). Since the expected boundary (4.29) is unique, so is the expected critical region. Hence, just as a nonrandomized test based on continuous statistics is identified with a unique critical region (rather than a critical function) in introductory statistics texts (e.g., Hogg and Craig, 1970, p. 272), a randomized test based on discrete statistics can be identified with a unique expected critical region.

In scientific applications, the experimentalist would reject if the observed value of $t$ were $\leq t''(s)$, and not reject, otherwise.

The set (4.29) is a concise representation of the critical function $f$, in that both $t'(s)$ and $R(s)$ can be derived from $t''(s)$ by the relations

$$t'(s) = \min\{t \in W_s : t > t''(s)\},$$

$$R(s) = \frac{t''(s) - [t'(s) - 2]}{t'(s) - [t'(s) - 2]}$$

$$= \frac{t''(s) - [t'(s) - 2]}{2}. $$
As an example, figure 4.2 is a graph of all the points in ST-space \( W_{ST} \) for \( n = 16 \) such that \( s \geq 0 \). The solid line is formed by joining the points (4.29) with straight line segments. The points in \( W_{ST} \) below this boundary are those for which \( f(s, t) = 1 \) (i.e., those which, if realized, result in rejection with probability 1). The points immediately above are those for which \( f(s, t) = R(s) \) (i.e., those which, if realized, result in rejection with probability \( R(s) < 1 \), or acceptance of \( H_0 \) in a nonrandomized test). The remaining points are those for which \( f(s, t) = 0 \) (i.e., those which, if realized, result in acceptance with probability 1).

We can completely close the gap between elementary treatments of hypothesis testing (restricted to nonrandomized tests and continuous statistics) and the advanced treatment of the general case by the following device.

For each \( s = -n, ..., n \), define a probability measure \( P_b^*|s \) on the interval \([|s|, |s|+2[(n-|s|)/2]]\) by

\[
P_b^*|s[T < t] = P_b^T|s[T < t^*] + P_b^T|s[T = t^*] [t-(t^*-2)]/2, \tag{4.30a}
\]

where

\[
t^* = t^*(s, t) = \min\{t_o \in W_s : t_o > t}\.
\tag{4.30b}
\]

\( P_b^*|s \), the "polygonalized version" of \( P_b^T|s \), agrees with \( P_b^T|s \) at every point of \( W_s \) and is linear in between, as illustrated in figure 4.3. \( P_b^*|s \) has been defined in such a way that

\[
P_b^*|s[T < t^n(s)] = w, \quad b = 1/4,
\]
Figure (4.29): The expected boundary and expected critical region which characterize the UMPU-.05 Neyman structure test (4.26) for n = 16. Those points below the expected boundary yield rejection.
Figure 4.3: The relationship between $P^T_b$ (step function, solid line) and $P^*_b$ (polygonalized function, broken line; explanation in text).
where $t^*(s)$ is the expected boundary of the size $w$ test (4.26). Thus, the size of the test $f_s$ is the $p_b^*$-probability that a sample point falls into the expected critical region, under the null hypothesis that $b = 1/4$. In fact, for each $s = -n, \ldots, n$, $p_b^*\{T < t^*(s)\}$ is, as a function of $b$, the power of the conditional test $f_s$.

The results of the UMPU-.05 Neyman structure test (4.26) applied to the data of Gould et al. (1980) are listed in table 4.1, where the critical level is calculated for each strain according to the exact distribution. The critical level of a realization is the size of the smallest sized test which would result in rejection for that realization.

<table>
<thead>
<tr>
<th>Strain</th>
<th>n</th>
<th>s</th>
<th>t</th>
<th>$t^*(s)$</th>
<th>Critical level</th>
</tr>
</thead>
<tbody>
<tr>
<td>COL</td>
<td>166</td>
<td>18</td>
<td>36</td>
<td>72.232</td>
<td>0.000000</td>
</tr>
<tr>
<td>LDB</td>
<td>337</td>
<td>105</td>
<td>111</td>
<td>169.981</td>
<td>0.000000</td>
</tr>
<tr>
<td>WES</td>
<td>128</td>
<td>14</td>
<td>44</td>
<td>54.283</td>
<td>0.000202</td>
</tr>
<tr>
<td>BRO</td>
<td>512</td>
<td>-3</td>
<td>31</td>
<td>236.092</td>
<td>0.000000</td>
</tr>
<tr>
<td>FAR</td>
<td>138</td>
<td>23</td>
<td>25</td>
<td>60.191</td>
<td>0.000000</td>
</tr>
<tr>
<td>RIV</td>
<td>55</td>
<td>6</td>
<td>6</td>
<td>20.435</td>
<td>0.000000</td>
</tr>
<tr>
<td>TAL</td>
<td>74</td>
<td>20</td>
<td>22</td>
<td>32.003</td>
<td>0.000004</td>
</tr>
<tr>
<td>ART</td>
<td>225</td>
<td>5</td>
<td>53</td>
<td>98.996</td>
<td>0.000000</td>
</tr>
<tr>
<td>ARC</td>
<td>228</td>
<td>49</td>
<td>79</td>
<td>106.123</td>
<td>0.000000</td>
</tr>
<tr>
<td>CLA</td>
<td>281</td>
<td>187</td>
<td>199</td>
<td>194.048</td>
<td>0.297602</td>
</tr>
<tr>
<td>WHI</td>
<td>1073</td>
<td>744</td>
<td>756</td>
<td>779.574</td>
<td>0.000000</td>
</tr>
</tbody>
</table>
Now we shall turn our attention to the power of this test, \( B_f(a) \).

It follows from the definition (4.3) that

\[
B_f(a) = \sum_{s,t} f(s,t) p_{ST}^{a}, \quad a \text{ in } A.
\]

Note that this function depends not only on \( b = a_0 a_2 / a_1 \), but also on an auxiliary parameter, for which \( a_0 \) will be used here. Thus, the power of the test (4.26) can be computed as a function \( B'(a_0, b) = B[a(a_0, b)] \) of \( a_0 \) and \( b \) where \( a(a_0, b) = [a_0(a_0, b), a_1(a_0, b), a_2(a_0, b)] \) is defined by

\[
a_0(a_0, b) = a_0, \quad (4.31a)
\]

\[
a_1(a_0, b) = \{-a_0 + [a_0^2 + 4a_0b(1 - a_0)]^{1/2}\} / (2b), \quad (4.31b)
\]

\[
a_2(a_0, b) = \{2b(1 - a_0) + a_0 - [a_0^2 + 4a_0b(1 - a_0)]^{1/2}\} / (2b). \quad (4.31c)
\]

This is the inverse of the relationship

\[
a_0(a) = a_0, \quad b(a) = a_0 a_2 / a_1^2.
\]

Thus, \( B'(a_0, b) = B_f(a(a_0, b)) \) is the power function of the Neyman structure test \( f \) as a function of \( b \) and the auxiliary parameter \( a_0 \).

We also could consider the power of the conditional test \( f_s \) given \( S = s \) (the conditional power)

\[
B(b|s) = E_b f_s(T), \quad b > 0,
\]

where expectation is with respect to \( p_{b|s}^T \). In other words,

\[
B(b|s) = \sum_t f_s(t) p_{b|s}^T(t). \quad (4.32)
\]
The conditional power has the apparent advantage over the unconditional power of being independent of any auxiliary parameter. Moreover, in the sense that

\[ E[B(b|s)] = E_a[E_b[f_s(T)]] = E_a[E[f(S, T) ; S = s]] \]

\[ = E_a f(S, T) = B_f(a), \]

the conditional power \( B(b|s) \) is an unbiased estimator of the unconditional power \( B_f(a) \). Finally, the unconditional power \( B'(a_0, b) \) is related to the conditional power by

\[ B'(a_0, b) = E_a [B(b|s)] = \sum_{s} B(b|s) P^S(s). \quad (4.33) \]

(This is the reason for expressing the unconditional power as a function of \( b \) and an auxiliary parameter rather than as a function of \( a \).)

Since the hypotheses of the test involve only the parameter \( b \), it appears that the conditional power is the more useful measure of the test's efficiency.

But these are exclusively statistical considerations.

If the conditional power of the test against some specific biological alternative is sought, say \( q = q' < 1 \), then the relationship between \( b \) and \( q \) must be considered.

In fact, it follows from \( b = a_0 a_2 / a_1^2 \) and (3.8) that

\[ b(p, q) = (1 - p)^2 q p^2 / [2p(1 - p) + p^2(1 - q)]^2. \]

Thus, the biological conditional power function is

\[ B'(p, q|s) = B(b(p, q)|s), \]
which may be thought of as a function of q with p as an auxiliary parameter.

Likewise, the unconditional biological power is

\[ B^u(p, q) = B'(a_0(p, q), b(p, q)) \]

where \( a_0(p, q) = (1 - p)^2 \). The choice of \( a_0 \) as the auxiliary parameter is arbitrary theoretically, but convenient computationally because \( a_0 \) and \( p \) are in one-to-one correspondence.

The shortcoming of the conditional power function, even though it is an unbiased estimator of the unconditional power function, is that it cannot be computed until after the experiment has been performed, and therefore cannot be used to design the experiment; specifically, to determine the size of the sample (number of cups \( n \)) to achieve a given power against any specified alternative.

The unconditional power can be used to choose the sample size \( n \). Since \( p \) as well as \( q \) must be used to make this determination, a preliminary experiment to estimate \( p \) raising one larva per cup would be useful in this regard.

Figure 4.4 shows the power function of the Neyman structure size \( w = 0.05 \) test \( f \) (4.26) as a function of \( b \) for three different values of the auxiliary parameter \( a_0 \) and \( n = 74 \) (TAL strain). The trinomial parameter \( a_0 \) was chosen as the auxiliary parameter because of the relationship (3.9) between \( p \) and \( a \). That is, since biologically we are interested in the power as a function of \( q \) for different values of \( p \), we compute \( B_f(b, a_0) \), and then reparametrize to get \( B_f(p, q) \) (figure 4.5).
Figure 4.4: The power of the UMPU-w test (4.26) as a function of b for \( a_0 = 0.013514 \) (broader dashed line), \( a_0 = 0.04 \), and \( a_0 = 0.16 \), for \( n = 74 \) (as in TAL strain; 0.013514 = \( \hat{a} \) for TAL, table 2.1).
Figure 4.5: The power function of the UMPU-.05 Neyman structure test (4.26) as a function of the biological parameter $q$ for two values of the biological parameter $p$, $p = .6$ and $p = .883750$, for $n = 74$ (TAL strain; $.883750 = \hat{p}$).
As explained in the previous chapter, a test of the validity of the model (3.8) is feasible with the present data (table 1.1). The null and alternative hypotheses for this test are, respectively,

\[ H_0: \quad b \leq 1/4 \quad (a_0 a_2 / a_1^2 \leq 1/4 \quad \text{or} \quad a \in \text{A''}), \]

\[ H_1: \quad b > 1/4 \quad (a_0 a_2 / a_1^2 > 1/4 \quad \text{or} \quad a \in \text{A'' complement}). \]

Thus, the choice is between two subfamilies of the trinomial pdf (2.13).

It is meaningless to state these hypotheses in terms of the biological parameters \( p \) and \( q \) (i.e., \( H_0: \quad q \leq 1 \) vs. \( q > 1 \)) because \( q > 1 \) makes no sense mathematically (because \( q \) represents a probability) and has no biological interpretation.

By the same reasoning used to construct the UMPU-w test (4.26) of \( H_0 \) (3.13) against \( H_1 \) (3.14), we arrive at the following Neyman structure test \( g \) of \( H_0 \) against \( H_1 \).

Define a critical function \( g \) mapping \( W^ST \) into \([0, 1]\) by

\[
g(s, t) = 1 \quad \text{if} \quad t > t^*(s),
\]

\[
g(s, t) = R^*(s) \quad \text{if} \quad t = t^*(s),
\]

\[
g(s, t) = 0 \quad \text{if} \quad t < t^*(s),
\]

where

\[
t^*(s) = \min \{t \in W_s : P_{b}^{T} S \{T > t \} \leq w \}
\]

and where

\[
R^*(s) = \frac{(w - P_{b}^{T} S \{T > t^*(s)\})}{P_{b}^{T} S \{T = t^*(s)\}},
\]

\[ b = 1/4. \]
The test so constructed is the unique UMPU-w test of $H_0$ against $H_1$ (Lehmann, 1959, p. 136, theorem 3).
5. EXACT DISTRIBUTION

In chapter 7, algorithms and a complex computer program for the numerical evaluation of $P_b^{T|s}$ for any $b > 0$ (i.e., for any $s$ in $A$) are presented. These are used to compute the conditional and unconditional power functions. In order to construct the Neyman structure test $f$ (4.26) it is necessary and sufficient to compute $P_b^{T|s}$ (4.16) only for $b = 1/4$ (i.e., only under the null hypothesis (4.19)). The results of this chapter point the way to a less complex computation for the special case $b = 1/4$, and also provide convenient formulas for the moments of $P_b^{T|s}$.

For each $s = -n, ..., n$, we shall compute $P_b^{T|s}(s + 2x)$

$$= P_b\{X_0 = x \mid S = s\} \text{ for each integer } x \text{ such that}$$

$$|s| - s / 2 = \max(0, -s) \leq x \leq [(n - s) / 2], \quad (5.1)$$

which is equivalent to computing $P_b^{T|s}(t)$ for each $t$ in the domain $W_s$ (4.14).

By changing variables from $t = s + 2x$ to $x = (t - s)/2$, we conclude from (4.12) that

$$P_{a}^{\tilde{a}}\{X_0 = x, X_1 = n - s - 2x, X_2 = s + x\} = P_{a}^{ST}(s, s + 2x)$$

$$= n! a_1^n (a_2/a_1)^s (a_0 a_2/a_1^2)^x / [x! (n-s-2x)! (s+x)!],$$

$$s = -n, ..., n, \quad \max(0, -s) \leq x \leq [(n-s)/2];$$

by the same change of variables, it follows from (4.13) that
\[ P_a^S(s) = n! a_1^n (a_2/a_1)^S \sum_x b^x / \{x! (n-s-2x)! (s+x)!\} \]  
\[ s = -n, \ldots, n, \]

where summation is over all integers \( x \) in the range

\[ \max(0, -s) \leq x \leq [(n-s)/2]; \] and by the same change of variables, it follows from (4.16) that

\[ P_b \{ X_0 = x \mid S = s \} = P_b \{ T_1^S(s+2x) = \frac{b^x / \{x! (n-s-2x)! (s+x)!\}}{\sum b^x / \{x! (n-s-2x)! (s+x)!\}} \] 
\[ s = -n, \ldots, n, \quad \max(0, -s) \leq x \leq [(n-s)/2], \]

where summation is over all integers \( x \) in the range

\[ \max(0, -s) \leq x \leq [(n-s)/2]. \]

**Fact 5.1:** Define

\[ Z(s, n, b) = \sum_x b^x / \{x! (n-s-2x)! (s+x)!\}, \quad (5.4) \]

\[ b > 0, \quad s = -n, \ldots, n. \]

Then

\[ Z(s, n, 1/4) = 2^S \prod_{j=1}^n (2j - 1) / [(n-s)! (n+s)!], \quad (5.5) \]
\[ s = -n, \ldots, n. \]

The first step is to find a generating function (Chiang, 1968, p. 24 ff.) for the sequence \( Z(s, n, b) \), \( s = -n, \ldots, n \). Since

\[ \sum_{s} P_a^S(s) = 1, \]

it follows from (5.2) that
\[ a_1^n n! \sum_s (a_2/a_1)^s Z(s, n, b) = 1, \]

where summation is over all integers \( s = -n, \ldots, n \). Hence

\[ a_1^n n! \sum_s (a_2/a_1)^s Z(s, n, b) = (a_0 + a_1 + a_2)^n \]

and, dividing by \( a_1^n \), which is always positive,

\[ \sum_s (a_2/a_1)^s Z(s, n, b) = (a_0/a_1 + 1 + a_2/a_1)^n / n!. \]

Letting \( k = a_2/a_1 \) and using \( b = a_0 a_2/a_1^2 \) yields \( a_0/a_1 = b/k \) and

\[ \sum_s k^s Z(s, n, b) = (b/k + 1 + k)^n / n!. \]

or, multiplying both sides by \( k^n \),

\[ \sum_s k^{n+s} Z(s, n, b) = (b + k + k^2)^n / n!. \]

Taking the \((n + s)\)-th derivative and evaluating at \( k = 0 \) yields

\[ Z(s, n, b) = D_k^{n+s} (b + k + k^2)^n / [n! (n + s)!]|_{k=0}. \]

For \( b = 1/4 \) this is

\[ Z(s, n, 1/4) = D_k^{n+s} (1/4 + k + k^2)^n / [n! (n + s)!]|_{k=0}, \]

or

\[ Z(s, n, 1/4) = D_k^{n+s} (1/2 + k)^{2n} / [n! (n + s)!]|_{k=0}, \]

or

\[ Z(s, n, 1/4) = (1/2 + k)^{n-s} (2n)! / [(n - s)! n! (n + s)!]|_{k=0}, \]

\[ = 2^s (2n)! / [2^n n! (n - s)! (n + s)!]. \]
However,

\[
\frac{(2n)!}{n!} = 2^n \prod_{j=1}^{n} (2j - 1).
\]

Therefore

\[
Z(s, n, 1/4) = 2^s \prod_{j=1}^{n} (2j - 1) / [(n - s)!(n + s)!],
\]

\[s = -n, \ldots, n,\]

the desired result.

Let the \(k\)th factorial moment of \(X_0\) given \(S = s\) be denoted and defined \(m_k(s, n, b) = E(X^{(k)}) = E[X(X - 1)(X - 2)\ldots(X - k + 1)].\)

Fact 5.2: For \(s = -n, \ldots, n\), the \(k\)th factorial moment of

\[
P_b\{X_0 = x \mid S = s\}
\]

is

\[
m_k(s, n, b) = b^k \frac{Z(s + k, n - k, b)}{Z(s, n, b)},
\]

\[k = 0, 1, 2, \ldots, \lfloor(n - s)/2\rfloor'.\]

For \(k > \lfloor(n - s)/2\rfloor', m_k(s, n, b) = 0.

By definition

\[
m_k(s, n, b) = \sum_x x^{(k)} P\{X_0 = x \mid S = s\}
\]

where summation is over all \(x\) such that \(\max(0, -s) \leq x \leq \lfloor(n-s)/2\rfloor'.\)

However, \(x^{(k)} = x(x - 1)(x - 2)\ldots(x - k + 1) = 0\) for all \(x\) such that \(x\)
\[ = 0, 1, 2, \ldots, (k - 1). \] Therefore \( m_k(s, n, b) = 0 \) for 
\[ k > [(n-s)/2]' \], and

\[ m_k(s, n, b) = \sum_x x^k P[X_0 = x \mid S = s], \quad k \leq [(n-s)/2]', \]

where summation is over all \( x \) such that \( \max(k, -s) \leq x \leq [(n-s)/2]' \). By (5.3),

\[ m_k(s, n, b) \]

\[ = \sum_x (s) b^x / [(x-k)(n-s-2x)! (s+x)! Z(s, n, b)] \]

where summation is over all \( x \) such that \( \max(k, -s) \leq x \leq [(n-s)/2]' \). Therefore

\[ m_k(s, n, b) \]

\[ = \sum_x b^x / [(x-k)(n-s-2x)! (s+x)! Z(s, n, b)] \]

\[ = b^k \sum_x b^{x-k} / Z(s, n, b) \]

\[ = b^k \sum_x \frac{b^{x-k}}{Z(s, n, b)} \]

where summation is over all integers \( x \) such that

\[ \max(k, -s) \leq x \leq [(n-s)/2]' \]. Therefore,

\[ m_k(s, n, b) = b^k \sum_x \frac{b^x}{Z(s, n, b)} \]

where summation is over all integers \( x \) such that

\[ \max(k, -s) - k \leq x \leq [(n-s)/2]' - k \], that is, where summation is over all \( x \) such that \( \max[0, -(s+k)] \leq x \leq [(n-k)-(s+k)]/2]' \). This is the desired conclusion.
By combining the results of the two previous facts; specifically, by substituting the formula for \( Z(s, n, 1/4) \) (5.5) into (5.6) the following theorem is obtained.

**Theorem 5.1:**

\[
m_k(s, n, 1/4) = \frac{(n-s)!}{[2^k \prod_{j=1}^{k} (n-s-2j)!]} \prod_{j=1}^{k} (2n-2j+1).
\]  

Equation (5.7) is equivalent to

\[
m_k(s, n, 1/4) = \prod_{j=1}^{k} (n-s-2j+2) (n-s-2j+1) / [2 (2n-2j+1)].
\]

From this the following recurrence relationship for \( m_k(s, n, 1/4) \) with respect to \( k \) is apparent:

\[
m_k(s, n, 1/4) = m_{k-1}(s, n, 1/4) (n-s-2k+2) (n-s-2k+1) / [2 (2n-2k+1)],
\]

\[k = 2, 3, \ldots, \lfloor (n - s) / 2 \rfloor.\]

Adopting the conventional definition \( x^{(0)} = 1 \), we have

\[E(X^{(0)} | S = s) = E(1 | S = s) = 1,\]

that is

\[
m_0(s, n, 1/4) = 1,
\]

and the recurrence relation (5.8) is true for \( k = 1 \) as well. Thus we have the ingredients of an algorithm for computation of the first four factorial moments of \( X_0 \) given \( S = s \). This algorithm is formulated in chapter 7.
Given the first four factorial moments $m_{(k)}(s, n, 1/4)$ of the conditional distribution of $X_0$ given $S = s$ (5.3), the mean, variance, and third and fourth central moments can be computed using well known and easily derived formulas (Johnson and Kotz, 1969, pp. 18-19). From these, the mean, variance, skewness, and kurtosis of $p_T^b | S$ (4.16) ($b = 1/4$) are computed as follows.

Denote $m_T(s, n, 1/4) = E_{b=1/4}(T | S = s)$. Since $T = S + 2X_0$, $E(T | S = s) = s + 2E(X_0 | S = s)$. From this and theorem 5.1 it follows that

$$m_T(s, n, 1/4) = \frac{[n(n - 1) + s^2]}{(2n - 1)}$$

$$= \frac{(n^2 + s^2 - n)}{(2n - 1)}.$$  \hspace{1cm} (5.10)

Likewise, letting $s_T^2(s, n, 1/4)$ denote the variance of $p_T^b | S$ ($b = 1/4$), we have $s_T^2(s, n, 1/4) = \text{Var}_{b=1/4}(X_0 | S = s)$. Applying theorem 5.1 we obtain

$$s_T^2(s, n, 1/4) = \frac{2(n + s)(n + s - 1)(n - s)(n - s - 1)}{(2n - 1)^2 (2n - 3)}$$

$$= \frac{2(n^2 - s^2)^2 - (2n - 1)(n^2 - s^2)}{(2n - 1)^2 (2n - 3)}.$$  \hspace{1cm} (5.11)

In a like fashion, the skewness and kurtosis of $p_T^b | S$ ($b = 1/4$) can be computed. Let $m_k(s, n, b)$ denote the $k$th central moment of the conditional distribution of $X_0$ given $S = s$. The index of skewness (skewness) is defined by $m_3(s, n, b) / [m_2(s, n, b)]^{3/2}$. The index of kurtosis (kurtosis) is defined by $m_4(s, n, b) / [m_2(s, n, b)]^2$. These
are also the skewness and kurtosis of \( P_{b}^{T|s} \), however, because given 
\( S = s, T = s + 2X_0 \) is a linear function of \( X_0 \) (Johnson and Kotz, 1969, p. 18).

Figures 5.1 and 5.2 are graphs of the mean, variance, skewness, and kurtosis of \( P_{b}^{T|s} \) as a function of \( s \) for \( b = 1/4 \) and \( n = 74 \) (TAL strain), computed as described above.

Formula (5.5) of fact 5.1 is the basis of recurrence relations for 
\( P_{b}^{T|s} \) (\( b = 1/4 \)) and, thereby, of algorithms for the computation of the Neyman structure test \( f(4.26) \) (i.e., its critical level given an observation and/or its expected critical region boundaries given \( n \)), which are formulated in chapter 7.

Indeed, substituting (5.4) into (5.3) yields

\[
P_{b}\{X_0 = x \mid S = s\} = b^x / [x! (n-s-2x)! (s+x)! Z(s, n, b)].
\]

For \( b = 1/4 \) this is equal to

\[
1 / [2^{2x} x! (n-s-2x)! (s+x)! Z(s, n, 1/4)].
\]

Substituting from formula (5.5) yields

\[
P_{b=1/4}\{X_0 = x \mid S = s\} = (n-s)! (n+s)! / [2^{3+2x} x! (n-s-2x)! (s+x)! \prod_{j=1}^{n} (2j-1)].
\]

From this we find a recurrence relation for \( P_{b=1/4}\{X_0 = x \mid S = s\} \) with respect to \( x \).
Figure 5.1: The mean and variance of the conditional distribution of $T$ given $S = s$ (4.16) as a function of $s$ for $n = 74$ (TAL strain) and $b = 1/4$ (4.19).
Figure 5.2: The skewness and kurtosis of the conditional distribution of $T$ given $S = s$ (4.16) as a function of $s$ for $n = 74$ (TAL strain) and $b = 1/4$ (4.19).
\[ P_{b=1/4} \{ X_0 = x \mid S = s \} \]
\[ = P_{b=1/4} \{ X_0 = x - 1 \mid S = s \} \frac{(n-s-2x+2)(n-s-2x+1)}{4x(s+x)}, \]
\[ x = 1, 2, \ldots, \left\lfloor \frac{(n-s)/2} \right\rfloor. \]  
(5.12)

It also follows that
\[ P_{b=1/4} \{ X_0 = 0 \mid S = s \} = \frac{(n+s)!}{2^s s! \prod_{j=1}^{n} (2j-1)}. \]

This yields a recurrence relation with respect to s,
\[ P_{b=1/4} \{ X_0 = 0 \mid S = s \} \]
\[ = P_{b=1/4} \{ X_0 = 0 \mid S = s - 1 \} \frac{(n + s)}{(2s)}, \]  
(5.13)
\[ s = 1, 2, \ldots n. \]

Finally, it follows that
\[ P_{b=1/4} \{ X_0 = 0 \mid S = 0 \} = \prod_{j=1}^{n} \frac{j}{(2j-1)}. \]  
(5.14)
6. APPROXIMATE TESTS

Gould, et. al. (1980) used a chi-squared goodness-of-fit procedure to test the null hypothesis that there is no nonrandom mortality (Sokal and Rohlf, 1969, pp. 550 ff.).

The assumption that there is no cannibalism, and that there is no other form of mortality which is affected by interaction (i.e., that all mortality is random), implies that the number of pupae per cup is a binomially distributed random variable with parameters 2 and r, where r is the probability that an individual larva reaches the pupation stage, as defined in (2.18).

Recall that \( \hat{r} \) as defined by (2.17) is an unbiased, consistent, maximum likelihood estimator for the biological parameter r. This still holds under the null hypothesis that all mortality is random. Using \( X_0 + X_1 + X_2 = n \) and \( S = X_2 - X_0 \), (2.17) takes the form

\[
\hat{r} = \frac{n + S}{2n}.
\] (6.1)

The average survivorship r is an auxiliary parameter which is estimated from the data by \( \hat{r} \) in the classical goodness-of-fit procedure (Pearson, 1900; Lancaster, 1969, pp. 142 ff.).

Recalling that \( a_i \) is the probability of i pupae in any one cup, we find that, under the biological assumption that \( U \) is a binomially distributed random variable with parameters 2 and r, the relationship between the biological parameter r and the statistical parameters \( a_0 \), \( a_1 \), and \( a_2 \) is

\[
a_0 = (1 - r)^2, \quad a_1 = 2r(1 - r), \quad a_2 = r^2.
\] (6.2)
Thus, the statistical hypothesis corresponding to the biological hypothesis that pupation is random is (4.19), and the alternative is $H_1: b \neq 1/4$. Since $\bar{a}_0$, $\bar{a}_1$, and $\bar{a}_2$ are continuous functions of $r$, and $\hat{r}$ is a maximum likelihood estimator for $r$ under $H_0$ (4.19), the statistics $\bar{a}_0$, $\bar{a}_1$, and $\bar{a}_2$ defined by

\[
\bar{a}_0 = (1 - \hat{r})^2 = (n - S)^2 / (2n)^2,
\]

\[
\bar{a}_1 = 2\hat{r}(1 - \hat{r}) = (n^2 - S^2) / (2n^2),
\]

\[
\bar{a}_2 = \hat{r}^2 = (n + S)^2 / (2n)^2,
\]

are maximum likelihood estimators for $a_0$, $a_1$, $a_2$ under $H_0$ (4.19). Likewise, under the restriction $b = 1/4$, the maximum likelihood estimators for $n\hat{a}_i = E(X_i)$ are

\[
\bar{X}_i = n \bar{a}_i, \quad i = 0, 1, 2,
\]

or

\[
\bar{X}_0 = n(1 - \hat{r})^2 = (n - S)^2 / (4n),
\]

\[
\bar{X}_1 = 2n\hat{r}(1 - \hat{r}) = (n^2 - S^2) / (2n),
\]

\[
\bar{X}_2 = n\hat{r}^2 = (n + S)^2 / (4n).
\]

Finally, it follows from (6.2) that

\[
\bar{a}_0 + \bar{a}_1 + \bar{a}_2 = 1.
\]

and from (6.4) that

\[
\bar{X}_0 + \bar{X}_1 + \bar{X}_2 = n.
\]
Now Pearson's (1900) famous chi-squared statistic

\[ K^2 = \sum_{i=0}^{2} \frac{(\text{observed}_i - \text{expected}_i)^2}{\text{expected}_i} \]

can be computed. Here, observed_i = X_i, expected_i = \bar{x}_i. We get

\[ K^2 = \sum_{i=0}^{2} \frac{(X_i - \bar{x}_i)^2}{\bar{x}_i}. \] (6.7)

We are now in a position to prove

**Lemma 6.1:** Let \( K^2 \) and \( \bar{x}_1 \) be defined as above. Then

\[ K^2 = n \frac{(X_1 - \bar{x}_1)^2}{\bar{x}_1^2}. \] (6.8)

To prove this lemma, first note that

\[ \bar{x}_2 - \bar{x}_0 = S. \] (6.9)

Indeed, from (6.4) we obtain

\[ \bar{x}_2 - \bar{x}_0 = n \frac{\hat{\rho}^2 - (1 - \hat{\rho})^2}{n} \]
\[ = n \left[ \frac{\hat{\rho}^2 - (1 - \hat{\rho})^2}{n} \right] = n (2\hat{\rho} - 1). \]

Substituting from (6.1) yields

\[ \bar{x}_2 - \bar{x}_0 = n \left\{ \frac{2(n + S)}{(2n)} - 1 \right\}, \]

which proves (6.9).

The identity

\[ 4 \bar{x}_0 \bar{x}_2 = \bar{x}_1^2, \] (6.10)

which follows by substitution from (6.4), is also needed.

From (6.9) we have \( \bar{x}_2 - \bar{x}_0 = X_2 - X_0 \), or

\[ X_2 - \bar{x}_2 = X_0 - \bar{x}_0. \] (6.11)

Now it follows that

\[ (X_2 - \bar{x}_2)^2 = (X_0 - \bar{x}_0)^2 = (X_1 - \bar{x}_1)^2 / 4. \] (6.12)
Indeed, by (6.6), \( \bar{x}_1 = n - \bar{x}_2 - \bar{x}_0 \). Since \( n = x_0 + x_1 + x_2 \),

\[ \bar{x}_1 = x_0 + x_1 + x_2 - \bar{x}_2 - \bar{x}_0, \]
or

\[ \bar{x}_1 - x_1 = x_0 - \bar{x}_0 + x_2 - \bar{x}_2. \]

By (6.11)

\[ \bar{x}_1 - x_1 = 2 (x_0 - \bar{x}_0), \]

and squaring both sides yields \((x_1 - \bar{x}_1)^2 = 4 (x_0 - \bar{x}_0)^2\). By applying (6.11) once again, (6.12) obtains.

It follows from (6.7) by applying (6.12) that

\[ K^2 = (x_1 - \bar{x}_1)^2 [1/(4\bar{x}_0) + 1/\bar{x}_1 + 1/(4\bar{x}_2)], \]

or, finding a common denominator,

\[ K^2 = (x_1 - \bar{x}_1)^2 (\bar{x}_1 \bar{x}_2 + 4\bar{x}_0 \bar{x}_2 + \bar{x}_0 \bar{x}_1) / (4\bar{x}_0 \bar{x}_1 \bar{x}_2). \]

Substituting from (6.12) yields

\[ K^2 = (x_1 - \bar{x}_1)^2 (\bar{x}_1 \bar{x}_2 + \bar{x}_1^2 + \bar{x}_0 \bar{x}_1) / (\bar{x}_1^3). \]

or

\[ K^2 = (x_1 - \bar{x}_1)^2 (\bar{x}_2 + \bar{x}_1 + \bar{x}_0) / (\bar{x}_1^2). \]

The lemma follows from (6.6).

Using \( T = x_0 + x_2 \) and the definition of \( \bar{x}_1 \) (6.4), it follows that in terms of \( T \) and \( S \),

\[ K^2 = \frac{[T - (n^2 + S^2) / (2n)]^2}{[(n^2 - S^2) / (2n^{3/2})]^2}. \]

(6.13)

For a size \( w \) test, the procedure is to reject the null hypothesis \((b = 1/4)\) that all mortality is random (binomial) if and only if

\[ K^2 > k^2_{1-w}, \]

where \( k^2_{1-w} \) is such that if \( K^2 \) is distributed as a chi-squared random variable with one degree of freedom, then \( P(K^2 \leq k^2_{1-w}) = 1 - w. \)
Consequently, the null hypothesis is rejected if and only if $K < -k_{1-w}$ or if $k_{1-w} < K$, where $k_{1-w}$ denotes the positive square root of $k_{1-w}^2$, and $K$ is defined by

$$
K = \frac{T - (n^2 + S^2) / (2n)}{(n^2 - S^2) / (2n^{3/2})}.
$$

(6.14)

This means that the null hypothesis is rejected if

$$
T \leq (n^2 + S^2) / (2n) - k_{1-w} (n^2 - S^2) / (2n^{3/2})
$$
or if

$$
T \geq (n^2 + S^2) / (2n) + k_{1-w} (n^2 - S^2) / (2n^{3/2}).
$$

However, $k_{1-w} = z_{1-w/2}$ and $-k_{1-w} = z_{w/2}$, where for any $u$ such that $0 < u < 1$, $z_u$ is the number such that if a random variable $N$ is normally distributed with mean 0 and variance 1, then $P(N \leq z_u) = u$.

Thus, the size $w$ one-sided chi-squared goodness-of-fit test based on the assumption that $K^2$ is distributed as a chi-squared random variable with one degree of freedom defines the same critical region as the size $w$ two-sided "normal" test based on the assumption that $K$ is normally distributed with mean 0 and variance 1, provided that in the latter test the null hypothesis is rejected if and only if $K < z_{w/2}$ or $K > z_{1-w/2}$, where $z_{w/2}$ and $z_{1-w/2}$ are defined as above.

However, the assertion that $K^2$ is distributed as a chi-squared random variable with one degree of freedom does not imply that $K$ is normally distributed with mean 0 and variance 1. Indeed, suppose that $N$ is normal with mean 0 and variance 1, and that $K = |N|$. Then $K$ is not
normal, yet $K^2 = N^2$ is chi-squared with one degree of freedom. More generally, assume that $N$ is normal with mean 0 and variance 1; let $p$ be any number between 0 and 1; and let $K = |N|$ with probability $p$, $K = -|N|$ with probability $1 - p$. Then $K^2 = N^2$ and $K^2$ is chi-squared with 1 degree of freedom, yet $K$ is not normal.

In practice the chi-squared test is performed conditionally upon the observed value of $S$. $p_{ST}^*$ is such that all of the maximum likelihood estimators that have been discussed thus far are unique. Under the hypothesis that $b = 1/4$, all unique maximum likelihood estimators are necessarily determined by the value of the sufficient statistic $S$ (Hogg and Craig, 1970, p. 256, theorem 2). Thus, the chi-squared test amounts intuitively to a conditional test given $S = s$ of the simple hypothesis

$$a_0 = \bar{a}_0, \quad a_1 = \bar{a}_1, \quad a_2 = \bar{a}_2,$$

(which implies $b = 1/4$) against the composite alternative

$$b \neq 1/4, \quad a_2 - a_0 = \bar{a}_2 - \bar{a}_0 (= s/n).$$

Let

$$t_w(s) = \frac{(n^2 + s^2)}{(2n)} + z_w \frac{(n^2 - s^2)}{(2n^{3/2})}, \quad (6.15)$$

$$0 < w < 1.$$

Table 6.1 lists the actual size (calculated according to the exact distribution of $T$ given $S = s$)

$$w = p^*_b \{T \leq t_{.025} \} + p^*_b \{T \geq t_{.975} \}, \quad b = 1/4.$$
of the two-sided randomized test based on $T$, conditional on $S = s$, which would generate an expected critical region identical to that generated by the supposedly size .05 chi-squared test for the data of Gould, et al. (1980; table 1.1). The fact that these numbers are different from .05 indicates that $K^2$ actually is not distributed as a chi-squared random variable with one degree of freedom. The difference $(w - .05)$ between the actual size $w$ and .05 is a measure of the inaccuracy of the approximation achieved by the chi-squared test; precisely, $(w - .05)$ is a measure of the difference between the the actual distribution of $K^2$ and the chi-squared distribution at the point $k_{.95}^2$.

Table 6.1: The actual size of the critical region generated by the supposedly size .05 chi-squared test applied to the data of Gould et al. (1980; explanation in text).

| Strain | n | $s$ | $|s|/n$ | Size $w$ |
|--------|---|-----|--------|---------|
| RIV    | 55| 6   | 0.109  | 0.0698428|
| TAL    | 74| 20  | 0.270  | 0.0613851|
| WES    | 128| 14 | 0.109  | 0.0562521|
| FAR    | 138| 23 | 0.167  | 0.0579774|
| COL    | 166| 18 | 0.108  | 0.0556397|
| ART    | 225| 5  | 0.022  | 0.0542764|
| ARC    | 228| 49 | 0.215  | 0.0536008|
| CLA    | 281| 187| 0.665  | 0.0573228|
| LDB    | 337| 105| 0.312  | 0.0525627|
| BRO    | 512| -3 | 0.006  | 0.0520654|
| WHI    | 1073| 744| 0.698  | 0.0514937|
With the exception of the data (i.e., \(n\) and \(s\) values) for the CLA strain (and to a lesser extent for the FAR strain), the inaccuracy \(w - .05\) of the chi-squared test decreases as the sample size (number of cups) \(n\) increases. In this respect, the result of the test applied to the data of the CLA strain reveals that the inaccuracy of the chi-square approximation is not solely a function of the sample size, that is, the accuracy cannot be expected to improve automatically if the sample size is increased. In fact, as will be shown conclusively, the accuracy of the chi-squared and normal approximations depends not only on the sample size \(n\), but also on the relative magnitude of \(s\) with respect to \(n\) (i.e., \(|s|/n\), which also is listed in table 6.1 for each strain).

We shall now consider the normal approximation of the distribution of \(K\) as the basis of a (one-sided) test of the biological hypothesis of no cannibalism (3.11) against the alternative (3.12), and compare it to the UMPU-W test (4.26).

The assumption that \(K\) is normally distributed with mean \(0\) and variance \(1\) is equivalent to the assumption that, given \(S = s\), \(T\) is normally distributed with mean

\[
\bar{m}_T(s, n, 1/4) = \frac{n^2 + s^2}{2n}
\]

and variance

\[
\bar{s}^2_T(s, n, 1/4) = \frac{(n^2 - s^2)^2}{4n^3}.
\]

Assuming that under \(H_0\) (3.11), given \(S = s\), \(T\) is normally distributed with mean \(\bar{m}_T(s, n, 1/4)\) and variance \(\bar{s}^2_T(s, n, 1/4)\) the test procedure is to reject \(H_0\) if and only if \(T \leq t_{.05}\), where \(t_{.05}\) is defined
as in (6.15). The actual size \( w = P_b \left( S / \frac{t}{0.05} \right) (b = 1/4) \) of the supposedly size .05 normal test applied to the data of Gould et al. (1980; table 1.1) is listed in table 6.2. The inaccuracy \( w - .05 \) measures the inaccuracy of the approximation achieved by the normal test; precisely, it is a measure of the actual (exact) distribution of \( K \) and the normal mean 0 variance 1 distribution at the point \( z_{0.05} = 1.645 \).

Table 6.2: The actual size of the critical region generated by the supposedly size .05 normal test applied to the data of Gould et al. (1980; explanation in text).

| Strain | \( n \) | \( s \) | \(| s |/n \) | Size \( w \) | Size \( w_c \) | R. I.* |
|--------|------|-----|-------|-------|-------|-------|
| RIV    | 55   | 6   | 0.109 | 0.0974065 | 0.0878286 | 0.202037 |
| TAL    | 74   | 20  | 0.270 | 0.0955111 | 0.0855980 | 0.217818 |
| WES    | 128  | 14  | 0.109 | 0.0789967 | 0.0730274 | 0.205859 |
| FAR    | 138  | 23  | 0.167 | 0.0784917 | 0.0710976 | 0.259516 |
| COL    | 166  | 18  | 0.108 | 0.0749264 | 0.0697713 | 0.206812 |
| ART    | 225  | 5   | 0.022 | 0.0710142 | 0.0664897 | 0.215307 |
| ARC    | 228  | 49  | 0.215 | 0.0709935 | 0.0657160 | 0.251390 |
| CLA    | 281  | 187 | 0.665 | 0.0743032 | 0.0704934 | 0.156760 |
| LDB    | 337  | 105 | 0.312 | 0.0668229 | 0.0626619 | 0.247344 |
| BRO    | 512  | -3  | -0.006 | 0.0628057 | 0.0597484 | 0.238744 |
| WHI    | 1073 | 744 | 0.693 | 0.0625024 | 0.0604520 | 0.163997 |

* relative improvement, \( (w - w_c)/(w - .05) \)

Here, a type I error amounts to accepting the hypothesis that cannibalism occurs when, in fact, it does not. It has already been decided that avoidance of precisely this error is the highest priority.
in constructing the test; use of the normal test would be "dangerous" in this respect.

If the inaccuracy were in the other direction, that is if the actual size \( w \) of the equivalent randomized test were less than .05, then the inaccurate normal test could be used as a conservative substitute for the UMP-.05 test \( f_s \).

Another way of assessing the difference between the UMP\(-.05\) Neyman structure test (4.26) and the supposedly size .05 normal test, is to compare their respective critical regions. This is done in figures 6.1, 6.2, and 6.3, for \( n = 16, 64, 256 \), respectively. Figure 6.1 (\( n = 16 \)) shows the boundaries of the critical regions superimposed on \( W^\text{ST} \). The points between the boundaries represent those outcomes for which the two tests result in different decisions. Figure 6.2 shows the same thing for \( n = 64 \), except only the points at which the two tests differ are shown. In figure 6.3, for \( n = 256 \), the points are so dense that they are indistinguishable at the level of resolution of the figure; at the same time they are so dense that many lie between the boundaries of the two tests.

Finally, we consider a "corrected normal" approximation of the UMP\( -w \) tests. The correction entails using the exact values \( m_T(s, n, 1/4) \) and \( s_T^2(s, n, 1/4) \) of the mean and variance of \( p_b^{T|s} \), \( b = 1/4 \), rather than the estimators \( \bar{m}_T(s, n, 1/4) \) and \( s_T^2(s, n, 1/4) \) which obtain from the chi-squared test. Thus, it is assumed that the statistic \( K_c = [T - m_T(s, n, 1/4)]/s_T(s, n, 1/4) \) is a standard normal random variable. The actual size

\[
w_c = p_b^{T|s}\{T \leq m_T(s, n, 1/4) + z_{.05} s_T(s, n, 1/4)\}
\]
Figure 6.1: The critical region generated by the supposedly size .05 normal test (lower broken line) and that of the UMPU-.05 test f (4.26) (lower solid line). The points of ST-space (n = 16) are included (symbol +) to illustrate where contrary decisions would occur (squares). The upper two lines are the approximate mean $\bar{m}$ (broken) and exact mean $m$ (solid).
Figure 6.2: The critical region generated by the supposedly size .05 normal test (broken line) and that of the UMPU-.05 test f (4.26) (solid line). The points of ST-space (n = 64) that lie between the boundaries of the two critical regions are included (symbol +) to illustrate where contrary decisions would occur.
Figure 6.3: The critical region generated by the supposedly size .05 normal test (broken line) and that of the UMPU-.05 test f (4.26) (solid lines). The points of ST-space (n = 256) are too dense to be distinguished at this level of resolution.
of this supposedly size .05 test (i.e., the size $w_c$ of the UMP-$w_c$ randomized conditional test $f_s$ which would produce the same expected critical region) is listed in the next to the last column of table 6.2. Also listed there is the relative improvement $(w - w_c)/(w - .05)$ of the corrected normal test over the uncorrected normal test.

The constructions of these approximate tests entail approximating the exact $P_b^{* \mid s}$-distribution of $T$ given $S = s$, $b = 1/4$. As already mentioned, the various measures of inaccuracy for the respective tests originate from the disparity between the exact $P_b^{* \mid s}$-distribution of $T$ given $S = s$ and the normal approximation at $t_{.025}$, $t_{.975}$, and $t_{.05}$; $w_c - .05$ is a measure of the disparity between $P_b^{* \mid s}$ and the corrected normal cdf at the point $t = m_T(s, n, 1/4) - 1.645 s_T(s, n, 1/4)$ ($1.645 = z_{.05}$).

Further insight can be gained by comparing the exact cdf of $K$ with the cdf of a truly normal mean 0 variance 1 random variable. For $s/n$ fixed at 0, 1/4, 1/2, and 3/4, these comparisons are illustrated in figures 6.4 through 6.7, respectively. Each of these figures is a graph of the exact conditional distribution of $K$ given $S/n = s/n$, for $n = 16$, 64, and 256, and the cdf of a truly normal mean 0 variance 1 random variable.

Comparison of the exact cdf of $K_c$ with the cdf of a truly normal mean 0 variance 1 random variable produces very similar results, except the curves are somewhat closer together.

Figures 6.8 through 6.11 show the exact cdf of $T$ given $S = s$ $P_b^{T \mid s}[T \leq t]$ and the normal and corrected normal approximation for $n = 16$, 64, and for $s/n = 0$, 1/2. It is apparent that the corrected
Figure 6.4: The exact conditional distribution of K given S/n = 0 (polygonalized) for n = 16, 64, 256 (broken lines from left to right, respectively) and the cdf of a standard normal random variable (solid line).
Figure 6.5: The exact conditional distribution of $K$ given $S/n = 1/4$ (polygonalized) for $n = 16, 64, 256$ (broken lines from left to right, respectively) and the cdf of a standard normal random variable (solid line).
Figure 6.6: The exact conditional distribution of $K$ given $S/n = 1/2$ (polygonalized) for $n = 16, 64, 256$ (broken lines from left to right, respectively) and the cdf of a standard normal random variable (solid line).
Figure 6.7: The exact conditional distribution of $K$ given $S/n = 3/4$ (polygonalized) for $n = 16, 64, 256$ (broken lines from left to right, respectively) and the cdf of a standard normal random variable (solid line).
Figure 6.8: The exact distribution of $T$ given $S/n = 0$ (solid line, step function), the corrected and uncorrected normal approximations (broken lines from left to right), all for $n = 16$. 
Figure 6.9: The exact distribution of $T$ given $S/n = 0$ (solid line, step function), the corrected and uncorrected normal approximations (broken lines from left to right), all for $n = 64$. 

---

Figure 6.9: The exact distribution of $T$ given $S/n = 0$ (solid line, step function), the corrected and uncorrected normal approximations (broken lines from left to right), all for $n = 64$. 

---
Figure 6.10: The exact distribution of $T$ given $S/n = 1/2$ (solid line, step function), the corrected and uncorrected normal approximations (broken lines from left to right), all for $n = 16$. 
Figure 6.11: The exact distribution of $T$ given $S/n = 1/2$ (solid line, step function), the corrected and uncorrected normal approximations (broken lines from left to right), all for $n = 64$. 
normal approximation is only slightly better than the uncorrected normal approximation, even for \( n = 16 \).

These numerical analyses, based on computation of the exact values of \( P_b(t) \) (which is explained in chapter 7), suggest that for fixed values of \( s/n \), the conditional distribution of

\[
K = \frac{T/n - [1 + (S/n)^2]/2}{[1 - (S/n)^2]/(2n^{1/2})}
\]

is asymptotically normal(0, 1).

In fact, in the next two paragraphs it is shown that

\[
m_T(s, n, 1/4)/n \text{ converges to } m_T(s, n, 1/4)/n = [1+(s/n)^2]/2 \text{ as } n \text{ approaches infinity with } s/n \text{ held constant},
\]

and

\[
s_T^2(s, n, 1/4)/n \text{ converges to } s_T^2(s, n, 1/4)/n = [1-(s/n)^2]/2 \text{ as } n \text{ approaches infinity with } s/n \text{ held constant}.
\]

Numerical results suggest the conclusions that the index of skewness of \( P_b(t) \) approaches 0 as \( n \) approaches infinity, and that the index of kurtosis of \( P_b(t) \) approaches 3 as \( n \) approaches infinity.

In chapter 5 it is shown that \( m_T(s, n, 1/4) \) is given by (5.10). Therefore,

\[
m_T(s, n; 1/4) = \frac{n(n - 1) + s^2}{n} - \frac{n^2 + s^2 + o(n^2)}{n(2n-1)} = \frac{1 + (s/n)^2}{2n^2 + o(n^2)} \to \frac{1 + (s/n)^2}{2}
\]
as \( n \) increases with \( s/n \) fixed. The convergence is illustrated in figure 6.12, where \( m_T(s, n, 1/4)/n \) is graphed as a function of \( s/n \) for \( n = 16, 64, 256 \).

It was also shown in chapter 5 that the exact variance of \( P_b^{T_1} \), \( b = 1/4 \), is \( s_T^2(s, n, 1/4) \) as given in equation (5.11). It follows that

\[
\frac{s_T^2(s, n; 1/4)}{n} = \frac{2(n^2 - s^2)^2 - 2(2n - 1)(n^2 - s^2)}{n} \frac{n}{(2n - 1)^2 (2n - 3)}
\]

\[
= \frac{2n^4[1 - (s/n)^2]^2 + o(n^4)}{8n^4 + o(n^4)} \quad \rightarrow \quad \frac{[1 - (s/n)]^2}{4}
\]

The convergence is illustrated in figure 6.13, where \( s_T^2(s, n, 1/4)/n \) is graphed as a function of \( s/n \) for \( n = 16, 64, 256 \), along with the graph of \( s_T^2(s, n, 1/4)/n \).

Figures 6.14 and 6.15 are graphs of the index of skewness and the index of kurtosis for \( P_b^{T_1} \) (computed as describe in chapters 5 and 7) as a function of \( (s/n) \) for \( n = 16, 64, 256 \). The same data are listed in table 6.3.

Since the actual sizes of the approximate tests are greater than the supposed size, the probability of a type I error (rejecting \( H_0 \) when it is true) is greater than it's supposed to be, a situation which was to be avoided with the highest priority. In light of the ease and low cost with which the exact UMPU-w test (4.26) can be performed using the technique of chapter 5 and 7, there is really no reason to use any of the approximations discussed here.
Figure 6.12: \( m_T(s, n, 1/4)/n \) as a function of \(|s|/n\) for \( n = 16, 64, 256 \) (broken lines), and \( m_T(s, n, 1/4)/n \) (solid line), illustrating the convergence (explanation in text).
Figure 6.13: $s_T(s, n, 1/4)/n$ as a function of $|s|/n$ for $n = 16, 64, 256$ (broken lines), and $s_T^2(s, n, 1/4)/n$ (solid line), illustrating the convergence (explanation in text).
Figure 6.14: The skewness of $P_b^T$ as a function of $|s|/n$ for $n = 16, 64, 256$; $b = 1/4$. 
Figure 6.15: The kurtosis of $P^T_s$ as a function of $|s|/n$ for $n = 16, 64, 256; b = 1/4$. 
Table 6.3: The indices of skewness and kurtosis of $p_b^{Tls} (b = 1/4)$.

<table>
<thead>
<tr>
<th>s/n</th>
<th>Index of Skewness</th>
<th>Index of Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 16$</td>
<td>$n = 64$</td>
</tr>
<tr>
<td>0.0000</td>
<td>-0.00118</td>
<td>-0.00003</td>
</tr>
<tr>
<td>0.0625</td>
<td>0.00354</td>
<td>0.00202</td>
</tr>
<tr>
<td>0.1250</td>
<td>0.01793</td>
<td>0.00826</td>
</tr>
<tr>
<td>0.1875</td>
<td>0.04274</td>
<td>0.01901</td>
</tr>
<tr>
<td>0.2500</td>
<td>0.07934</td>
<td>0.03483</td>
</tr>
<tr>
<td>0.3125</td>
<td>0.12992</td>
<td>0.05659</td>
</tr>
<tr>
<td>0.3750</td>
<td>0.19781</td>
<td>0.08566</td>
</tr>
<tr>
<td>0.4375</td>
<td>0.28817</td>
<td>0.12408</td>
</tr>
<tr>
<td>0.5000</td>
<td>0.40910</td>
<td>0.17501</td>
</tr>
<tr>
<td>0.5625</td>
<td>0.57392</td>
<td>0.24354</td>
</tr>
<tr>
<td>0.6250</td>
<td>0.80595</td>
<td>0.33828</td>
</tr>
<tr>
<td>0.6875</td>
<td>1.14978</td>
<td>0.47498</td>
</tr>
<tr>
<td>0.7500</td>
<td>1.70281</td>
<td>0.68578</td>
</tr>
<tr>
<td>0.8125</td>
<td>2.72772</td>
<td>1.04767</td>
</tr>
<tr>
<td>0.8750</td>
<td>5.29465</td>
<td>1.80305</td>
</tr>
<tr>
<td>0.9375</td>
<td>4.32656</td>
<td>1.88885</td>
</tr>
<tr>
<td>1.0000</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>
7. COMPUTATIONS

In this chapter three computations are explained: (1) the computation of the conditional and unconditional power functions (4.32) and (4.33); (2) computation of the mean, variance, skewness, and kurtosis of $T_b^1$'s; (3) computation of the Neyman structure test (4.26).

The high level computer language SAS (statistical analysis system) is used exclusively (SAS Institute, 1979). The primary reason for using such a high level language is to take advantage of the easily implemented graphing procedures: PROC PLOT, PROC GPLOT, and PROC G3D (SAS Institute, 1979; 1980). Comparisons of distribution functions, test functions, and critical regions—especially in an exploratory context—is eased significantly by the use of these tools. In addition, the use of a high level language virtually eliminates the technical details of input and output, which generally are a feature of lower level languages, by definition.

SAS is designed for data manipulation, statistical analysis, and report writing. Each specific task is accomplished by a separate module of which there are two types, data steps (DATA) and procedures (PROC). Data steps employ input/output statements, declaration statements and programming statements similar to PL/1. Data steps create and/or manipulate datasets (computer files) consisting of series of records called "observations". Each observation is a list of values, one for each SAS "variable" in the dataset. All the statements of a data step are executed for one observation at a time, in the order in which they appear in the dataset (file).
Most of the SAS PROC's are statistical procedures for data analysis. None of that type is used here. Only PROC SORT, which reorders observations of a dataset according to the value of one or more specified variables, and the graphics procedures PROC PLOT, PROC GPLOT, and PROC G3D which plot specified variables of a dataset, both two- and three-dimensionally, are utilized. Thus, SAS datasteps functioning as program modules are used predominantly.

The computation of the power functions (4.32) and (4.33) requires the computation of $P_{b}^{T}S$ and $P_{a}^{S}a = (a_{0}, b)$, for each value of $(a_{0}, b)$ at which the power is desired. To compute discrete distributions numerically requires arithmetic manipulation of many very large (e.g., $n!$) and small (e.g., $1/n!$) numbers. In order to maintain precision, a special method for representing, adding, and multiplying such numbers is required (see Arden, 1963, pp. 93-104). This requirement is fulfilled by three subprograms, SCALE, SUM, and CONVERT (figure 7.1), that perform the necessary operations when they are called from the main program.

Subprograms are constructed within SAS datasteps using the SAS programming statements LINK and RETURN, which accomplish that which would be performed by a PL/1 subprocedure call or a FORTRAN subroutine call (SAS Institute, 1979, p. 115). The main program, which must appear first in the datastep, and subprograms which follow, are separated from each other by RETURN statements. The subprograms are called from the main program and from other subprograms by LINK statements.

The subprogram used most often is called (and labeled) SCALE. The procedure is necessary to prevent overflow and underflow when numbers...
SCALE:
IF Z = . THEN DO;
   PUT 'TRIED TO SCALE MISSING VALUE AT STEP ' _N_ =;
   STOP;
   END;
IF Z = 0 THEN RETURN;
DOWN:
   IF ABS(Z) >= 4 THEN DO;
      Z = Z / 16; E_Z = E_Z + 1;
      GO TO DOWN;
   END;
UP:
   IF ABS(Z) < 1/4 THEN DO;
      Z = Z * 16; E_Z = E_Z - 1;
      GO TO UP;
   END;
RETURN;

SUM:
IF E_V1 > E_V2 THEN DO;
   BIG = V1; E_BIG = E_V1;
   SML = V2; E_SML = E_V2;
   END;
ELSE DO;
   BIG = V2; E_BIG = E_V2;
   SML = V1; E_SML = E_V1;
   END;
SCLE_SML:
   IF E_BIG = E_SML THEN GO TO ADD;
   SML = SML / 16; E_SML = E_SML + 1;
   GO TO SCLE_SML;
ADD:
   Z = BIG + SML; E_Z = E_BIG; LINK SCALE;
   SUM = Z; E_SUM = E_Z;
   RETURN;
CONVERT: LINK SCALE;
   IF ABS(E_Z) > 6 THEN Z = 0;
   ELSE Z = Z * (16 ** E_Z);
   RETURN;

Figure 7.1: Subprograms SCALE, CONVERT, and SUM.
greater than $10^{73}$ or less than $10^{-73}$ arise in computations. The IBM 360/370 system which was used for this project cannot accommodate nonzero numbers of magnitude greater than $10^{73}$ or less than $10^{-73}$ (SAS Institute, 1979). The number $56!$ is greater than $10^{73}$; the number $10^{73}$! must be computed to construct the Neyman structure test for the data of Whiteville, NC strain (n = 1073 cups).

The method used here is analogous to the floating point storage technique of high speed electronic digital computers (Arden, 1963 pp. 97-104, Kennedy and Solomon, 1977, p. 546). Any real variable which might attain an extremely large or small order of magnitude is represented in "canonical base-sixteen duplex form" (CBSDF). By a base-sixteen duplex form (BSDF) of a number is meant any ordered pair $(a, b)$ satisfying

$$x = a \cdot 16^b$$

(7.1)

The CBSDF of a nonzero number $x$ is the unique ordered pair which also satisfies

$$\frac{1}{4} < a < 4.$$

(7.2)

The elements of a BSDF $(a, b)$ are called the mantissa and exponent, respectively (Kennedy and Solomon, 1977, p. 546).

Values are assigned to the SAS variables Z and _Z in the main program before SCALE is called with the statement LINK SCALE. If the purpose is to convert a number $x$ to CBSDF, then the SAS program statements

```
Z = x; _Z = 0; LINK SCALE;
```
greater than $10^{73}$ or less than $10^{-73}$ arise in computations. The IBM 360/370 system which was used for this project cannot accommodate nonzero numbers of magnitude greater than $10^{73}$ or less than $10^{-73}$ (SAS Institute, 1979). The number $56!$ is greater than $10^{73}$; the number $1073!$ must be computed to construct the Neyman structure test for the data of Whiteville, NC strain ($n = 1073$ cups).

The method used here is analogous to the floating point storage technique of high speed electronic digital computers (Arden, 1963 pp. 97-104, Kennedy and Solomon, 1977, p. 546). Any real variable which might attain an extremely large or small order of magnitude is represented in "canonical base-sixteen duplex form" (CBSDF). By a base-sixteen duplex form (BSDF) of a number is meant any ordered pair $(a, b)$ satisfying

$$x = a \cdot 16^b$$  \hspace{1cm} (7.1)

The CBSDF of a nonzero number $x$ is the unique ordered pair which also satisfies

$$\frac{1}{4} < a < 4.$$  \hspace{1cm} (7.2)

The elements of a BSDF $(a, b)$ are called the mantissa and exponent, respectively (Kennedy and Solomon, 1977, p. 546).

Values are assigned to the SAS variables $Z$ and $E_Z$ in the main program before SCALE is called with the statement LINK SCALE. If the purpose is to convert a number $x$ to CBSDF, then the SAS program statements

$$Z = x; E_Z = 0; \text{LINK SCALE;}$$
would appear in the main program. If the purpose is to convert a BSDF 
(c, d) \( (c \stackrel{16}{=} d = x) \) to CBSDF, then the statements

\[
Z = c; E_Z = d; \text{LINK SCALE;}
\]

would appear in the main program. The result is an ordered pair \((a, b)\) 
assigned to the SAS variables \((Z, E_Z)\) that satisfies (7.1) and (7.2).

Accurate multiplication (and division) of base sixteen duplex forms 
(BSDF's) is quite simple with the help of subprogram SCALE. Consider 
multiplication of a BSDF \((a, b)\) by a real number \(c\) or by another BSDF 
\((c, d)\). (The real number \(c\) is equivalent to a BSDF \((c, d)\) with \(d = 0\).) 
The product \((u, v)\) of \((a, b)\) and \((c, d)\) is obtained by first forming the 
duplex \((ac, b + d)\) and then calling SCALE. The result is a CBSDF \((u, v)\) 
such that \(u \stackrel{16}{=} v = ac \stackrel{16}{=} b + d\) as desired.

Addition of two BSDF's \((a, b), (c, d)\) is performed by first 
repeatedly increasing by one the exponent of the lesser of the two 
numbers, \((a, b)\) say, and concurrently dividing its mantissa by 16, until 
the two exponents are equal; then adding the mantissas \(a\) and \(c\); and 
then, finally, by SCALEing the result. This procedure is performed by 
the subprogram SUM, listed in figure 7.1.

The input for SUM is the four SAS variables \(V1, E_V1, V2, E_V2,\) 
which represent respectively the mantissa and exponent of "value 1" and 
the mantissa and exponent of "value 2". Continuing the above example, 
the statements

\[
V1 = a; E_{V1} = b; V2 = c; E_{V2} = d; \text{LINK SUM;}
\]

would appear in the main program to call SUM.

Also listed in figure 7.1 is subprogram CONVERT which changes a 
BSDF \((a, b)\) into a real number \(x\). CONVERT merely computes \(x = a \stackrel{16}{=} b\) if
the absolute value of \( b \) is greater than or equal to 6, and \( x = 0 \) otherwise. The algorithm is adequate because it operates only on numbers that represent probabilities, that is, numbers between zero and one.

As mentioned above, computation of the power functions (4.32) and (4.33) requires the computation of \( p_{b}^{T_{1}s} \) and \( p_{a}^{S} \) at each point in parameter space at which the power is desired. Regarding \( a_{0} \) as an auxiliary parameter, these functions are computed for \( b = .01, .02, \ldots, .04, \) and for \( a_{0} = .01, .04, .013514 \) (for TAL strain; figure 4.4). This means that for each value of \((a_{0}, b)\), \( p_{b}^{T_{1}s}(t) \) must be computed for each \((s, t)\) in \( W_{ST} \) (4.9), and \( p_{a}^{S}(s) \) (4.13) must be computed for \( s = -n, \ldots, n \).

Note the symmetry revealed in equation (4.16). For each fixed value of \( t \) in the range (4.14), \( P\{ T = t \mid S = s \} \) is an even function of \( s \). The numerator of (4.16) depends on \( s \) only in the factors \( [(t-s)/2]! \) and \( [t+s)/2]! \). The numerator is therefore completely determined by the magnitude (absolute value) of \( s \), regardless of the sign (+ or -). The same is true for the summand and the range of summation (4.14) in the denominator of (4.16). Thus, the entire expression (4.16) depends only on the absolute value of \( s \), that is,

\[
P\{ T = t \mid S = s \} = P\{ T = t \mid S = -s \} = P\{ T = t \mid S = |s| \},
\]

or equivalently,

\[
P\{X_{0} = (t-s)/2 \mid S = s \} = P\{X_{0} = (t+s)/2 \mid S = -s \}.
\]
We therefore need compute \( P(T = t \mid S = s) \) only for nonnegative values of \( s \). The scheme that is actually used is to compute

\[
P_b^{T \mid S}(s + 2x) = P_b\{T = s + 2x \mid S = s\} = P_b\{(T - S)/2 = x \mid S = s\}
\]

for all \( x \) in the range

\[
(|s| - s)/2 = \max(0, -s) \leq x \leq \lceil(n - s)/2\rceil,
\]

for each \( s \) such that \( s = 0, 1, 2, \ldots, n \). Thus, letting

\[
F_s(x) = \frac{b^{(s+2x)/2}}{x! (n-s-2x)! (s+x)!},
\]

\[
G(s) = \sum_x F_s(x),
\]

\[
H(s) = \left(\frac{a_2}{a_0}\right)^{s/2},
\]

and

\[
M = n! a_1^n,
\]

we obtain

\[
P_b^{T \mid S}(s + 2x) = P_b\{X_0 = x \mid S = s\} = F_s(x)/G(s),
\]

\[
P_a^S(s) = M H(s) G(s).
\]

It therefore follows from formula (7.8), definition (7.4) and the discussion of symmetry in the numerator and denominator of (4.16) that

\[
G(s) = G(-s) = G(|s|).
\]

The densities and power functions are computed in a program composed of three modules, SAS datasteps FREQ, TOTAL, and POWER. The first module, FREQ, computes \( F_s \).
An algorithm to compute $F_s(x)$ for each fixed $s = 0, 1, 2, ..., n$ is constructed as follows. For $x = 1, 2, 3, ..., [(n - s) / 2]'$,

$$F_s(x-1) = b^{[s+2(x-1)]/2} / \{(x-1)! (s+x-1)! [n-s-2(x-1)]!\}$$

$$= b^{[s+2(x-1)]/2} / ((x-1)! (s+x-1)! (n-s-2x+2)!)) \cdot$$

Thus,

$$F_s(x) = b^{(s+2x)/2} / ((x) (s+x)! (n-s-2x)!)$$

$$= \frac{b^{[s+2(x-1)]/2}}{x(x-1)! (s+x)(s+x-1)! [n-s-2(x-1)]! / [(n-s-2x+2)(n-s-2x+1)]} \cdot$$

that is,

$$F_s(x) = F_s(x-1) b (n-s-2x+2)(n-s-2x+1) / [x(s+x)]. \quad (7.10)$$

The algorithm is

**Algorithm 7.1:** Calculate the numbers $F_s(0), F_s(1), ..., F_s([(n - s) / 2]')$ recursively by the relation

$$F_s(0) = b^{s/2} / [s! (n-s)!], \quad (7.11)$$

$$F_s(x) = F_s(x-1) b (n-s-2x+2)(n-s-2x+1) / [x(s+x)], \quad (7.12)$$

$$x = 1, 2, ..., [(n - s) / 2]'. \quad$$

Formula (7.11) follows from (7.4), and formula (7.12) is the same as (7.10).
The process of calculating $F_s(x)$ for each $s$ and $x$ is further streamlined by the following algorithm for computing $F_s(0)$ for each $s = -n, \ldots, n$.

Algorithm 7.2: Calculate the numbers $F_0(0), F_1(0), \ldots, F_n(0)$ recursively by the relations

\[
F_0(0) = \frac{1}{n!},
\]

\[
F_s(0) = F_{s-1}(0) b^{1/2} \frac{(n - s + 1)}{s},
\]

\[s = 1, 2, \ldots, n.\]

Algorithm 7.2 follows from the definition (7.11) of the function $F_s(0)$.

These two algorithms are implemented in SAS datastep FREQ (listed in figure 7.2) to compute the non-normalized "frequencies" $F_s(x)$.

The version of FREQ given in figure 7.2 computes $F_s(x)$ for $b = .01, .02, \ldots, .25$. This is accomplished by employing ARRAYS (subscripted variables) and DO-END loops. The 25 values of $F_s(x)$ are stored in CBSDF in a pair of SAS ARRAYS called F and E_F ("E" for "exponent"), with elements F1, F2, ..., F25, and E_F1, E_F2, ..., E_F25, indexed by SAS variable K, respectively. In datastep FREQ the values of $b$ are not stored in an ARRAY; rather, they are generated at each iteration by $b = .01 \times K$. Table 7.1 is a translation table for the SAS variables used in the various programs being described here and their representation in the text.
DATA EXACT.FREQ;
N = 74;
ARRAY F0 (K) F0 1-F0 25; ARRAY E_F0 (K) E_F0 1-E_F0 25;
ARRAY F (K) F1-F25; ARRAY E_F (K) E_F1-E_F25;
RETAIN F0 1—E_F0 25 F1—E_F25;
DO S = 0 TO N;
  IF S = 0 THEN DO;
    Z = 1; E_Z = 0;
    DO I = 1 TO N;
      Z = Z * I; LINK SCALE;
    END;
    N_FAC = Z; E_N_FAC = E_Z;
  END;
  DO X = 0 TO FLOOR((N-S)/2);
    T = S + 2*X;
    DO K = 1 TO 25;
      IF T = S THEN DO;
        IF S = 0 THEN DO;
          F0 = 1 / N_FAC; E_F0 = -E_N_FAC;
        END;
        ELSE DO;
          F0 = F0 * SQRT(.01*K) * (N-S+1) / S;
        END;
        Z = F0; E_Z = E_F0; LINK SCALE; F0 = Z; E_F0 = E_Z;
        F = F0; E_F = E_F0;
      END;
      ELSE DO;
        F = F * K * .01 * (N-S-2*X+2) * (N-S-2*X+1) / (X*(X+S));
      END;
      Z = F; E_Z = E_F; LINK SCALE; F = Z; E_F = E_Z;
    END;
    OUTPUT;
  END;
END;
KEEP S T F1—E_F25;
RETURN;

Figure 7.2: Datastep (program module) FREQ, which computes $F_s(x)$ (7.4).
The second program module (SAS datastep TOTAL, figure 7.3) computes \( G(s) \) and \( P^S_a(s) \) for \( s = -n, \ldots, n \).

SAS datastep FREQ, the first program module, produces a dataset containing, for each \( s \geq 0 \), \( f_s(x) \), \( x = 0, \ldots, \left\lceil \frac{(t-s)}{2} \right\rceil' \). This dataset is sorted in such a way that for each \( s \) the records are in ascending order of \( f_s(x) \) (PROC SORT). Control then passes to datastep TOTAL. Datastep TOTAL computes \( G(s) \) and \( P^S_a(s) \) simultaneously as it creates a new dataset (TOTAL) from the input dataset FREQ, sorted as above.

For each \( s \), \( G(s) \) is computed simply by accumulating the sum of \( f_s(x) \). Thus \( G(s) \) is computed as prescribed in (7.5) with the added provision that the order of addition is from smallest to largest, as a result of the way the input dataset (FREQ) was sorted. This often used trick minimizes truncation error in the summation (D. F. McAllister, J. F. Monahan, personal communication).

For \( s = -n, \ldots, n \), \( P^S_a(s) \) is computed by the following algorithms based on equation (7.6), (7.7), and (7.9).

The number \( M \) (7.7) is computed once at the beginning of the datastep because it doesn't depend on \( s \). The following algorithm is used.

Algorithm 7.3: Compute the numbers \( M_0, M_1, M_2, \ldots, M_n \) recursively by the relations

\[
M_0 = 1, \quad M_i = i a_1 M_{i-1}, \quad i = 1, 2, \ldots, n.
\]

The result is

\[
M_n = \prod_{i=1}^{n} i a_1 = n! a_1^n = M.
\]
PROC SORT DATA=EXACT.FREQ; BY S E F25 F25;
DATA EXACT.TOTAL; SET EXACT.FREQ; BY S;
IF _N_ = 1 THEN DO;
   N = 74;
   ARRAY F (K) F1-F25; ARRAY E_F (K) E_F1-E_F25;
...

ARRAY THO (J) TH01-TH04;
TH01 = .16; TH02 = .04; TH03 = .0135141; TH04 = .0001;
DO J = 1 TO 4; DO K = 1 TO 25;
   PHI = .01 * K;
   TH1 = (-TH0 + SQRT(TH0**2 + 4*PHI*TH0*(1 - TH0))) / (2*PHI);
   M = 1; E_M = 0;
   DO I = 1 TO N; M = M * I * TH1;
      Z = M; E_Z = E_M; LINK SCALE; M = Z; E_M = E_Z;
   END; END;

RETAIN N H101--E G25; *** ITERATE ***;
IF FIRST.S THEN DO K = 1 TO 25; G = F; E_G = E_F; END;
ELSE DO K = 1 TO 25;
   V1 = F; E_V1 = E_F; V2 = G; E_V2 = E_G; LINK SUM;
   G = SUM; E_G = E_SUM;
END;
IF LAST.S THEN DO;
   DO J = 1 TO 4; DO K = 1 TO 25;
      PHI = .01 * K;
      TH1 = (-TH0 + SQRT(TH0**2 + 4*PHI*TH0*(1 - TH0))) / (2*PHI);
      TH2 = 1 - TH0 - TH1;
      IF S = 0 THEN DO; H=M; E_H=E_M; L=M; E_L=E_M; END;
      ELSE DO; H=H*SQRT(TH2/TH0); L=L/SQRT(TH2/TH0); END;
      Z = H; E_Z = E_H; LINK SCALE; H = Z; E_H = E_Z;
      Z = L; E_Z = E_L; LINK SCALE; L = Z; E_L = E_Z;
      PH = G * H; E_PH = E_G + E_H; PL = G * L; E_PL = E_G + E_L;
   END; END;
OUTPUT;
END;
KEEP S G1--E_G25 PH101--E_PH425 PL101--E_PL425;
RETURN;

Figure 7.3: Datasep (program module) TOTAL, which computes G(s) (7.5),
H(s) (7.6), M (7.7), and P^S(s) for s = -n, ..., n; four values of the
auxiliary parameter a_0 (SAS variables TH01 - TH04); and 25 values of b
(SAS variable PHI; see table 7.1).
The computation is accurate because at each iteration the product is scaled (SCALEd) to maintain precision.

Next, for each fixed $s = -n, \ldots, n$, the product $M H(s)$ is computed by the following algorithm.

**Algorithm 7.4:** Compute the numbers $H_0, H_1, \ldots, H_n; L_0, L_1, \ldots, L_n$ recursively by the relations

\[
H_0 = L_0 = M;
\]

\[
H_s = (a_2 / a_0)^{1/2} H_{s-1}, \quad L_s = (a_2 / a_0)^{-1/2} L_{s-1},
\]

$s = 1, 2, \ldots, n$.

Thus,

\[
H_s = M \prod_{j=1}^{s} (a_2 / a_0)^{1/2} = M (a_2 / a_0)^{s/2} = M H(s),
\]  

\[ s = 0, 1, 2, \ldots, n, \tag{7.15a} \]

\[
L_s = M \prod_{j=1}^{s} (a_2 / a_0)^{1/2} = M (a_2 / a_0)^{s/2} = M H(-s),
\]  

\[ s = 0, 1, 2, \ldots, n. \tag{7.15b} \]

Like $L_s$ and $H_s$, $F^S_a(s)$ is computed when the datastep reaches a record (SAS observation) in the input dataset (FREQ) where

$x = [(n-s)/2]'$ (LAST.S). At this point the modified record of the dataset being created (TOTAL) contains the variables $G(s) = G(-s)$, $L_s = M H(-s)$, $H_s = M H(s)$, and $M$. $F^S_a(s)$ is computed as prescribed in (7.9) by
Only this record is output, one for each \( s \), so the created dataset 
(TOTAL) contains the values for \( G(s) \), \( P_a^S(s) \), \( P_a^S(-s) \), \( s = 0, 1, \ldots, n \).
The datasets FREQ and TOTAL are stored on disk for reuse later.

The third module of the program can be used to compute the conditional pdf of \( T \), \( p_{b}^{T|s}(t = s + 2x) \); and/or the joint pdf of \( S \) and \( T \), 
\( p_{a}^{ST}(s, t = s + 2x) \), and/or the corresponding cumulative distribution functions (cdf's), depending on the desired result. The desired result, be it the densities or distributions themselves, the Neyman structure critical function, the conditional or unconditional power function of the Neyman structure test, the corresponding critical regions and confidence levels, and so forth, can be computed in this module also.

Only the desired results are output and stored; the distributions and densities are recomputed from the data of FREQ and TOTAL whenever they are needed. It is more economical to recompute the distributions based on the data stored in FREQ and TOTAL than to output and store on disk what would be a very large dataset.

The input for the third module is the data of FREQ and TOTAL, sorted and merged by \( n \), \( s \), and \( t \). \( p_{b}^{T|s}(t) \), \( p_{b}^{T|s}(T \leq t) \) and \( B'(a_0, b) \) are computed simply by formula (7.8),

\[
p_{b}^{T|s}(T \leq s + 2x) = \sum_x p_{b}^{T|s}(s + 2x), \quad (7.17)
\]
(where summation is over (7.3)), and (4.33), respectively.
Table 7.1: SAS variables and ARRAYs used in datasteps FREQ, TOTAL, and POWER and their symbolic representation in the text.

<table>
<thead>
<tr>
<th>ARRAY or variable</th>
<th>Subscript, if an ARRAY</th>
<th>Representation in text</th>
<th>Text reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td></td>
<td>n</td>
<td>(2.12)</td>
</tr>
<tr>
<td>FO, E_FO</td>
<td>K</td>
<td>F_s(0)</td>
<td>(7.13), (7.14)</td>
</tr>
<tr>
<td>F, E_F</td>
<td>K</td>
<td>F_s(x)</td>
<td>(7.4)</td>
</tr>
<tr>
<td>N_FAC, E_N_FAC</td>
<td>K</td>
<td>n!</td>
<td>(2.4)</td>
</tr>
<tr>
<td>X, S, T</td>
<td></td>
<td>x (value of X_0), s, t</td>
<td></td>
</tr>
<tr>
<td>TH0</td>
<td>J</td>
<td>a_0</td>
<td>(2.4)</td>
</tr>
<tr>
<td>PHI</td>
<td></td>
<td>b</td>
<td>(4.18)</td>
</tr>
<tr>
<td>TH1, TH2</td>
<td></td>
<td>a_1, a_2</td>
<td>(2.4)</td>
</tr>
<tr>
<td>M, E_M</td>
<td>J, K</td>
<td>M</td>
<td>(7.7)</td>
</tr>
<tr>
<td>G, E_G</td>
<td>K</td>
<td>G(s)</td>
<td>(7.5)</td>
</tr>
<tr>
<td>H, E_H, L, E_L</td>
<td>J, K</td>
<td>H_s, L_s</td>
<td>(7.15)</td>
</tr>
<tr>
<td>PH, E_PH, PL, E_PL</td>
<td>J, K</td>
<td>P_b^s(s)</td>
<td>(7.16)</td>
</tr>
<tr>
<td>P, E_P</td>
<td>K</td>
<td>P_b^{Tls}(t)</td>
<td>(7.8)</td>
</tr>
<tr>
<td>CD, E_CD</td>
<td>K</td>
<td>P_b^{Tls}[T &lt; t]</td>
<td>(7.17)</td>
</tr>
<tr>
<td>LAG, E_LAG</td>
<td>K</td>
<td>P_b^{Tls}[T ≤ t-2]</td>
<td></td>
</tr>
<tr>
<td>PWS, E_PWS</td>
<td>J, K</td>
<td>B(b</td>
<td>s)</td>
</tr>
<tr>
<td>PW, E_PW, POWER</td>
<td>J, K</td>
<td>B'(a_0, b)</td>
<td>(4.33)</td>
</tr>
<tr>
<td>THETA_0, THETA_1, THETA_2</td>
<td></td>
<td>a_0, a_1, a_2</td>
<td>(2.4)</td>
</tr>
<tr>
<td>PARAM_P, PARAM_Q</td>
<td>J, K</td>
<td>p, q</td>
<td>(3.9), (3.10)</td>
</tr>
</tbody>
</table>
For the computation of the power functions, $P_{14}^{T|s}$ and $P_{a}^{ST}$ must be computed for each value of $a$ at which the power is desired. A module (DATA POWER) to compute these for $n = 74$, and 100 values of $a$ is listed in figure 7.4; $B'(a_0, b)$ is graphed in figure 4.4. If $P_{a}^{ST}(s, t)$, the Neyman structure test, and so forth are desired, additional statements are added to datastep POWER, sometimes creating several different datasets in the same datastep, to produce the desired function(s).

The conditional power function $B_0(a)$ of the Neyman structure test $f$ have been computed for $n = 74$, 25 values of $b$ and 4 values of $a_0$ (thus 100 values of $a$). These are illustrated in figures ZZ.

The computation of the Neyman structure test (specifically, the expected boundary of the critical region) requires computation of $P_{b}^{T|s}$ only under the null hypothesis (4.19) (i.e., only for $b = 1/4$). This can be accomplished in one data step using algorithms based on fact 5.1.

Formula (5.5) of fact 5.1 leads the way to a one-datastep computation of $P_{b}^{T|s}$ for $b = 1/4$, and thereby a one-datastep computation of the Neyman structure tests (i.e., its critical level given an observation and/or its expected critical region boundaries given $n$, etc.).

**Algorithm 7.5:** Compute the numbers $P\{X_0 = 0 \mid S = s\}$, $s = 0, 1, \ldots, n$ recursively by the relations (5.14) and (5.13); for each $s = 0, 1, 2, \ldots, n$, compute the numbers $P\{X_0 = x \mid S = s\}$, $x = 1, 2, \ldots, [(n-s)/2]'$ recursively by the relation (5.12).
PROC SORT DATA=EXACT.FREQ; BY S T; PROC SORT DATA=EXACT.TOTAL; BY S;
DATA EXACT.POWER (KEEP=PHI THETA_0-THETA_2 PARAM_P PARAM_Q POWER);
MERGE EXACT.FREQ EXACT.TOTAL; BY S;
*** INITIALIZE ***; N = 74; A = .05;
TH01 = .16; TH02 = .04; TH03 = .0135141; TH04 = .0001;
*** declaration of ARRAYS ...
RETAIN CD1--E CD25 PWS1--E PWS5 PW101--E PW425 T_ALPHA;
DO K = 1 TO 25; *** ITERATE - COMPUTE PDF(T;S) AND CDF(T;S) ***;
    Z = F / G; E_Z = E_F - E_G; LINK SCALE; P = Z; E_P = E_Z;
    IF T = S THEN DO; CD = P; E_CD = E_P; LAG = 0; E_LAG = 0; END;
    ELSE DO; LAG = CD; E_LAG = E_CD;
        V1 = CD; E_V1 = E_CD; V2 = P; E_V2 = E_P; LINK SUM; CD = SUM; E_CD = E_SUM;
    END;
END;  
*** COMPARE CONDITIONAL CDF WITH A=SIZE W ***;
Z = LAG25; E_Z = E_LAG25; LINK CONVERT; L = Z;
Z = CD25; E_Z = E_CD25; LINK CONVERT; H = Z;
IF L <= A AND H > A THEN DO; U = 1;
    RANDOM = (A - L) / (H - L); T A = T - 2 + RANDOM;
    DO K = 1 TO 25; *** CONDITIONAL POWER GIVEN S = PWS ***;
        Z = P * RANDOM; E_Z = E_P * RANDOM; LINK SCALE;
        V1 = Z; E_V1 = E_Z; V2 = LAG; E_V2 = E_LAG; LINK SUM;
        PWS = SUM; E_PWS = E_SUM;
    DO J = 1 TO 4; **** POWER ****;
        IF S = 0 THEN DO; *** PW=UNCONDITIONAL (THO, PHI) ***;
            Z = PH * PWS; E_Z = E_PH * E_PWS; LINK SCALE;
            PW = Z; E_PW = E_Z;
        END;
        ELSE DO;
            V1 = PH; E_V1 = E_PH; V2 = PL; E_V2 = E_PL; LINK SUM;
            Z = SUM * PWS; E_Z = E_SUM * E_PWS; LINK SCALE;
            V1 = PW; E_V1 = E_PW; V2 = Z; E_V2 = E_Z; LINK SUM;
            PW = SUM; E_PW = E_SUM;
    IF N = S THEN DO; *** OUTPUT POWER ***;
        Z = PW; E_Z = E_PW; LINK CONVERT; POWER = Z;
        LINK TRANS;
        OUTPUT EXACT POWER;
    END; END; END; END; END; END;
RETURN;
TRANS:  *** SUBPROGRAMS ***;
    PHI = .01 * K;
    THETA_0 = THO;
    THETA_1 = (-THO + SQRT(THO ** 2 + 4 * PHI * THO * (1 - THO))) / (2 * PHI);
    THETA_2 = THETA_0 - THETA_1;
    PARAM_P = 1 - SQRT(THETA_0); PARAM_Q = THETA_2 / (PARAM_P ** 2); RETURN;

Figure 7.4: The third datastep (program module) which computes the cdf's $P_b^{T_s}(t)$, $P_a^{ST}(s, t)$; the conditional and unconditional power functions of the Neyman structure test.
This algorithm is used in two programs: one which computes the
critical levels, and expected critical region boundaries of the Neyman
structure tests for given realizations of $S$ and $T$ (eg., datastep
NULL.SAMPLE, figure 7.5); another that computes the expected critical
region boundaries of the Neyman structure test (4.26) for each value of
$s$, for given $n$ and $w$ (the size of the test; eg., datastep NULL.REGIONS,
figure 7.6). The former of these produced the last two columns of of
table 4.1; the latter produced the data for figures 6.4-6.7.

We also have the following algorithm for computation of the first
four factorial moments of $X_0$ given $S = s$.

**Algorithm 7.6:** Compute the numbers $m_{(k)}(s, n, 1/4)$, $k = 1, 2, 3,$
4, recursively by the relations (5.9) for $k = 0$ and (5.8) for $k = 1, 2,$
3, 4.

This algorithm is the basis of the program SAS datastep MOMENTS listed
in figure 7.7 In that procedure the mean, variance, skewness and
kurtosis of the distribution of $T$ given $S = s$, $P^{T:s}_b$, $b = 1/4$, are
computed as functions of the factorial moments of the distribution of $X_0$
given $S = s$. The relationships used are well known (Johnson and Kotz,
1969, pp. 18, 19).
DATA NULL; SAMPLE; INPUT STRAIN $ DO D1 D2;
N = DO + D1 + D2; T_OBS = DO + D2; S_OBS = D2 - DO;
DO S = 0 TO ABS(S_OBS);
   IF S = 0 THEN DO;
      Z = 1; E_Z = 0;
      DO J = 1 TO N;
         Z = Z * J / (2 * J - 1); LINK SCALE; END; END;
   ELSE DO;
      Z = PO * (N + S) / 2 / S; E_Z = E_PO; LINK SCALE; END;
      PO = Z; E_PO = E_Z; END;
S = ABS(S_OBS);
   DO X = 0 TO FLOOR((N - S) / 2);
      T = S + 2*X;
      IF T = S THEN DO;
         P = PO; E_P = E_PO;
         CD = P; E_CD = E_P;
         LAG = 0; E_LAG = 0; END;
      ELSE DO;
         Z = P * (N-S-2*X+2) * (N-S-2*X+1) / (4*X*(S+X)); E_Z = E_P;
         LINK SCALE; P = Z; E_P = E_Z;
         LAG = CD; E_LAG = E_CD;
         V1 = CD; E_V1 = E_CD; V2 = P; E_V2 = E_P; LINK SUM;
         CD = SUM; E_CD = E_SUM; END;
      END;
   END;
   Z = LAG; E_Z = E_LAG; LINK CONVERT; L = Z;
   Z = CD; E_Z = E_CD; LINK CONVERT; H = Z;
   IF L <= .025 AND .025 < H THEN DO;
      R_X_025 = (.025 - L) / (H - L); T_R_X_025 = T;
      T_X_025 = T - 2 + 2 * R_X_025; END;
   IF L <= .05 AND .05 < H THEN DO;
      R_X_05 = (.05 - L) / (H - L); T_R_X_05 = T;
      T_X_05 = T - 2 + 2 * R_X_05; END;
   IF L <= .975 AND .975 < H THEN DO;
      R_X_975 = (H - .975) / (H - L); T_R_X_975 = T - 2;
      T_X_975 = T - 2 * R_X_975; END;
   IF T = T_OBS THEN DO;
      Z = CD; E_Z = E_CD; LINK CONVERT; C_L_XACT = Z; END; END;
CARDS;

Figure 7.5: Dataseq SAMPLE which computes the critical levels of the Neyman structure tests for given data. Here the data of table 1.1 (Gould, et al., 1980) has been used to produce the critical levels reported in table 4.1.
DATA NULL.REGIONS APPROX.CDF (KEEP=S_OVER_N N T XACT_CDF );
DO N = 16, 64, 256;
DO S = 0 TO 7 * N / 8;
    IF S = 0 THEN DO; * COMPUTE P0 = P(T=S;S=0) = P(X0=0;S=0);
        Z = 1; E_Z = 0;
        DO J = 1 TO N;
            Z = Z * J / (2 * J - 1); LINK SCALE; END; END;
    ELSE DO; * COMPUTE P0 = P(T=S;S>0) = P(X0=0;S>0);
        Z = P0 * (N + S) / 2 / S; E_Z = E_P0; LINK SCALE; END;
        P0 = Z; E_P0 = E_Z;
        DO X = 0 TO FLOOR((N - S) / 2);
            T = S + 2*X;
            IF T = S THEN DO;
                P = P0; E_P = E_P0;
                CD = P; E_CD = E_P;
                LAG = 0; E_LAG = 0; END;
            ELSE DO; -
                Z = P * (N-S-2*X)*((N-S-2*X+1)/(4*X*(S+X))); E_Z = E_P;
                LINK SCALE; P = Z; E_P = E_Z;
                LAG = CD; E_LAG = E_CD;
                V1 = CD; E_V1 = E_CD; V2 = P; E_V2 = E_P; LINK SUM;
                CD = SUM; E_CD = E_SUM; END;
        Z = LAG; E_Z = E_LAG; LINK CONVERT; L = Z;
        Z = CD; E_Z = E_CD; LINK CONVERT; H = Z;
        XACT_CDF = Z;
        IF L <= .025 AND .025 < H THEN DO;
            T_R_X025 = T; T_X_025 = T - 2 + 2*(.025 - L)/(H - L);
            END;
        IF L <= .05 AND .05 < H THEN DO;
            T_R_X05 = T; T_X_05 = T - 2 + 2*(.05 - L)/(H - L); END;
        IF L < .975 AND .975 <= H THEN DO;
            T_R_X975 = T - 2; T_X_975 = T - 2*(H - .975)/(H - L); END;
        * OUTPUT DATA APPROX.CDF * ;
            IF S/N=0 OR S/N=1/8 OR S/N=1/4 OR S/N=1/2 OR S/N=3/4 OR S/N=7/8
            THEN DO;
                S_OVER_N = S / N;
                OUTPUT APPROX.CDF; END; END; END; END;
RETURN;

Figure 7.6: Datastep REGIONS which computes the expected critical
region boundaries of the Neyman structure tests given n and w. This
program produces the data for figures 6.1-6.3.
DATA NULL.MOMENTS;
ARRAY MF (K) MF1-MF4;
DO N = 74;
  DO S = -N TO N;
    * COMPUTE FACTORIAL MOMENTS FOR X = S + 2X GIVEN S;
      Z = 1;
      DO K = 1 TO 4;
        Z = Z * (N-S-2*K+2)*(N-S-2*K+1) / (2*(2*N-2*K+1));
        MF = Z;
      END;
    * COMPUTE ABSOLUTE MOMENTS;
      MA1 = MF1;
      MA2 = MF2 + MF1;
      MA3 = MF3 + 3 * MF2 + MF1;
      MA4 = MF4 + 6 * MF3 + 7 * MF2 + MF1;
    * COMPUTE CENTRAL MOMENTS;
      MC2 = MA2 - MA1**2;
      MC3 = MA3 - 3 * MA2 * MA1 + 2 * MA1**3;
      MC4 = MA4 - 4 * MA3 * MA1 + 6 * MA2 * MA1**2 - 3 * MA1**4;
    * COMPUTE MEAN VARIANCE SKEWNESS AND KUTOSIS OF T;
      MEAN = S + 2 * MA1;
      VARIANCE = 4 * MC2;
      SKEWNESS = MC3 / (SQRT(MC2))**3;
      KURTOSIS = MC4 / MC2**2;
  OUTPUT;
END;
END;

Figure 7.7: SAS datastep MOMENTS to compute the mean, variance, skewness and kurtosis of P^Tis, b = 1/4. This particular version produced the data for figures 5.1 and 5.2.
8. REFERENCES CITED


Johnson, M. W., R. E. Stinner, and R. L. Rabb (1975), Ovipositional
response of Heliothis zea (Boddie) to its major hosts in North

Mifflin, Boston, 328 pp.

Kennedy, M. and M. B. Solomon (1977), Structured PL/ZERO Plus PL/ONE,
Prentice Hall, Englewood Cliffs, N. J.

Kirkpatrick, T. W. (1957), Insect Life in the Tropics, Longmans, London,
311 pp.

Lancaster, H. O. (1969), The Chi-squared Distribution, Wiley, New York,
356 pp.

Laster, M. L. (1972), Intraspecific hybridization of Heliothis virescens
and Heliothis subflexa, Environ. Entomol., 1, 682-687.

369 pp.

Neyman, J. and E. S. Pearson (1933), On the problem of the most
efficient tests of statistical hypotheses, Phil. Trans. Roy. Soc. A
231, 289-337.

Pearson, K. (1900), On a criterion that a given system of deviations
from the probable in the case of a correlated system of variables
is such that it can be reasonably supposed to have arisen from
random sampling, Philos. Mag., (5) 50, 157-175.

pp.


Francisco, 776 pp.

Spedding, C. R. W. (1980), Systems concepts in agriculture: an overview,
Concepts in Agriculture, Biomathematics Series No. 2, Institute of
Statistics Mimeo Series No. 1269, North Carolina State University,
Raleigh, 9-26.

Stinner, R. E., J. W. Jones, C. Tuttle, and R. E. Caron (1977),
Population mortality and cyclicity as affected by intraspecific
