Discrete Small Sample Asymptotics

Steven J. Kathman, Jr.

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

George R. Terrell, Chair
Eric P. Smith
Keying Ye
Robert V. Foutz
Marion R. Reynolds

December 7, 1999
Blacksburg, Virginia

Keywords: Poisson approximation, Tilting, Generating function

Discrete Small Sample Asymptotics

Steven J. Kathman Jr.

(ABSTRACT)

Random variables defined on the natural numbers may often be approximated by Poisson variables. Just as normal approximations may be improved by saddlepoint methods, Poisson approximations may be substantially improved by tilting, expansion, and other related methods. This work will develop and examine the use of these methods, as well as present examples where such methods may be needed.

Contents

List of Tables

Chapter 1

Introduction

One of the oldest limit theorems in probability is the Poisson 'law of small numbers'. In it's simplest form, it states that the binomial distribution converges to a Poisson distribution as n approaches infinity, if $p = \lambda/n$ for some $\lambda > 0$, see Barbour(1992). This suggests that the Poisson distribution may be used to approximate binomial probabilities, and that the approximation works better if n is moderate to large and p is small. Similar theorems exist for other discrete distributions. One of the most appealing features of Poisson approximations is that they only require the mean of the distribution being approximated. Unfortunately Poisson approximations often prove to be inadequate, especially for points that are not near the mean of the distribution.

It will be shown that given some extra information, it is possible to improve Poisson approximations. The extra information will usually be obtained from generating functions, and thus they are the focus of chapter 3. Since many of the techniques are analogous to those used to improve normal approximations, a brief literature review summarizing some of the advances in improving normal approximations is presented in chapter 2. Chapter 2 also mentions what has previously been done to improve Poisson approximations.

Chapter 4 is the beginning of the development of the techniques used to improve Poisson approximations. This chapter is devoted to expansion methods, one of which is analogous to the Edgeworth expansion. Chapter 5 continues the development of techniques by examining the use of tilting. One form of tilting that is discussed is exponential tilting. Chapter 6 explores the possibilities of combining the techniques of chapters 4 and 5. The saddlepoint methods are the result of combining these methods for normal approximations. Thus chapter 6 may be viewed as forming methods that are analogous to saddlepoint methods for Poisson approximations. Throughout these chapters simple distributions, such as the binomial or hypergeometric, are used to illustrate how the techniques work. Each technique is followed by a simple example.

Chapter 7 is devoted to more complicated examples. Most of the examples in chapter 7 are important to the area of applied probability. Chapter 8 discusses some possible future work, which includes the possibility of improving binomial approximations. Chapter 9 consists of the conclusion, which contains a brief summary of some of the key issues, as well as discusses which techniques to use in various circumstances.

An important issue in this work is approximating $P(X \geq y)$ or $P(X \leq y)$ where X is the random variable from the distribution being approximated. These two probabilities are referred to as right tail and left tail probabilities respectively. Also, points that are not close to the mean are referred to as being in the tails of the distribution. Since this is the region where Poisson approximations tend to be the least accurate, it receives the most attention in this work.

Chapter 2

Literature Review

The Gaussian distribution plays a large role in asymptotic theory. One of the oldest asymptotics results is the Central Limit Theorem, which in it's simplest form states that the distribution of the sample mean is asymptotically Gaussian. As is well known in statistics, the sample mean is not the only statistic with an asymptotic Gaussian distribution. Asymptotic normality is used in almost every area of statistics. However for finite sample sizes, Gaussian approximations often prove to be inadequate. In these situations statisticians have found methods for improving them.

One of the oldest methods for improving Gaussian approximations is the Edgeworth expansion. The Edgeworth expansion is discussed thoroughly in Hall (1992), Bederick and Hill (1992), and Barndorff-Nielsen and Cox (1979). It is most useful for improving approximations for points near the mean of the density. The approximation is easy to implement since all that is required are the cumulants, which may be readily obtained from the moments of the distribution. Unfortunately, the approximations due to the Edgeworth expansion are usually not accurate for points in the tails of the distribution, and using it may also lead to negative probability estimates. The Edgeworth expansion is mentioned again briefly in Chapter 4.

One method for obtaining accurate approximations in the tails of a distribution is exponential tilting. Exponential tilting is discussed thoroughly in Terrell (1999), Reid (1988), Daniels (1987), and Barndorff-Nielsen and Cox (1979). The idea behind exponential tilting appears to have originated with Esscher (1932), see Reid (1988). The basic idea behind exponential tilting is to transform the density to have a mean at the point of interest. Then approximate the transformed density with some common distribution, usually the Gaussian. After the approximation, transform the density back to obtain an approximation to the original density. Exponential tilting has been shown to lead to accurate approximations for points in the tails of the distribution. Jin and Robinson (1999) discuss performing exponential tilting, and choosing different distributions to approximate the transformed density. Exponential tilting will be discussed more thoroughly in Chapter 5.

Daniels (1954) was one of the first to use saddlepoint methods to improve Gaussian approximations. Daniels' methods were derived using methods of integration through the complex plane. Later it was discovered that saddlepoint methods may be derived by combining exponential tilting with the Edgeworth expansion. Reid (1988), Daniels (1987) and many others discuss this approach. The idea behind combining exponential tilting and the Edgeworth expansion will be discussed again in Chapter 6.

The original application for saddlepoint methods was to improve the Gaussian approximation to the distribution of the sample mean. Since then saddlepoint methods have been used in many applications, and advances have been made in the methods used. Lugannani and Rice (1980) developed a method for improving Gaussian approximations to tail probabilities. Their approximation is much simpler in form then the methods developed by combining exponential tilting and the Edgeworth expansion. Skovgaard (1987) generalized the work of Lugannani and Rice so that it may be applied to conditional distributions. Wood, Booth, and Butler (1993) generalized the Lugannani and Rice approach so that other continuous distributions may be used in place of the Gaussian. There they use the Inverse Gaussian and the Chi-Square distributions. Easton and Ronchetti (1986) and Wang (1992) discuss methods for generalizing the saddlepoint methods so that the only data needed are the moments of the distribution being approximated.

Although the methods for improving Gaussian approximations have advanced in the past forty years, little has been done to improve Poisson approximations. Some progress has been made using continuity corrections with the saddlepoint methods, see Daniels (1987). Barbour and Jensen (1989) discuss the development of an expansion method that is based on the Poisson distribution as opposed to the Gaussian. They also discuss combining it with exponential tilting to obtain accurate point probabilities, and then computing all the necessary point probabilities to obtain tail probabilities. Barbour (1992) reviews this information. The focus of this work is to discuss settings for which the Poisson approximation is appropriate. An alternative derivation of the expansion will be presented in Chapter 4. Other improvements to the techniques will be presented in Chapter 6.

Chapter 3

Generating Functions

Generating functions serve many useful purposes in both statistics and mathematics. In statistics they are used primarily to represent distributions. They often lead to simple computations of probabilities, moments, or factorial moments. The purpose of this chapter is to review and develop the generating functions needed for the approximations to be presented in later chapters. Generating functions that will be omitted from this chapter include the moment generating function and the cumulant generating function. These generating functions are important to the improvement of Gaussian approximations, but do not serve a useful purpose for improving Poisson approximations.

3.1 Probability Generating Function

Perhaps one of the oldest and most popular generating functions is the probability generating function. The probability generating function may be presented in the following way:

$$\pi_x(q) = E(q^x) = \sum_x q^x f(x).$$

Many common distributions on the non-negative integers have probability generating functions that may be expressed in terms of elementary functions. For example, the probability generating function for the binomial distribution is $(1-p+pq)^n$. The probability generating function for the Poisson is important for the development of the techniques in later chapters. It is simply $e^{\lambda(q-1)}$. Note that the form is an exponential of a linear expression, which will be important later.

When the probability generating function is known, and is expressed in its simplest form, it may be expanded in a Taylor series about $q = 0$ to obtain the point probabilities. This leads to the following relationship for point probabilities:

$$f(x) = \pi_x^{(x)}(0)/x!$$

where $\pi_x^{(x)}(0)$ refers to the x^{th} derivative of the probability generating function evaluated at $q = 0$. This may be useful for cases where the probability generating function is a simple expression, but the density does not have a simple form. However, it may be tedious for large values of x, where higher derivatives are needed.

Another important feature of the probability generating function is that it may be used to obtain the mean for the distribution. It is easy to verify that $E(X) = \pi_x'(1)$. The probability generating function may also be used to generate other factorial moments by the following relationship:

$$E(X^{(k)}) = \pi_x^{(k)}(1)$$

where $X^{(k)} = \frac{X!}{(X-k)!}$. This suggests that factorial moments are usually easy to obtain for discrete distributions.

An important feature of probability generating functions involves the sum of random variables. Consider the following independent discrete random variables: $X_1, X_2, \cdots, X_n$. Let $Z = X_1 + X_2 + \cdots + X_n$. Then the probability generating function for Z is the following:

$$\pi_Z(q) = \pi_{X_1}(q)\pi_{X_2}(q)\cdots\pi_{X_n}(q).$$

An example where this is useful is for computing probabilities for the sum of independent Bernoulli random variables with different values for p. The probability generating function for a Bernoulli random variable is simply $(1 - p + pq)$. Using the relation given above for a sum, and then expanding in a Taylor series about the point $q = 0$ gives all of the point probabilities.

Probability generating functions are easy to obtain for many discrete random variables. They are often used for random variables in Engineering applications as well as many others. Some probability generating functions will be discussed, as needed, in the later chapters.

3.2 Factorial Moment Generating Function

As seen above, the probability generating function may be used to obtain factorial moments. For this reason, little attention has been given to the factorial moment generating function. It will be shown here that factorial moments may be used to obtain point probabilities.

Let $m^{(k)}$ represent the k^{th} factorial moment. Again let $\pi_x(t)$ represent the probability generating function. Now let $\psi_x(t)$ represent the factorial moment generating function. The factorial moment generating function takes the following form:

$$\psi_x(t) = 1 + m^{(1)}t + \frac{m^{(2)}t^2}{2!} + \frac{m^{(3)}t^3}{3!} + \cdots .$$

Now the probability generating function and the factorial moment generating function are related by $\psi_x(t-1) = \pi_x(t)$. Thus the probability generating function takes the following form:

$$\pi_x(t) = 1 + m^{(1)}(t-1) + \frac{m^{(2)}(t-1)^2}{2!} + \frac{m^{(3)}(t-1)^3}{3!} + \cdots .$$

Consider the following basic observations:

$$\begin{aligned}
(t-1) &= t-1 \\
(t-1)^2 &= t^2 - 2t + 1 \\
(t-1)^3 &= t^3 - 3t^2 + 3t - 1 \\
&\vdots \\
(t-1)^k &= \sum_{i=0}^{k} t^{k-i} \binom{k}{i} (-1)^i .
\end{aligned}$$

Matching the powers of t yields the following:

$$\begin{aligned}
P(X=y) &= \frac{m^{(y)}}{y!} - \frac{m^{(y+1)}\binom{y+1}{y}}{(y+1)!} + \frac{m^{(y+2)}\binom{y+2}{y}}{(y+2)!} + \cdots \\
&= \sum_{i=0}^{n^*} \frac{m^{(y+i)}(-1)^i}{y!i!} \qquad (3.1)
\end{aligned}$$

where n^* is the population size, or the maximum value for X, minus y. Now consider a trivial example to illustrate this computation. Compute the probability that $X = 3$ where X follows a binomial distribution with $n = 5$ and $p = .1$. For the binomial distribution $m^{(k)} = n^{(k)}p^k$. For this computation: $m^{(3)}$, $m^{(4)}$, and $m^{(5)}$ are needed. These values are .06, .012, and .0012 respectively. So the probability that X equals 3 is as follows:

$$P(X=3) = \frac{.06}{3!} - \frac{.012}{3!} + \frac{.0012}{3!2!} = .0081.$$

Notice that this is the exact probability, not an estimate.

Consider getting point probabilities for the hypergeometric distribution. This is an important example since it relates to contingency tables. The k^{th} factorial moment for a hypergeometric random variable may be written as:

$$\frac{W^{(k)}n^{(k)}}{(W+B)^{(k)}}.$$

So the point probability at y may be obtained by:

$$f(y) = \sum_{i=0}^{n^*} \frac{W^{(y+i)}n^{(y+i)}(-1)^i}{(W+B)^{(y+i)}y!i!}.$$

where n^* is the minimum of $W-y$ and $n-y$. If $W=10$, $B=10$, $n=8$, and $y=6$, then the point probability obtained by using this technique is .0750179. Again notice that this is the exact point probability. In this example $n^*=2$.

This method of obtaining point probabilities may be used repeatedly to obtain tail probabilities. However, this approach may become tedious if many point probabilities are needed.

3.3 Factorial Cumulant Generating Function

Another lesser known, but very important generating function is the factorial cumulant generating function. The importance of it will be demonstrated in later chapters. The factorial cumulant generating function, which will be denoted by $k_x(q)$, is simply the log of the probability generating function. This generating function should not be confused with the cumulant generating function, which is the log of the moment generating function.

Expanding the factorial cumulant generating function about the point $q=1$ leads to the following expression:

$$k_x(q) = (q-1)k^{(1)} + \frac{(q-1)^2k^{(2)}}{2!} + \frac{(q-1)^3k^{(3)}}{3!} + \frac{(q-1)^4k^{(4)}}{4!} + \cdots$$

where $k^{(i)}$ is the i^{th} factorial cumulant. Now noting the relationship between the factorial cumulant generating function and the probability generating function leads to the following relationship between the factorial cumulants and factorial moments:

$$k^{(1)} = m^{(1)}$$

$$
\begin{aligned}
k^{(2)} &= m^{(2)} - (m^{(1)})^2 \\
k^{(3)} &= m^{(3)} - 3m^{(1)}m^{(2)} + 2(m^{(1)})^3 \\
k^{(4)} &= m^{(4)} - 4m^{(1)}m^{(3)} + 12m^{(2)}(m^{(1)})^2 - 3(m^{(2)})^2 - 6(m^{(1)})^4 \\
&\vdots
\end{aligned}
$$

Thus the factorial cumulants may be obtained directly from the factorial moments.

The factorial cumulant generating function for the Poisson distribution is $\lambda(q-1)$. Note that it is linear in q, making all of the factorial cumulants beyond the first one equal to 0. Thus factorial cumulants are important in the study of Poisson approximations. One way to check whether or not a distribution is nearly Poisson is to examine the factorial cumulants. The fact that the factorial cumulant generating function for the Poisson is linear will be important in the development of the Poisson approximations to be presented in later chapters.

The factorial cumulant generating function for the binomial distribution is $n\ln(1-p+pq)$. The factorial cumulants are as follows:

$$
\begin{aligned}
k^{(1)} &= np \\
k^{(2)} &= -np^2 \\
k^{(3)} &= 2np^3 \\
&\vdots \\
k^{(j)} &= (-1)^{j-1}(j-1)!np^j.
\end{aligned}
$$

From this it may be seen that the Poisson approximation is better for small p since this will lead to small factorial cumulants. Notice that the second factorial cumulant is simply the variance minus the mean. For the binomial this is negative, thus revealing that the binomial distribution has shorter tails.

3.4 Finite Population Probability Generating Function

Finite population settings are common in many areas of statistics, especially in Categorical data analysis. Other problems under the heading of finite populations include Urn models and Occupancy problems. Unfortunately the probability generating function for most finite population distributions cannot be represented by simple or elementary functions. For example, the probability generating function for the hypergeometric distribution is a hypergeometric function. The probability generating function for a cell of a two-way contingency table is therefore a hypergeometric function. The purpose of this section is present a different probability generating function, which will be referred to as the Finite Population Probability Generating Function from now on, that may be used to represent distributions associated with finite populations.

3.4.1 Forming the Generating Function

The generating function presented here is a variation of the one from Terrell (Personal Communication, 1999). The sampling model being considered here is as follows: an urn has W white marbles and B black marbles that are indistinguishable in every way except color. A single trial will consist of removing a marble at random from those remaining in the urn, without replacing it. If n marbles are removed one at a time without replacement, then the number of white marbles selected follows a hypergeometric distribution. Thus the focus here is on finite populations where the trials are not necessarily independent.

Recall that the probability generating function may be written as $\pi_x(q) = E(q^x)$. For the hypergeometric distribution and the distribution for the cell of a contingency table, the parameters are related to the size of the population or a part of the population. For the hypergeometric described above, the parameters are $W+B$, W, and n. The form of the generating function being considered here will modify one of these parameters, say W, to a new value Q. The generating function is defined as:

$$\varphi_x(Q) = E\left(\frac{Q^{(W-X)}}{W^{(W-X)}}\right) \tag{3.2}$$

where

$$Q^{(W-X)} = \frac{Q!}{(Q-(W-X))!}.$$

Consider $\varphi_x(W-1)$:

$$\begin{aligned}\varphi_x(W-1) &= E\left(\frac{(W-1)^{(W-X)}}{W^{(W-X)}}\right) = E\left(\frac{(W-1)!X!}{(X-1)!W!}\right) \\ &= E\left(\frac{X}{W}\right) = \frac{E(X)}{W}.\end{aligned}$$

In general:

$$E(X^{(k)}) = W^{(k)} \varphi_x(W-k). \tag{3.3}$$

Probabilities can be generated from this in a manner similar to that of a probability generating function. By definition, φ_x can be written as:

$$\varphi_x(Q) = \sum_x \frac{Q^{(W-X)} f(X)}{W^{(W-X)}}.$$

Using Newton's forward difference formula, φ_x can also be expressed as:

$$\begin{aligned} \varphi_x(Q) &= \sum_x \frac{Q^{(X)} \triangle^X \varphi_x(0)}{X!} \\ &= \sum_x \frac{Q^{(W-X)} \triangle^{W-X} \varphi_x(0)}{(W-X)!} \end{aligned}$$

where $\triangle$ is being used to represent a forward difference. Newton's forward difference formula is presented briefly in section A.3. Matching the appropriate terms leads to:

$$f(x) = \begin{pmatrix} W \\ x \end{pmatrix} \triangle^{W-x} \varphi_x(0).$$

The generating function, $\varphi_x(Q)$, behaves very much like a probability generating function. However, it is designed to represent densities for finite populations in a simple form. It is also designed in such a way that it will be simple to use for tilting. Tilting is the topic of Chapter 5.

3.4.2 Example: Hypergeometric

The hypergeometric distribution will be considered here to illustrate the use of the generating function $\varphi_x(Q)$. Recall that the probability generating function for the hypergeometric distribution is a hypergeometric function, and that this function cannot be expressed in terms of elementary functions. The form of the probability density function being considered here is as follows:

$$f(x) = \frac{\begin{pmatrix} W \\ x \end{pmatrix} \begin{pmatrix} B \\ n-x \end{pmatrix}}{\begin{pmatrix} W+B \\ n \end{pmatrix}}.$$

The $\varphi_x(Q)$ generating function is derived as follows:

$$\begin{aligned}
\varphi_x(Q) &= E\left(\frac{Q^{(W-X)}}{W^{(W-X)}}\right) = \sum_x \left(\frac{Q^{(W-X)}}{W^{(W-X)}}\right) \frac{\binom{W}{X}\binom{B}{n-X}}{\binom{W+B}{n}} \\
&= \sum_x \left(\frac{Q!X!}{[Q-(W-X)]!W!}\right)\left(\frac{W!B!n!(W+B-n)!}{X!(W-X)!(n-X)!(B-n+X)!(W+B)!}\right) \\
&= \sum_x \frac{\binom{Q}{X-W+Q}\binom{B}{n-X}}{\binom{Q+B}{n-W+Q}} \left(\frac{(Q+B)!n!}{(W+B)!(n-W+Q)!}\right) \\
&= \frac{(Q+B)!n!}{(W+B)!(n-W+Q)!} \\
&= \frac{n^{(W-Q)}}{(W+B)^{(W-Q)}}.
\end{aligned}$$

The expected value of X may be computed as follows:

$$\begin{aligned}
E(X) &= W\varphi_x(W-1) = \frac{W(W-1+B)!n!}{(W+B)!(n-W+W-1)!} \\
&= \frac{Wn}{W+B}.
\end{aligned}$$

Thus this generating function gives a simple way of representing the hypergeometric distribution.

Chapter 4

Expansion Methods

4.1 Edgeworth Expansion

As mentioned, the Gaussian distribution plays a large role in asymptotic theory. However, Gaussian approximations often prove to be inadequate. In these situations statisticians tend to rely on methods for improving them. One such method is the Edgeworth expansion.

A derivation of the Edgeworth expansion is given in Hall(1992). The derivation involves an expansion of the characteristic function. Then an inverse expansion is used which leads to the Gaussian approximation plus some correction terms. The Edgeworth expansion is used to approximate the distribution of a statistic, which is usually expressed in the form of a sum. Thus it is usually expressed in terms of $n^{1/2}(\hat{\theta} - \theta)$, where n is the sample size or the number of terms in the sum. Using the notation of Hall(1992), the Edgeworth expansion may be expressed as:

$$P(n^{1/2}(\hat{\theta} - \theta)/\sigma \leq x) = \Phi(x) + n^{-1/2}\zeta_1(x)\phi(x) + \cdots + n^{-j/2}\zeta_j(x)\phi(x) + \cdots \tag{4.1}$$

where $\phi(x)$ is the standard normal density and $\Phi(x)$ is the cdf for the normal density. The ζ_j functions are polynomials with coefficients depending on the cumulants of $(\hat{\theta} - \theta)$. These polynomials often involve the Hermite polynomials. Bedrick and Hill (1992) and Barndorff-Nielsen and Cox (1979) also discuss the Edgeworth Expansion.

4.2 An Edgeworth-Like Expansion for Nearly-Poisson Probabilities

4.2.1 Derivation of the Expansion

A second expansion has been developed that is similar to the Edgeworth expansion, but is based on the Poisson distribution as opposed to the Gaussian distribution. The expansion is presented in Barbour (1992), but a different derivation and form will be presented here.

For the derivation, recall the following:

$$\begin{aligned}\pi_x(q) &= e^{k_x(q)} = e^{k^{(1)}(q-1)+k^{(2)}\frac{(q-1)^2}{2}+\cdots} \\ &= e^{k^{(1)}(q-1)}e^{k^{(2)}\frac{(q-1)^2}{2}+\cdots}.\end{aligned}$$

Notice that $e^{k^{(1)}(q-1)}$ is just the probability generating function for a Poisson random variable. The remaining exponential term can be expanded in another Taylor Series.

An important issue arises at this stage of the derivation. A rule is needed for the size of the cumulants. In the Edgeworth expansion, this is determined by working with the distribution of a sum of i.i.d random variables. The approach here will be to think of approximating binomial probabilities. The rule of thumb when approximating binomial probabilities with Poisson probabilities is that it is appropriate when p is small and np is moderate. This will be taken to mean $O(np) = 1$ or that $O(n) = p^{-1}$. Then $O(k^{(2)}) = p$, $O(k^{(3)}) = p^2$, and in general, $O(k^{(k)}) = p^{k-1}$. Thus the terms in the expansion will be grouped so that the terms of the same order in p are put together.

Notice that leaving out the terms from the expanded exponential leads to the usual Poisson approximation. Including correction terms involves multiplying the Poisson probability generating function by powers of $(q-1)$. For $(q-1)^k$, this will look like:

$$\sum_{y=0}^{\infty} q^y \sum_{j=0}^{k} (-1)^j \binom{k}{j} p(y-j).$$

Notice that the inner sum is the k^{th} backward difference (See equation (A.1)). Matching the coefficients with equal powers of q yields the following approximation:

$$f(x) = p(x) + \frac{k^{(2)}}{2}\nabla^2 p(x) + \left[-\frac{k^{(3)}}{6}\nabla^3 p(x) + \frac{(k^{(2)})^2}{8}\nabla^4 p(x)\right] + \cdots \tag{4.2}$$

where $f(x)$ is the true density, and $p(x)$ is the Poisson density with the mean taken to be the mean for $f(x)$. $k^{(i)}$ represents the i^{th} factorial cumulant, and $\nabla^j p(x)$ represents the

j^{th} backward difference of $p(x)$. See section A.1 for information regarding the backward difference formula.

In statistics, tail probabilities are usually more important than point probabilities. Tail probabilities can easily be obtained by summing both sides of equation(4.2). Using the information in section A.1 regarding the sum of a backward difference leads to the following approximations:

$$P(X \leq y) = R(y) + \frac{k^{(2)}}{2}\nabla^1 p(y) + \left[-\frac{k^{(3)}}{6}\nabla^2 p(y) + \frac{(k^{(2)})^2}{8}\nabla^3 p(y)\right] + \cdots \quad (4.3)$$

$$P(X \geq y) = R^*(y) - \frac{k^{(2)}}{2}\nabla^1 p(y-1) + \left[\frac{k^{(3)}}{6}\nabla^2 p(y-1) - \frac{(k^{(2)})^2}{8}\nabla^3 p(y-1)\right] + \cdots \quad (4.4)$$

where $R(y) = \sum_{x=0}^{y} p(x)$, and $R^*(y) = \sum_{x=y}^{\infty} p(x)$.

These expansion methods work best when computing probabilities near the mean. Caution should be used when using these methods in the far tails of a distribution. It is possible to obtain negative probability estimates for points far enough out in the tails. This will be illustrated in an example in the next section.

4.2.2 Example: Binomial Distribution

The binomial distribution will be used here to illustrate the methods. More examples and applications of the methods presented above will occur later. The binomial distribution is being presented here for several reasons. One is to show that the methods work at least for a simple case. Another is that the Poisson approximation to the binomial is discussed in many textbooks. So it is interesting to see if the approximation can be improved.

In order to use the expansion, the factorial cumulants are needed. For the binomial distribution the factorial cumulants are $k^{(j)} = (-1)^{j-1}(j-1)!np^j$. Table 4.1 shows the results for approximating the point probabilities of a binomial distribution with $n = 50$ and $p = .1$. This table and the following tables present the approximations with the relative errors in parenthesis. The relative errors are computed by subtracting the approximation from the exact probability, and then dividing by the exact probability. In this table Expand1 and Expand2 refer to the expansion using one and two correction terms respectively. In this example $k^{(1)} = 5$, $k^{(2)} = -.5$, and $k^{(3)} = .1$.

From the table it is easy to see that the approximations due to the expansion are better than the usual Poisson approximation. Although in this example the approximations resulting from the expansion are accurate in the tails, the accuracy is still far better near the mean. To illustrate the problem with negative probability estimates, consider approximating the probability that y equals 16. The approximation using the expansion with one correction term would be -.000002.

Table 4.1: Approximating the Binomial Distribution with $n = 50$ and $p = .1$ using the Expansion Method

y	Binomial	Poisson	Expand1	Expand2
2	.077943	.084224 (-.08)	.078329 (-.005)	.077957 (-.0002)
4	.180903	.175467 (.03)	.180731 (.001)	.180904 $(-5.3 * 10^{-6})$
5	.184923	.175467 (.05)	.184240 (.004)	.184869 (.0003)
6	.154102	.146223 (.05)	.153534 (.004)	.154058 (.0003)
8	.064277	.065278 (-.02)	.064625 (-.005)	.064324 (-.0007)
10	.015183	.018133 (-.19)	.015413 (-.02)	.015188 (-.0003)
12	.002215	.003434 (-.55)	.002164 (.02)	.002200 (.007)

As mentioned throughout, interest is usually in obtaining tail probabilities. Table 4.2 displays the results for right tail approximations using the binomial $n = 50$ and $p = .1$. The notation for this table is the same as that used in the previous table.

From table 4.2, it becomes clear that the expansion methods continue to improve the Poisson approximations for tail probabilities. It also becomes clear that the accuracy of the expansion methods diminishes, along with the Poisson approximation, for points in the far tails. Again the expansion method may lead to negative probability estimates.

Table 4.3 displays the results for left tail probability approximations. A binomial with $n = 50$ and $p = .4$ is used. This table shows the same trends previously seen. Here the expansion methods perform very well near the mean with the accuracy diminishing in the tails. In this example the accuracy of the expansion method is much less favorable compared to the accuracy seen in the previous tables. The reason for this is that the conditions are less favorable for the Poisson approximation. Here p is no longer small. Thus it appears that the accuracy of the expansion method diminishes in the tails at a faster rate as the conditions become less favorable for the Poisson approximation.

In the previous two examples it has been shown that the expansion method with one correction term can give unreasonable estimates. However, in these cases the expansion method with two correction terms has given at least positive estimates. To show that the expansion method with two correction terms can also give negative estimates, consider a binomial with $n = 50$ and $p = .5$. The right tail approximation for $y = 35$ using the expansion methods

Table 4.2: Approximating the Right Tail of a Binomial Distribution with $n = 50$ and $p = .1$ using the Expansion Method

y	Binomial	Poisson	Expand1	Expand2
6	.383877	.384039 (-.0004)	.384039 (-.0004)	.383893 (-.00004)
8	.122145	.133372 (-.09)	.122927 (-.006)	.122196 (-.0004)
10	.024538	.031828 (-.30)	.024575 (-.0015)	.024509 (.0012)
12	.003220	.005453 (-.69)	.002980 (.075)	.003200 (.006)
14	.000285	.000698 (-1.45)	.000170 (.40)	.000289 (-.014)
16	.000017	.000069 (-3.06)	$-9.61 * 10^{-6}$ (1.57)	.000022 (-.29)
18	$7.60 * 10^{-7}$	$5.42 * 10^{-6}$ (-6.13)	$-3.26 * 10^{-6}$ (5.29)	$2.12 * 10^{-6}$ (-1.79)

with one and two correction terms are -.002018 and -.001932 respectively. The true probability is .003302. In section 4.2.1, it was shown that the order of the correction terms depends on p. Thus as p increases or approaches one, the accuracy of the expansion methods will also start to diminish. This was also seen in the previous examples. One possible method for obtaining better approximations is to add more correction terms.

4.2.3 Comments: Edgeworth-Like Expansion

The Edgeworth-Like expansion has been shown to lead to an improvement over the usual Poisson approximations. The idea behind the use of the expansion is to simply add back some of the error incurred while using the Poisson approximation. The hope is that the portion of the error added back is the larger portion. Another way to view the use of the expansion is to recall that the factorial cumulants, beyond the first one, are zero for the Poisson distribution. Thus using the expansion is accounting for the fact that the factorial cumulants may not all be zero for the distribution being approximated.

Perhaps the most appealing feature of the expansion is it's simplicity. All that is required to perform the approximation are the factorial moments. For most discrete distributions, the factorial moments are easy to obtain, even if the density is complicated.

The only major disadvantage is that the accuracy of the approximation diminishes in the tails of the distribution. It may also lead to negative probability estimates. This presents a

Table 4.3: Approximating the Left Tail of a Binomial Distribution with $n = 50$ and $p = .4$ using the Expansion Method

y	Binomial	Poisson	Expand1	Expand2
18	.335613	.381422 (-.13)	.347665 (-.04)	.339338 (-.01)
16	.156091	.221074 (-.42)	.169425 (-.09)	.159096 (-.02)
14	.053955	.104864 (-.94)	.058380 (-.08)	.053629 (.006)
12	.013251	.039012 (-1.94)	.010812 (.18)	.011893 (.10)
10	.002197	.010812 (-3.92)	-.000821 (1.37)	.001738 (.21)
8	.000231	.002087 (-8.03)	-.001054 (5.56)	.000335 (-.45)
6	.000014	.000255 (-17.21)	-.000258 (19.42)	.000104 (-6.43)

serious problem since cut-off values for p-values are usually in the tails of the distribution.

4.3 Density-Based Tail Probability Approximation

4.3.1 Derivation

The methods obtained above adjust Poisson approximations using information involving factorial cumulants. An alternative method of deriving an approximation to a tail probability is to approximate the density rather than the probability generating function. Here the focus will be on the upper tail. The derivation will only be outlined here. See section B.2 in the appendix for detailed derivations for the upper and lower tails.

The idea here is to write $f(x)$ as a product of $p(x)$ and $g(x)$, where $p(x)$ is the Poisson density and $g(x)$ is a function that is nearly constant. So approximating $\sum_{x=y}^{\infty} f(x)$ will involve working with $\sum_{x=y}^{\infty} p(x)g(x)$. To prevent $g(x)$ from being negative, $g(x)$ will be set equal to $e^{h(x)}$. Thus $f(x)$ is being set equal to $p(x)e^{h(x)}$.

Solving for $h(x)$ yields $h(x) = \ln(f(x)) - \ln(p(x))$. So $\nabla h(x) = \nabla \ln(f(x)) - \nabla \ln(p(x))$. Now -$\nabla \ln(p(x)) = \ln(x) - \ln(\lambda)$. This yields the following formula for the backward difference of $h(x)$:

$$\nabla h(x) = \nabla \ln(f(x)) + \ln(x) - \ln(\lambda). \tag{4.5}$$

In general, the formula for the k^{th} backward difference of $h(x)$ is as follows:

$$\nabla^k h(x) = \nabla^k \ln(f(x)) - \nabla^{k-1} \ln(x). \tag{4.6}$$

Later it will be beneficial to have $\nabla h(y)$ equal to zero. This can be accomplished by setting equation (4.5) equal to zero and solving for λ. This leads to:

$$\lambda = \frac{yf(y)}{f(y-1)}.$$

This is a case where it is useful to take the mean of the Poisson to be something other than the mean associated with $f(x)$.

Now expand $h(x)$ about y using Newton's backward difference formula (See Appendix A section A.3). Newton's backwards difference formula expands the function using the backward difference of the function as opposed to the derivatives. The resulting form is the following:

$$f(x) = p(x) \exp\left[h(y) + (x-y)\nabla h(y) + \binom{x-y+1}{2} \nabla^2 h(y) + \cdots \right].$$

Recall that λ can be chosen to make $\nabla h(y)$ equal to zero. So this can be simplified to the following expression.

$$f(x) = p(x)g(y) \exp\left[\binom{x-y+1}{2} \nabla^2 h(y) + \binom{x-y+2}{3} \nabla^3 h(y) + \cdots \right].$$

In order to make it easier to take summations of both sides, the exponential can be expanded in the usual power series.

After expanding with the power series, the issue of how to group the terms arises again. That is, what is the relative size of each term. As with the Edgeworth-like expansion method, this will be addressed by considering the approximation to the binomial distribution. The details for this technique are addressed in section B.2.3.

The next issue deals with summing both sides of the expression in order to obtain approximations to the tail probabilities. Thus the following expression needs to be computed:

$$\sum_{x=y}^{\infty} p(x)g(y)\left(\left[1+\binom{x-y+1}{2}\nabla^2 h(y)+\binom{x-y+2}{3}\nabla^3 h(y)+\cdots\right]\right.$$
$$\left.+1/2\left[\binom{x-y+1}{2}^2(\nabla^2 h(y))^2+\cdots\right]+\cdots\right).$$

This will be accomplished by summing term by term and using summation by parts. Let $L_k^*(y)=\sum_{x=y}^{\infty}(x-y)^k p(x)$. Also let $z=y-\lambda$. Then the first few L^* functions are as follows:

$$\begin{aligned}
L_1^*(y) &= \lambda p(y-1) - zR_\lambda^*(y)\\
L_2^*(y) &= \lambda R_\lambda^*(y) - zL_1^*(y)\\
L_3^*(y) &= \lambda R_\lambda^*(y) + 2\lambda L_1^*(y) - zL_2^*(y)\\
L_4^*(y) &= \lambda R_\lambda^*(y) + 3\lambda L_1^*(y) + 3\lambda L_2^*(y) - zL_3^*(y).
\end{aligned} \tag{4.7}$$

Notice that the L^* functions may be obtained recursively using the previous L^* functions and the binomial coefficients. Using the L^* functions and the fact that $g(y)=f(y)/p(y)$ leads to the following approximation:

$$\begin{aligned}
P(X\geq y) &= (f(y)/p(y))[R_\lambda^*(y)+(1/2)(\nabla^2\ln(f(y))+\nabla\ln(y))(L_2^*(y)+L_1^*(y))\\
&+ [(1/6)(\nabla^3\ln(f(y))+\nabla^2\ln(y))(L_3^*(y)+3L_2^*(y)+2L_1^*(y))\\
&+ (1/8)(\nabla^2\ln(f(y))+\nabla\ln(y))^2(L_4^*(y)+2L_3^*(y)+L_2^*(y))]\\
&+ \cdots].
\end{aligned} \tag{4.8}$$

Approximating the left tail probabilities only requires adjustments to the L^* functions. Let $L_k=\sum_{x=0}^{y}(x-y)^k p(x)$. Again let $z=y-\lambda$. The first few L functions are as follows:

$$\begin{aligned}
L_1(y) &= -\lambda p(y) - zR_\lambda(y)\\
L_1(y-1) &= -\lambda p(y-1) - zR_\lambda(y-1)\\
L_2(y) &= \lambda R_\lambda(y-1) - zL_1(y)\\
L_2(y-1) &= \lambda R_\lambda(y-1) - zL_1(y-1)\\
L_3(y) &= \lambda R_\lambda(y-1) + 2\lambda L_1(y-1) - zL_2(y)\\
L_4(y) &= \lambda R_\lambda(y-1) + 3\lambda L_1(y-1) + 3\lambda L_2(y-1) - zL_3(y).
\end{aligned} \tag{4.9}$$

Notice that these L functions may also be derived recursively in a way similar to the L^* functions above. The only adjustments needed to obtain approximations to the left tail is to replace the L^* functions with the respective L functions, and replace R^*_λ with R_λ. The approximation is as follows:

$$\begin{aligned} P(X \leq y) &= (f(y)/p(y))[R_\lambda(y) + (1/2)(\nabla^2 \ln(f(y)) + \nabla \ln(y))(L_2(y) + L_1(y)) \\ &+ [(1/6)(\nabla^3 \ln(f(y)) + \nabla^2 \ln(y))(L_3(y) + 3L_2(y) + 2L_1(y)) \\ &+ (1/8)(\nabla^2 \ln(f(y)) + \nabla \ln(y))^2(L_4(y) + 2L_3(y) + L_2(y))] \\ &+ \cdots]. \end{aligned} \tag{4.10}$$

4.3.2 Example: Binomial Distribution

The binomial distribution is examined in the appendix in section B.2.3,where it was used to determine the relative size of the correction terms. Here the binomial distribution will be used to assess how well the density based approach works. The results from the density based approach will be compared to the results using the Edgeworth-like expansion method.

First consider the example from section 4.2.2 for right tail approximations. There the methods were used to approximate the tail probabilities for a binomial with $n = 50$ and $p = .1$. Table 4.4 displays the results for the techniques of section 4.2.2 as well as the density based approximations. Den0, Den1, and Den2 will represent the density based approximation with no correction terms, one correction term, and two correction terms respectively. Recall that λ was chosen to make what would have been the first correction term equal to zero. So no correction terms refers to using the approximation in equation (4.8) up to the $R^*_\lambda(y)$ term. No additional correction terms are included for the approximations labeled Den0 in the table.

Table 4.4 reveals that the density based methods are very accurate in the tails of the distribution. The density based method with no extra correction terms was reasonably accurate. This suggests that given two point probabilities, $f(y)$ and $f(y-1)$, it is possible to obtain a reasonable estimate for the tail probabilities. Given a little extra information, such as the recursive formula for obtaining other probabilities, or one or two more point probabilities, then the approximations become very accurate. These methods become more accurate for points further out in the tails. Recall that the basic idea behind this method is to match the Poisson density to the density being approximated. This is more feasible in the tails of the density.

Table 4.5 compares the methods for approximating the left tail of a binomial distribution. As in section 4.2, the binomial being approximated will be the binomial with $n = 50$ and $p = .4$. The same trends appear to occur in the left tail as in the right. The density based methods are very accurate in the tails. Their accuracy increases as the point moves further out in the tails.

In section 4.2 it became clear that the accuracy of the expansion methods would start to diminish slightly as p increased. To determine the reaction of the density based approaches

Table 4.4: Comparing Density Based Methods and Expansion Methods for the Right Tail of a Binomial Distribution with $n = 50$ and $p = .1$

y	Bin.	Poi.	Exp1	Exp2	Den0	Den1	Den2
6	.383877	.384039 (-.0004)	.384039 (-.0004)	.383893 (-.00004)	.404738 (-.05)	.382499 (.004)	.384002 (-.0003)
7	.229773	.237817 (-.04)	.230505 (-.003)	.229835 (-.0003)	.239320 (-.04)	.229272 (.002)	.229814 (-.0002)
8	.122145	.133372 (-.09)	.122927 (-.006)	.122196 (-.0004)	.126153 (-.03)	.121977 (.001)	.122158 (-.0001)
9	.057867	.068094 (-.18)	.058302 (-.008)	.057872 (-.00009)	.059409 (-.03)	.057815 (.0009)	.057871 (-.00007)
10	.024538	.031828 (-.30)	.024575 (-.002)	.024509 (.001)	.025082 (-.02)	.024523 (.0006)	.024539 (-.00004)
11	.009355	.013695 (-.46)	.009162 (.02)	.009321 (.004)	.009531 (-.02)	.009351 (.0004)	.009355 (0.00)
12	.003220	.005453 (-.69)	.002980 (.07)	.003200 (.006)	.003272 (-.02)	.003219 (.0003)	.003220 (0.00)
14	.000285	.000698 (-1.45)	.000170 (.40)	.000289 (-.014)	.000289 (-.01)	.000285 (0.00)	.000285 (0.00)
16	.000017	.000069 (-3.06)	$-9.61 * 10^{-6}$ (1.57)	.000022 (-.29)	.000017 (0.00)	.000017 (0.00)	.000017 (0.00)

to such a situation, consider approximating the right tail of a binomial distribution with $n = 60$, $p = .8$, and $y = 55$. The true probability and the Poisson approximation are .012106 and .173167 respectively. The approximations based on the expansion methods are .082430 and .038573. Thus the accuracy of the expansion methods has diminished. The approximations using the density based methods are .014173, .011523, and .012495 for no correction terms, one correction term, and two correction terms respectively. Thus it appears that the density based approaches are not affected by increasing p in the same way as the expansion methods.

4.3.3 Example: Hypergeometric Distribution

As another illustration, consider approximating the hypergeometric distribution. This example is important because it is a simple example of a distribution that is associated with a finite population. The form of the hypergeometric being considered here is as follows:

Table 4.5: Comparing Density Based Methods and Expansion Methods for the Left Tail of a Binomial Distribution with $n = 50$ and $p = .4$

y	Bin.	Poi.	Exp1	Exp2	Den0	Den1	Den2
18	.335613	.381422 (-.14)	.347665 (-.04)	.339338 (-.01)	.361504 (-.08)	.329415 (.02)	.337462 (-.006)
17	.236876	.297028 (-.25)	.251456 (-.06)	.240924 (-.02)	.249649 (-.05)	.234413 (.01)	.237484 (-.002)
16	.156091	.221074 (-.42)	.169425 (-.09)	.159096 (-.02)	.162011 (-.04)	.155165 (.006)	.156281 (-.001)
15	.095502	.156513 (-.64)	.104864 (-.10)	.096773 (-.013)	.098061 (-.03)	.095176 (.003)	.095558 (-.0006)
14	.053955	.104864 (-.94)	.058380 (-.08)	.053629 (.006)	.054979 (-.02)	.053848 (.002)	.053970 (-.0003)
13	.027988	.066128 (-1.36)	.028166 (-.006)	.026756 (.04)	.028364 (-.01)	.027956 (.001)	.027992 (-.0001)
12	.013251	.039012 (-1.94)	.010812 (.18)	.011893 (.10)	.013376 (-.009)	.013242 (.0007)	.013251 (0.00)
10	.002197	.010812 (-3.92)	-.00082 (1.37)	.001738 (.21)	.002207 (-.005)	.002197 (0.00)	.002197 (0.00)
8	.000231	.002087 (-8.03)	-.00105 (5.55)	.000335 (-.45)	.000231 (0.00)	.000231 (0.00)	.000231 (0.00)

$$f(x) = \frac{\binom{W}{x}\binom{B}{n-x}}{\binom{W+B}{n}}.$$

To use the Edgeworth-like approximation, the factorial moments are all needed. For the hypergeometric distribution described above, the factorial moments are as follows:

$$E(X^{(k)}) = \frac{W^{(k)}n^{(k)}}{(W+B)^{(k)}}.$$

The factorial moments are then used to obtain the factorial cumulants. Then the approximation is simply obtained from equation (4.4) and (4.3).

To obtain the density based approximation, $\nabla \ln(f(x))$ needs to be computed. For the hypergeometric distribution, this simplifies to the following:

$$\nabla \ln(f(x)) = \ln\left(\frac{(W-x+1)(n-x+1)}{x(B-n+x)}\right).$$

So the pieces needed for the approximation may be computed in the following way:

$$\begin{aligned}
\nabla^2 \ln(f(y)) + \nabla \ln(y) &= \nabla \ln\left(\frac{(W-y+1)(n-y+1)}{(B-n+y)}\right) \\
\nabla^3 \ln(f(y)) + \nabla^2 ln(y) &= \nabla^2 \ln\left(\frac{(W-y+1)(n-y+1)}{(B-n+y)}\right) \\
&\vdots
\end{aligned}$$

Once these are computed, the method proceeds as in the binomial examples.

Table 4.6 displays the results for the right tail approximation to the hypergeometric distribution with $W = 40$, $n = 25$, and $B = 60$. Note that the mean for this distribution is 10.

Table 4.6: Comparing Density Based Methods and Expansion Methods for Approximating the Right Tail of a Hypergeometric Distribution with $W = 40$, $n = 25$, and $B = 60$

y	Hyp.	Poi.	Exp1	Exp2	Den0	Den1	Den2
11	.404086	.416960 (-.03)	.41696 (-.03)	.41203 (-.02)	.580647 (-.44)	.256220 (.37)	.561096 (-.38)
13	.119710	.208444 (-.74)	.156745 (-.31)	.134018 (-.12)	.141853 (-.18)	.110413 (.08)	.125378 (-.05)
14	.050277	.135536 (-1.70)	.075884 (-.51)	.055212 (-.10)	.056840 (-.13)	.048241 (.04)	.051256 (-.02)
15	.017484	.083459 (-3.77)	.026647 (-.52)	.013113 (.25)	.019158 (-.10)	.017091 (.02)	.017638 (-.008)
16	.004978	.048740 (-8.79)	.001398 (.72)	-.00369 (1.74)	.005340 (-.07)	.004913 (.01)	.005000 (-.004)
17	.001147	.027042 (-22.58)	-.00847 (8.38)	-.00675 (6.88)	.001212 (-.06)	.001138 (.008)	.001150 (-.003)
18	.000211	.014278 (-66.70)	-.01009 (48.82)	-.00449 (22.28)	.000220 (-.04)	.000210 (.005)	.000211 (0.00)
19	.000030	.007187 (-238.6)	-.00829 (277.3)	-.00151 (51.33)	.000032 (-.06)	.000030 (0.00)	.000030 (0.00)

The example in Table 4.6 reveals that the density based methods are at their best in the tails. The first row of the table shows that these methods can have problems at points very close to the mean. This example also reveals that the Poisson approximation may lead to incorrect upper 5%, 10%, and 1% values. These values are important to statisticians carrying out a test of a hypothesis. Thus the improvement over the Poisson approximation in the tails of the distribution is crucial. The Edgeworth-like expansion performs well near the mean of the distribution, but quickly breaks down for points in the tails of the distribution.

Table 4.7 presents the left tail approximations for the same distribution used in table 4.6. The results from this table once again reveal the accuracy of the density based approach in the tails of the distribution. However, in this example, the approximation near the mean is at least reasonable. In fact, the density based approximations at $y = 9$ are more accurate than the Poisson approximation. Thus it appears that the density based approach is more accurate when estimating the left tail probabilities of a distribution, especially near the mean. The Edgeworth-like approximations show the same trends seen previously. The methods are accurate near the mean, but break down for points in the tails.

Table 4.7: Comparing the Density Based Methods and Expansion Methods for Approximating the Left Tail of a Hypergeometric Distribution with $W = 40$, $n = 25$, and $B = 60$

y	Hyp.	Poi.	Exp1	Exp2	Den0	Den1	Den2
9	.409988	.457930 (-.12)	.423809 (-.034)	.414725 (-.012)	.450240 (-.10)	.395386 (.04)	.416150 (-.02)
8	.241306	.332820 (-.38)	.271402 (-.125)	.253077 (-.05)	.253305 (-.05)	.238271 (.01)	.242224 (-.004)
7	.118515	.220221 (-.86)	.146520 (-.24)	.127719 (-.078)	.121508 (-.03)	.117973 (.005)	.118633 (-.001)
6	.047404	.130141 (-1.75)	.061354 (-.29)	.049488 (-.044)	.048002 (-.01)	.047324 (.002)	.047416 (-.0003)
5	.015040	.067086 (-3.46)	.015495 (-.03)	.012513 (.17)	.015131 (-.006)	.015031 (.0006)	.015041 (-.00007)
4	.003667	.029253 (-6.98)	-.0017 (1.46)	.00098 (.73)	.003677 (-.003)	.003666 (.0003)	.003667 (0.00)
3	.000658	.010336 (-14.71)	-.0041 (7.24)	-.00036 (1.55)	.000659 (-.002)	.000658 (0.00)	.000658 (0.00)

4.3.4 Comments: Density Based Method

The density based approach has shown remarkable accuracy in the examples presented. Perhaps the strongest advantage of this method is its simplicity, especially since the correction terms may be computed recursively from the previous terms. This makes it simple to add additional correction terms. The backwards differences of the log of the density may sometimes be computed from a recursive relationship for successive point probabilities. This was done for the binomial and hypergeometric example. If such a relationship does not exist, then it may be computed from using a few point probabilities. Usually only two to four point probabilities will be required to obtain accurate results.

Another advantage of the density based approach is that the basic idea leading to it is simple to understand. This approach is basically matching the density of a Poisson random variable to that of the density being approximated. This approach also allows for the flexibility of choosing λ in such a way to make the approximation easier to compute and more accurate. This is an idea that does not require a strong statistical background to understand. Most of the mathematical tools involved in the derivations are related to ones taught in a Calculus course.

The major disadvantage of the density based approach is that it does not perform well near the mean. However, most statistical applications require accurate approximations to tail probabilities. For example, in hypothesis testing, the null hypothesis is rejected if the test statistic is far enough out in the tails. At least in the examples considered up to this point, the density based approach has always lead to the correct .05, .1 , and .01 quantiles.

4.4 Discussion

Two different expansion methods have been presented in this chapter. The first expansion is based on approximating the probability generating function. The result was an expansion that led to accurate results near the mean of the distribution. Unfortunately the results were not usually accurate in the tails of the distribution. The point at which the expansion breaks down is related to the appropriateness of the Poisson approximation.

The second expansion is based on approximating the density. The result was an expansion that led to accurate results in the tails of the distribution. The density based expansion had it's problems for points near the mean of the distribution. Thus the two expansions have their strengths in totally different parts of the distribution. One advantage of the density based expansion is that it appears to work in some cases where the Poisson approximation may seem inappropriate.

Although these expansions seem to be very different, they are derived using similar techniques. The Edgeworth-like expansion was derived by first rewriting the probability generating function as an exponential of the factorial cumulant generating function. Then Taylor series expansions were used, followed by grouping the terms. The density based expansion was derived by first rewriting the density as a product of the Poisson density and an exponential term. The term in the exponential was expanded using Newton's Backward difference formula followed by a Taylor series expansion of the exponential. This was followed by grouping the terms. Thus the ideas behind the approximations are very similar. The main difference comes from the term being approximated. In the first case it was the probability generating function. In the second case it was the actual density. Also the Edgeworth-like expansion involves expanding around $q = 1$, which is associated with the center or mean of the distribution, as shall be seen in the next chapter. The density based expansion involves expanding around y, the point of interest.

Chapter 5

Tilting

The usual Poisson approximation and the Edgeworth-like expansion tend to work better for points near the mean of the density. These methods tend to lose their accuracy for points in the tails of the distribution. The tails of a distribution are important since they often relate to p-values. The density based expansion is one approach to obtaining accurate probability estimates for points in the tails of a distribution. Unfortunately it may have problems for points near the mean. Thus a method for obtaining accurate probabilities in the tails, while still yielding accurate estimates near the mean is needed.

The method to be explored in this chapter is tilting. As with the Edgeworth-like expansion, the main mathematical tool used here will be the generating functions. However the basic idea is similar to that of the density based expansion in that the focus is on the point of interest, not the center or mean of the distribution.

5.1 Exponential Tilting

5.1.1 Conjugate Density Approach

The idea behind exponential tilting appears to have originated with Esscher(1932), see Reid(1988). Since then it has been discussed in many sources including Barndorff-Nielsen and Cox (1979), Daniels (1987), and Reid (1988). Terrell (1999, Ch.13) gives a thorough discussion of exponential tilting, including its use in discrete cases. The basic idea behind exponential tilting is to transform the density so that the mean of the transformed density is the point of interest. If the distribution is a member of the exponential family, then the transformation is performed in such a way that the transformed density is from the same family as the original density. The transformed density is often called the conjugate density. The conjugate density takes the following form:

$$f_t(x) = \frac{e^{-xt} f(x)}{m_x(t)} \tag{5.1}$$

where again $f(x)$ is the density we are trying to approximate and $m_x(t)$ is the moment generating function. For discrete cases it may be more convenient to express this in the following way:

$$f_q(x) = \frac{q^x f(x)}{\pi_x(q)} \tag{5.2}$$

where $\pi_x(q)$ is the probability generating function. Since the emphasis here is on discrete distributions, the main focus will be on equation (5.2).

Let x_q be a random variable from the distribution f_q. Then Terrell (1999) derives the following p.g.f. for x_q:

$$\pi_{x_q}(r) = \frac{\pi_x(rq)}{\pi_x(q)}.$$

Now let $k_x(q) = \ln(\pi_x(q))$ represent the factorial cumulant generating function. Then the mean of x_q is $E(x_q) = qk'_x(q)$.

So to tilt the mean to the point of interest, say y, solve for q in the equation $y = qk'_x(q)$. Then approximate the conjugate density by some known distribution, usually the normal or Poisson distribution. The approximation to the transformed density should be fairly accurate since the approximation is at a point that is equivalent to the mean. This leads to:

$$\frac{q^y f(y)}{\pi_x(q)} = p_y(y) \tag{5.3}$$

where $p_y(y)$ is the Poisson density with mean y. Jin and Robinson(1999) discuss performing tilting, then using different distributions to approximate the transformed density. This work will focus exclusively on using the Poisson distribution. To obtain an approximation for $f(y)$, simply solve for it in equation (5.3). The resulting approximation is:

$$f(y) = \pi_x(q)(1/q)^y p_y(y). \tag{5.4}$$

After a little algebra, the approximation becomes:

$$f(y) = \pi_x(q)e^{-y(1-\frac{1}{q})} p_{y/q}(y). \tag{5.5}$$

As mentioned previously, statisticians are often interested in tail probabilities. Using the notation from equations (4.3) and (4.4), the tail probability approximations are obtained by summing over the necessary regions, and are as follows:

$$P(X \leq y) = \pi_x(q)e^{-y(1-\frac{1}{q})}R_{y/q}(y) \tag{5.6}$$

$$P(X \geq y) = \pi_x(q)e^{-y(1-\frac{1}{q})}R^*_{y/q}(y). \tag{5.7}$$

Exponential tilting works well for values that are in the tails of the distribution. Very little is gained by using these methods close to the mean. These methods reduce to the traditional Poisson approximations when the point of interest is the mean of the distribution.

5.1.2 Another View of Exponential Tilting

The previous section discussed exponential tilting from the view of forming a conjugate density. In this section, exponential tilting will be formed by means of an expansion. For this discussion, let $\hat{q}$ denote the solution to the expression $y = qk_x'(q)$.

Recall from chapter 3 that the factorial cumulant generating function for the Poisson distribution is linear. Thus when considering Poisson approximations, one possibility is to approximate the factorial cumulant generating function for the distribution of interest by a linear function. Note that expanding the factorial cumulant generating function with a first order Taylor series about the point $q = 1$ leads to the following:

$$\begin{aligned} k_x(q) &\approx k_x(1) + (q-1)k_x'(1) \\ &= (q-1)k^{(1)}. \end{aligned}$$

This suggests using a Poisson approximation with mean $k^{(1)}$, which is just the usual Poisson approximation.

Instead of expanding about the point $q = 1$, consider expanding about the point $q = \hat{q}$. This leads to the following:

$$\begin{aligned} k_x(q) &\approx k_x(\hat{q}) + (q-\hat{q})k_x'(\hat{q}) \\ &= k_x(\hat{q}) + qk_x'(\hat{q}) - \hat{q}k_x'(\hat{q}) \\ &= k_x(\hat{q}) - y + qk_x'(\hat{q}) + k_x'(\hat{q}) - k_x'(\hat{q}) \\ &= k_x(\hat{q}) - y + k_x'(\hat{q}) + k_x'(\hat{q})(q-1). \end{aligned}$$

Notice that taking the exponential of both sides leads to the following:

$$\begin{aligned}\pi_x(q) &\approx Exp[k_x(\hat{q}) - y + k_x'(\hat{q})](Exp[(q-1)k_x'(\hat{q})]) \\ &= \pi_x(\hat{q})e^{-y(1-\frac{1}{\hat{q}})}(e^{(q-1)k_x'(\hat{q})}).\end{aligned}$$

Now observe that $e^{(q-1)k_x'(\hat{q})}$ is just the probability generating function for a Poisson random variable with mean $k_x'(\hat{q}) = y/\hat{q}$. This suggests the following approximations:

$$\begin{aligned}f(y) &= \pi_x(\hat{q})e^{-y(1-\frac{1}{\hat{q}})}p_{y/\hat{q}}(y) \\ P(X \leq y) &= \pi_x(\hat{q})e^{-y(1-\frac{1}{\hat{q}})}R_{y/\hat{q}}(y) \\ P(X \geq y) &= \pi_x(\hat{q})e^{-y(1-\frac{1}{\hat{q}})}R^*_{y/\hat{q}}(y).\end{aligned}$$

Note that this is the same approximation obtained from exponential tilting. Thus exponential tilting may be thought of as expanding the factorial cumulant generating function about a value of q that is associated with the point of interest, instead of about the point $q = 1$, which is associated with the mean.

5.1.3 Example: Binomial

A starting place for using the approximation methods is to determine what q should be for the exponential tilting. Recall that this is done by solving for q in the equation $y = qk_x'(q)$. For the binomial distribution it turns out that:

$$q = \frac{(1-p)y}{p(n-y)}.$$

Recall that the probability generating function for the binomial is $(1-p+pq)^n$. The factorial cumulant generating function is just the log of this.

Table 5.1 displays the results for approximating point probabilities from a binomial with $n = 50$ and $p = .1$. Here Tilt refers to the approximation due to exponential tilting. Expand refers to the approximation due to a first order expansion.

From the table it is easy to see that exponential tilting continues to perform well for points in the far tails. It also performs well for points near the mean of the distribution. However, it does not perform as well as the expansion until the points are in the tails of the distribution.

Table 5.1: Approximating the Binomial Distribution with $n = 50$ and $p = .1$ using Exponential Tilting

y	Binomial	Poisson	Expand	Tilt
2	.077943	.084224 (-.08)	.078329 (-.005)	.076373 (.02)
4	.180905	.175467 (.03)	.180731 (.001)	.173543 (.04)
6	.154104	.146223 (.05)	.153534 (.004)	.144595 (.06)
8	.064278	.065278 (-.02)	.064625 (-.005)	.058930 (.08)
11	.006135	.008242 (-.34)	.006182 (-.008)	.005421 (.12)
13	.000719	.001321 (-.84)	.000647 (.10)	.000619 (.14)
15	.000056	.000157 (-1.80)	.000024 (.57)	.000047 (.16)
16	.000014	.000049 (-2.50)	-.000002 (1.14)	.000011 (.21)

Table 5.2 presents the results for right tail probabilities. Again the distribution being approximated is a binomial with $n = 50$ and $p = .1$. This table shows the same trends as the previous table. Exponential tilting appears to be at least reasonably accurate for points in the far tails of the distribution.

Table 5.3 presents the results for the left tail of the binomial distribution. Here the binomial with $n = 50$ and $p = .4$ is used. From the table it may again be seen that the approximation due to exponential tilting is highly accurate for points in the tails of the distribution. In this case they are much more accurate than the previous cases. To see why, recall that the conjugate density for the binomial distribution is also a binomial distribution, with the p being replaced by y/n. Thus when approximating points in the left tail of the distribution, the p gets smaller. It is well known that the Poisson approximation is better for small p, thus the approximation of the conjugate density is highly accurate. For right tail probabilities the method performs well since the conjugate density is being approximated at it's mean. However the $p = y/n$ is getting larger, and thus moving away from the ideal setting for a Poisson approximation.

The results in this section may be compared with the results from chapter 4 to see how tilting performs compared to the density based expansion. However, it should be noted that exponential tilting is easier to perform for this example. In general, if the probability generating function has a simple form, then exponential tilting will be easy to perform.

Table 5.2: Approximating the Right Tail of a Binomial Distribution with $n = 50$ and $p = .1$ using Exponential Tilting

y	Binomial	Poisson	Tilt	Expand
6	.383877	.384039 (-.0004)	.369148 (.04)	.384039 (-.0004)
8	.122145	.133372 (-.09)	.113397 (.07)	.122927 (-.006)
10	.024538	.031828 (-.30)	.022114 (.10)	.024575 (-.0015)
12	.003220	.005453 (-.69)	.002821 (.12)	.002980 (.075)
14	.000285	.000698 (-1.45)	.000243 (.15)	.000170 (.40)
16	.000017	.000069 (-3.06)	.000014 (.18)	$-9.61 * 10^{-6}$ (1.57)
18	$7.60 * 10^{-7}$	$5.42 * 10^{-6}$ (-6.13)	$6.09 * 10^{-7}$ (.20)	$-3.26 * 10^{-6}$ (5.29)

5.2 Tilting with Constraints

As seen previously, the usual Poisson approximation and Exponential tilting may come about by approximating the factorial cumulant generating function with a first order Taylor series. Consider another linear approximation to the factorial cumulant generating function. Let $L(q)$ represent the approximation to $k_x(q)$, and again let $\hat{q}$ denote the solution to $y = qk'_x(q)$. Now impose the following two constraints:

1. $L(1) = k_x(1) = 0$.
2. $L(\hat{q}) = k_x(\hat{q})$.

Then it may easily be found that $L(q) = \frac{k_x(\hat{q})(q-1)}{\hat{q}-1}$. Notice that the first constraint is satisfied by the approximation leading to the usual Poisson approximation. The second constraint is satisfied by the approximation leading to exponential tilting. Thus this approximation is an attempt to satisfy both of these conditions.

Approximating the factorial cumulant generating function in this way leads to the following approximation:

$$f(x) \approx P_{\lambda_1}(x) \tag{5.8}$$

Table 5.3: Approximating the Left Tail of a Binomial Distribution with $n = 50$ and $p = .4$ using Exponential Tilting

y	Binomial	Poisson	Tilt	Expand
18	.335613	.381422 (-.13)	.308519 (.08)	.347665 (-.04)
16	.156091	.221074 (-.42)	.139358 (.11)	.169425 (-.09)
14	.053955	.104864 (-.94)	.047985 (.11)	.058380 (-.08)
12	.013251	.039012 (-1.94)	.011886 (.10)	.010812 (.18)
10	.002197	.010812 (-3.92)	.001999 (.09)	-.000821 (1.37)
8	.000231	.002087 (-8.03)	.000214 (.07)	-.001054 (5.56)
6	.000014	.000255 (-17.21)	.000013 (.07)	-.000258 (19.42)

where $\lambda_1 = \frac{k_x(\hat{q})}{\hat{q}-1}$. Thus the approximation is similar to the usual Poisson approximation. The only difference is the choice for the mean. The resulting approximation due to tilting with the constraints is analogous to work done in Lugananni and Rice(1980).

Notice that the approximation to the factorial cumulant generating function associated with exponential tilting is not zero when $q = 1$. Thus the sum of the approximation to the density will not equal one. So the approximation due to tilting with the constraints is in essence renormalizing the approximation due to exponential tilting so that the approximation to the density will sum to one.

5.2.1 Example: Binomial Distribution

To see how this works, consider approximating the binomial distribution. All that is required is $\hat{q} = \frac{(1-p)y}{p(n-y)}$ and $k_x(\hat{q}) = n\ln(1 - p + p\hat{q})$. Table 5.4 presents the results for approximating point probabilities for a binomial with $n = 50$ and $p = .1$. Here New Tilt refers to the tilting with the constraints.

From table 5.4 it appears that the tilting with the constraints is comparable to exponential tilting. It is much more accurate than the usual Poisson approximation. Tilting with constraints is a little easier to implement than exponential tilting.

Table 5.5 presents the results for approximating the right tail probabilities of a binomial with $n = 50$ and $p = .1$. Again the tilting with the constraints is comparable to exponential

Table 5.4: Approximating the Binomial Distribution with $n = 50$ and $p = .1$ using Tilting with the Constraint

y	Binomial	Poisson	Expand	Tilt	New Tilt
2	.077943	.084224 (-.08)	.078329 (-.005)	.076373 (.02)	.076294 (.02)
4	.180905	.175467 (.03)	.180731 (.001)	.173543 (.04)	.173501 (.04)
6	.154104	.146223 (.05)	.153534 (.004)	.144595 (.06)	.144540 (.06)
8	.064278	.065278 (-.02)	.064625 (-.005)	.058930 (.08)	.058651 (.088)
11	.006135	.008242 (-.34)	.006182 (-.008)	.005421 (.12)	.005270 (.14)
13	.000719	.001321 (-.84)	.000647 (.10)	.000619 (.14)	.000582 (.19)
15	.000056	.000157 (-1.80)	.000024 (.57)	.000047 (.16)	.000042 (.25)
16	.000014	.000049 (-2.50)	-.000002 (1.14)	.000011 (.21)	.000010 (.29)

tilting. For points in the near tails, tilting with constraints is slightly more accurate. For points in the far tails, exponential tilting is more accurate. This suggests that imposing the constraints is helpful for points in the near tails, but not beneficial for points in the far tails.

Table 5.6 presents the results for approximating the left tail probabilities for a binomial with $n = 50$ and $p = .4$. The trend is the same as that seen in the previous table. Just as in Exponential tilting, the tilting with the constraints seems to perform better for left tail probabilities than right tail probabilities.

As seen in the examples in this section, the tilting with the constraints is comparable to exponential tilting. It sometimes performs better in the near tails. However, it does not perform as well as exponential tilting in the far tails. One benefit to this modified version of tilting is that it is a very simple approximation. It also shows that the mean of the distribution being approximated is not necessarily the best choice for the mean of the Poisson being used to approximate it.

5.3 Tilting for Finite Populations

In exponential tilting, the probability generating function was used to form the conjugate density. The conjugate density was formed in such a way that it was a member of the same

Table 5.5: Approximating the Right Tail of a Binomial Distribution with $n = 50$ and $p = .1$ using Tilting with the Constraint

y	Binomial	Poisson	Tilt	New Tilt	Expand
6	.383877	.384039 (-.0004)	.369148 (.04)	.374219 (.025)	.384039 (-.0004)
8	.122145	.133372 (-.09)	.113397 (.07)	.116181 (.05)	.122927 (-.006)
10	.024538	.031828 (-.30)	.022114 (.10)	.022525 (.08)	.024575 (-.0015)
12	.003220	.005453 (-.69)	.002821 (.12)	.002810 (.13)	.002980 (.075)
14	.000285	.000698 (-1.45)	.000243 (.15)	.000232 (.19)	.000170 (.40)
16	.000017	.000069 (-3.06)	.000014 (.18)	.000013 (.24)	$-9.61 * 10^{-6}$ (1.57)
18	$7.60 * 10^{-7}$	$5.42 * 10^{-6}$ (-6.13)	$6.09 * 10^{-7}$ (.20)	$4.89 * 10^{-7}$ (.36)	$-3.26 * 10^{-6}$ (5.29)

family as the original density, so long as the original density is a member of the exponential family. Here a conjugate density will be formed in the same way as in exponential tilting. The focus of this section will be on the hypergeometric distribution. This technique will be illustrated for other distributions in later sections.

Notice from the derivation of $\varphi_x(Q)$ for the hypergeometric distribution in Section 3.4.2 that:

$$\left(\frac{Q^{(W-X)}}{W^{(W-X)}}\right)\frac{Hypergeometric(W+B,W,n)}{\varphi_x(Q)} = Hypergeometric(Q+B,Q,n-W+Q).$$

Refer to this as the conjugate density. The mean of this conjugate density is $\frac{Q(n-W+Q)}{Q+B}$. The random variable with the hypergeometric distribution is now $X-W+Q$ as opposed to just X. Close inspection of the conjugate density reveals that it can be formed from the original density by removing $W-Q$ subjects from the cell of interest.

Consider a generalized version of this. Let $f(x)$ be the density of interest. Then let the following represent the conjugate density:

$$f_Q(x) = \left(\frac{Q^{(W-X)}}{W^{(W-X)}}\right)\frac{f(x)}{\varphi_x(Q)}.$$

The generating function for the conjugate density is derived as follows:

Table 5.6: Approximating the Left Tail of a Binomial Distribution with $n = 50$ and $p = .4$ using Tilting with the Constraint

y	Binomial	Poisson	Tilt	New Tilt	Expand
18	.335613	.381422 (-.13)	.308519 (.08)	.328291 (.02)	.347665 (-.04)
16	.156091	.221074 (-.42)	.139358 (.11)	.148992 (.045)	.169425 (-.09)
14	.053955	.104864 (-.94)	.047985 (.11)	.050164 (.07)	.058380 (-.08)
12	.013251	.039012 (-1.94)	.011886 (.10)	.012030 (.09)	.010812 (.18)
10	.002197	.010812 (-3.92)	.001999 (.09)	.00196 (.11)	-.000821 (1.37)
8	.000231	.002087 (-8.03)	.000214 (.07)	.000204 (.12)	-.001054 (5.56)
6	.000014	.000255 (-17.21)	.000013 (.07)	.0000124 (.11)	-.000258 (19.42)

$$\begin{aligned}
\varphi_{(X-W+Q)}(T) &= \sum_x \frac{T^{(Q-(X-W+Q))}}{Q^{(Q-(X-W+Q))}} f_Q(x) \\
&= \sum_x \left(\frac{T^{(W-X)} Q^{(W-X)}}{Q^{(W-X)} W^{(W-X)}} \right) \frac{f(x)}{\varphi_x(Q)} \\
&= \sum_x \frac{\frac{T^{(W-X)}}{W^{(W-X)}} f(x)}{\varphi_x(Q)} \\
&= \frac{\varphi_x(T)}{\varphi_x(Q)}.
\end{aligned}$$

Let X_Q represent a random variable from the conjugate density. The mean of X_Q is:

$$E(X_Q) = \frac{Q\varphi_x(Q-1)}{\varphi_x(Q)}.$$

For the hypergeometric this will yield $\frac{Q(n-W+Q)}{Q+B}$, the same expression as above. This seems to suggest tilting by choosing Q such that

$$\frac{Q\varphi_x(Q-1)}{\varphi_x(Q)} = x - W + Q \tag{5.9}$$

where x is the point where the density is to be estimated. Note that $x - W + Q$ is needed instead of x alone since, as mentioned above, the random variable for the conjugate density is $X - W + Q$, not just X. This is analogous to exponential tilting where q is chosen such that

$$\frac{q\pi_x'(q)}{\pi_x(q)} = y.$$

For the hypergeometric distribution, this simplifies to

$$Q = \frac{B(W-y)}{B-n+y}. \tag{5.10}$$

To perform tilting, approximate the conjugate density with the Poisson density. That is:

$$\left(\frac{Q^{(W-x)}}{W^{(W-x)}}\right)\frac{f(x)}{\varphi_x(Q)} = P_{x^*}(x-W+Q)$$

where $P_{x^*}(x-W+Q)$ is the Poisson density at point $x - W + Q$ with $x^* = x - W + Q$. Solving for $f(x)$ yields the following:

$$\begin{aligned} f(x) &= \varphi_x(Q)\frac{W^{(W-x)}}{Q^{(Q-x)}}P_{x^*}(x-W+Q) \\ &= \varphi_x(Q)\left(\frac{W!(Q-W+x)!}{x!Q!}\right)\frac{e^{-x^*}(x^*)^{(x-W+Q)}}{(x-W+Q)!} \\ &= \varphi_x(Q)\frac{W!}{Q!}(x^*)^{Q-W}\frac{e^{-x^*}(x^*)^x}{x!} \\ &= \frac{\varphi_x(Q)W!(x^*)^{Q-W}}{Q!}P_{x^*}(x). \end{aligned} \tag{5.11}$$

Tail probabilities may be found by summing both sides of the expression. This leads to the following approximations:

$$P(X \leq y) = \frac{\varphi_x(Q)W!(y^*)^{Q-W}}{Q!}R_{y^*}(y) \tag{5.12}$$

$$P(X \geq y) = \frac{\varphi_x(Q)W!(y^*)^{Q-W}}{Q!}R^*_{y^*}(y) \tag{5.13}$$

where $y^* = y - W + Q$ here.

5.3.1 Example: Hypergeometric

Now consider this method for approximating the hypergeometric distribution. Section 3.4.2 showed how to obtain the generating function. The previous section showed how to obtain the value of Q needed to obtain the desired mean. Table 5.7 displays the results for approximating right tail probabilities for $W = 20$, $B = 30$,and $n = 15$.

Table 5.7: Approximating Right Tail Probabilities of a Hypergeometric Distribution with $W = 20$, $B = 30$,and $n = 15$ using Tilting

y	Hypergeometric	Poisson	Tilted Approx.
7	.373808	.393697 (-.04)	.309841 (.18)
8	.172229	.25602 (-.49)	.137661 (.20)
9	.058293	.152736 (-1.62)	.045854 (.21)
10	.013985	.083924 (-5.00)	.012689 (.09)
11	.002288	.042621 (-18.12)	.001976 (.11)
12	.000243	.020092 (-81.78)	.000204 (.16)

Table 5.8 displays the results for approximating left tail probabilities for $W = 75$, $B = 100$, and $n = 25$. From the previous two tables, it can be seen that the tilting using this new generating function can lead to substantial improvements over the Poisson approximation. It may also be seen that the approximation works a little better for right tail probabilities than for left tail probabilities. The reason is similar to the reason that exponential tilting works better for approximating the left tail probabilities of a binomial. Recall that the Poisson approximation to a hypergeometric is better when W and n are less than B. For a discussion on this, see Terrell (1999). Here the conjugate density is hypergeometric with W being replaced by Q. For right tail probabilities Q is less than W, thus moving into a setting where the Poisson approximation is better. For left tail probabilities Q is greater than W. Thus the approximation benefits from being for a point at the mean of the distribution, but it suffers from moving away from the setting where Poisson approximations work their best. However, in both cases the tilting leads to a considerable improvement over the usual Poisson approximation.

Table 5.8: Approximating Left Tail Probabilities of a Hypergeometric Distribution with $W = 75$, $B = 100$, and $n = 25$ using Tilting

y	Hypergeometric	Poisson	Tilted Approx.
9	.300313	.372304 (-.24)	.22868 (.24)
8	.167092	.258345 (-.55)	.121977 (.27)
7	.078667	.162619 (-1.07)	.055816 (.29)
6	.030699	.091143 (-1.97)	.021348 (.30)
5	.009696	.044446 (-3.58)	.006416 (.34)
4	.002405	.018296 (-6.61)	.001575 (.35)

5.4 Discussion

Several different forms of tilting have been presented in this chapter. Exponential tilting has been shown to lead to accurate approximations for points in the tails of the distribution. It also performs well for points near the mean of the distribution. It is most useful when the probability generating function can be expressed in terms of simple functions. Although it is not always the most accurate method, it is one of the most consistent in that it does not lose its accuracy rapidly in any part of the distribution.

Viewing exponential tilting as the result of a linear expansion or approximation to the factorial cumulant generating function leads to the idea of tilting with the constraints. The approximation resulting from imposing the constraints is very simple. The accuracy of the method is comparable to that of exponential tilting. Like exponential tilting, it is also most useful when the probability generating function can be expressed in terms of simple functions.

Distributions associated with finite populations do not always have simple probability generating functions. In these cases the finite population probability generating function turns out to be useful. The form of tilting using this new generating function is very similar in form to exponential tilting. Like exponential tilting, it is accurate for points that are in the tails of the distribution. It also performs well for points near the mean of the distribution.

Observing the parameter that is being tilted may lead to knowing when tilting will be most effective. In some cases the parameter will be tilted in such a way that the Poisson approximation to the conjugate density will be more appropriate. This was the case for using exponential tilting in the left tail of the binomial distribution, and for using the finite population version of tilting in the right tail of the hypergeometric distribution. In some cases the parameter is being tilted in such a way that the Poisson approximation to the conjugate density is less appropriate. In these cases, the approximations of the original density due to tilting is still more accurate than the usual Poisson approximation.

Chapter 6

Combining Tilting and Expansion

Highly accurate approximations may be obtained by combining the ideas of tilting and expansion. When exponential tilting is combined with the Edgeworth Expansion, the saddlepoint approximations are obtained. Thus the work in this chapter will be analogous to the development of the saddlepoint approximations.

6.1 Exponential Tilting and Expansion

Barbour and Jensen (1989) suggest that exponential tilting and the Edgeworth-like expansion may be combined to yield highly accurate results. The basic idea behind combining the approaches begins with performing exponential tilting. However, instead of approximating the conjugate density with a Poisson density, it is approximated using equation (4.2). This approximation is as follows:

$$\frac{q^x f(x)}{\pi_x(q)} = \left(p(x) + \frac{k_q^{(2)}}{2}\nabla^2 p(x) + \left[-\frac{k_q^{(3)}}{6}\nabla^3 p(x) + \frac{(k_q^{(2)})^2}{8}\nabla^4 p(x) \right] + \cdots \right)$$

where $k_q^{(i)}$ is the i^{th} factorial cumulant of the transformed density. Solving for $f(x)$ leads to the following approximation for point probabilities:

$$f(x) = \pi_x(q) q^{-x} \left(p(x) + \frac{k_q^{(2)}}{2}\nabla^2 p(x) + \left[-\frac{k_q^{(3)}}{6}\nabla^3 p(x) + \frac{(k_q^{(2)})^2}{8}\nabla^4 p(x) \right] + \cdots \right). \qquad (6.1)$$

Usually when Poisson approximations are performed, the mean of the Poisson is taken to be the mean of the distribution being approximated. Equation (6.1) seems to suggest a different value for the mean of the Poisson. If the mean is chosen such that $\nabla^2 p(x) = 0$, then the first correction term will equal zero. Thus the accuracy obtained by using up to the first correction term in the expression in equation (6.1) may be obtained through exponential tilting alone. The expression $\nabla^2 p(x) = 0$ may be rewritten as:

$$\nabla^2 p(x) = p(x-2)\left(\frac{\lambda^2}{x(x-1)} - \frac{2\lambda}{x-1} + 1\right).$$

Setting this equal to zero and solving for λ leads to $\lambda = x - \sqrt{x}$ and $\lambda = x + \sqrt{x}$. Thus it appears that setting $qk'_x(q) = x - \sqrt{x}$ and solving for q, and then performing exponential tilting in the usual way will lead to a more accurate approximation. In fact, empirical results suggest that this method is more accurate than using equation (6.1) with one correction term. Using $x + \sqrt{x}$ leads to slightly less accurate results.

Tail probability approximations may be obtained by several methods. One method is to approximate the necessary point probabilities, and then add them. This is the approach discussed in Barbour and Jensen (1989), and is analogous to the normal setting where the density is approximated, and then numerical integration techniques are used. A simpler approach is to perform exponential tilting and the expansion method one time, and then take the summation of the results. This leads to a simple expression for the tail probability. The derivation is contained in section B.1. The derivation involves using the properties of backward differencing and summation by parts, which are discussed in Appendix A. The end results are as follows:

$$P(X \leq y) = \pi_x(q)e^{-(y-\frac{y}{q})}\left(R_{y/q}(y) + \frac{k_q^{(2)}}{2}L_2(q) - \frac{k_q^{(3)}}{6}L_3(q) + \frac{(k_q^{(2)})^2}{8}L_4(q) + \cdots\right) \quad (6.2)$$

$$P(X \geq y) = \pi_x(q)e^{-(y-\frac{y}{q})}\left(R^*_{y/q}(y) + \frac{k_q^{(2)}}{2}L^*_2(q) - \frac{k_q^{(3)}}{6}L^*_3(q) + \frac{(k_q^{(2)})^2}{8}L^*_4(q) + \cdots\right) \quad (6.3)$$

where $R_{y/q}(y)$ and $R^*_{y/q}(y)$ are as defined in equations (4.3) and (4.4). $L_k(q)$ and $L^*_k(q)$ are defined as:

$$\begin{aligned} L_k(q) = {} & \left(\frac{q-1}{q}\right)^k R_{y/q}(y) + \left(\frac{y!e^y}{qy^y}\right) P_{y/q}(y)[\nabla^{k-1}p(y) + \\ & \left(\frac{q-1}{q}\right)\nabla^{k-2}p(y) + \cdots + \left(\frac{q-1}{q}\right)^{k-1} p(y)] \end{aligned} \quad (6.4)$$

$$\begin{aligned} L^*_k(q) = {} & \left(\frac{q-1}{q}\right)^k R^*_{y/q}(y) - \left(\frac{y!e^y}{y^y}\right) P_{y/q}(y)[\nabla^{k-1}p(y-1) + \\ & \left(\frac{q-1}{q}\right)\nabla^{k-2}p(y-1) + \cdots + \left(\frac{q-1}{q}\right)^{k-1} p(y-1)] \end{aligned} \quad (6.5)$$

where $P_{y/q}(y)$ is the probability that a Poisson random variable with mean y/q is equal to y. Since highly accurate approximations are obtained with one or two correction terms, there is seldomly a need to compute the L functions beyond $L_4(q)$ or $L_4^*(q)$. Thus these tail probability approximations are fairly easy to compute.

An interesting observation about combining exponential tilting with an expansion in the Poisson setting is that the end result can be expressed as a series where the terms are expressed as a function involving Poisson information multiplied by the factorial cumulants. This holds for both the point probability approximations and the tail probability approximations. Thus the factorial cumulants are an important factor when considering these approximations. As mentioned previously, the factorial cumulants for a Poisson are all zero beyond the first one. The first is simply the mean of the Poisson distribution. Thus an important criteria for Poisson approximation methods to be successful is that the cumulants are at least decreasing. In deriving the expansion method in section 4.2, it was assumed that the cumulant were decreasing by some order of magnitude. An example is presented later which reveals that this method has problems when the cumulants are increasing as opposed to decreasing.

6.1.1 Example: Binomial

To illustrate the use of combining exponential tilting with the Edgeworth-Like expansion, consider approximating the binomial distribution. The information needed to perform the approximations is the same as what was needed to perform exponential tilting and the Edgeworth-like expansion. Only now for the expansion, the factorial cumulants from the conjugate density will be needed. These will not be difficult to obtain since the conjugate density for approximating a binomial is a binomial with p being replaced by y/n.

Table 6.1 shows results for approximating the point probabilities of a binomial distribution with $n = 50$ and $p = .1$. In the table Expand refers to the expansion method alone, Tilt refers to exponential tilting alone, Tilt2 refers to exponential tilting where the mean of the conjugate density is set to $y - \sqrt{y}$ as opposed to y, and Comb refers to combining exponential tilting with the expansion method. The combination method here only uses one correction term.

From the table it is easy to see that the second form of tilting and the combination method give highly accurate results, even far out in the tails. The second form of tilting yields a considerable improvement over exponential tilting, while still maintaining the simplicity. Unfortunately it does not lead to a useful approximation to tail probabilities.

Table 6.2 displays the results for right tail approximations using the binomial $n = 50$ and $p = .1$. Exp1 and Exp2 refer to the expansion method using one and two correction terms respectively. Comb1 refers to using the combination method with only one correction term, and comb2 refers to using two correction terms.

Table 6.1: Approximating the Binomial Distribution with $n = 50$ and $p = .1$ by Combining Tilting and Expansion

y	Binomial	Poisson	Expand	Tilt	Tilt2	Comb.
2	.077943	.084224 (-.08)	.078329 (-.005)	.076373 (.02)	.077950 $(-9 * 10^{-5})$	.077901 (.0005)
4	.180905	.175467 (.03)	.180731 (.001)	.173543 (.04)	.181044 (-.0008)	.180484 (.002)
6	.154104	.146223 (.05)	.153534 (.004)	.144595 (.06)	.154422 (-.002)	.153271 (.005)
8	.064278	.065278 (-.02)	.064625 (-.005)	.058930 (.08)	.064554 (-.004)	.0636448 (.01)
11	.006135	.008242 (-.34)	.006182 (-.008)	.005421 (.12)	.006195 (-.01)	.006017 (.02)
13	.000719	.001321 (-.84)	.000647 (.10)	.000619 (.14)	.000730 (-.02)	.000700 (.03)
15	.000056	.000157 (-1.80)	.000024 (.57)	.000047 (.16)	.000058 (-.04)	.000054 (.04)
16	.000014	.000049 (-2.50)	-.000002 (1.14)	.000011 (.21)	.000014 (0.00)	.000013 (.07)

From table 6.2, it becomes clear that the combination methods are highly accurate even in the tails. The expansion method with two correction terms does well until the values move further out in the tails. Once again, the expansion method with only one correction term has problems, which is evident from the negative values. Exponential tilting does well even in the tails. Although it is not the most accurate method, it always seems to perform at least reasonably well.

Table 6.3 displays results for left tail probability approximations. The table is set up in the same way as table 6.2. A binomial with $n = 50$ and $p = .4$ is used. The table shows the same trends previously seen. Again the combination methods perform very well. Here the expansion methods perform very well near the mean with the accuracy diminishing in the tails. In this example the accuracy of the expansion method is much less favorable compared to the combination method. The reason for this is that the conditions are less favorable for the Poisson approximation. Recall that the expansion method does not perform as well as the conditions become less favorable for the Poisson approximation. The combination methods tend to continue performing well. Also note that these are approximations for points in the left tails, thus the probability of a "success" for the conjugate density is getting smaller as the points move out into the left tail.

Table 6.2: Approximating the Right Tail of a Binomial Distribution with $n = 50$ and $p = .1$ by Combining Tilting and Expansion

y	Binomial	Poisson	Tilt	Exp1	Exp2	Comb1	Comb2
6	.383877	.384039 (-.0004)	.369148 (.04)	.384039 (-.0004)	.383893 (-.00004)	.382906 (.003)	.383796 (.0002)
8	.122145	.133372 (-.09)	.113397 (.07)	.122927 (-.006)	.122196 (-.0004)	.121227 (.008)	.122036 (.0009)
10	.024538	.031828 (-.30)	.022114 (.10)	.024575 (-.0015)	.024509 (.0012)	.024195 (.014)	.024484 (.0022)
12	.003220	.005453 (-.69)	.002821 (.12)	.002980 (.075)	.003200 (.006)	.003150 (.022)	.003206 (.004)
14	.000285	.000698 (-1.45)	.000243 (.15)	.000170 (.40)	.000289 (-.014)	.000276 (.032)	.000283 (.007)
16	.000017	.000069 (-3.06)	.000014 (.18)	$-9.61 * 10^{-6}$ (1.57)	.000022 (-.29)	.000017 (0.00)	.000017 (0.00)
18	$7.60 * 10^{-7}$	$5.42 * 10^{-6}$ (-6.13)	$6.09 * 10^{-7}$ (.20)	$-3.26 * 10^{-6}$ (5.29)	$2.12 * 10^{-6}$ (-1.79)	$7.18 * 10^{-7}$ (.055)	$7.47 * 10^{-7}$ (.017)

6.2 Tilting and Expansion with a Constraint

Consider rewriting the factorial cumulant generating function in the following way:

$$k_x(q) = (q-1)m(q)$$

where $m(q) = \frac{k_x(q)}{(q-1)}$ is a function of q. Now approximate $m(q)$ by expanding it with a Taylor series about the point $q = \hat{q}$, where $\hat{q}$ is again the solution to $qk_x'(q) = y$. For now, consider only a linear expansion. This leads to the following:

$$k_x(q) \approx (q-1)[m(\hat{q}) + (q-\hat{q})m'(\hat{q})].$$

The approximation to the factorial cumulant generating function now has the constraint that it is equal to zero when q is equal to one. Now this leads to approximating the probability generating function in the following way:

$$\pi_x(q) \approx \exp[(q-1)m(\hat{q}) + (q-1)(q-\hat{q})m'(\hat{q})].$$

Notice that $m(\hat{q}) = \frac{k_x(\hat{q})}{\hat{q}-1}$, which was called λ_1 in section 5.2. This leads to the following:

$$\pi_x(q) \approx \pi_{\lambda_1}(q)\exp[(q-1)(q-\hat{q})m'(\hat{q})]$$

Table 6.3: Approximating the Left Tail of a Binomial Distribution with $n = 50$ and $p = .4$ by Combining Tilting and Expansion

y	Binomial	Poisson	Tilt	Exp1	Exp2	Comb1	Comb2
18	.335613	.381422 (-.13)	.308519 (.08)	.347665 (-.04)	.339338 (-.01)	.330462 (.02)	.33434 (.004)
16	.156091	.221074 (-.42)	.139358 (.11)	.169425 (-.09)	.159096 (-.02)	.152980 (.02)	.155391 (.004)
14	.053955	.104864 (-.94)	.047985 (.11)	.058380 (-.08)	.053629 (.006)	.052915 (.02)	.053743 (.004)
12	.013251	.039012 (-1.94)	.011886 (.10)	.010812 (.18)	.011893 (.10)	.013038 (.02)	.013212 (.003)
10	.002197	.010812 (-3.92)	.001999 (.09)	-.000821 (1.37)	.001738 (.21)	.002171 (.01)	.002193 (.002)
8	.000231	.002087 (-8.03)	.000214 (.07)	-.001054 (5.56)	.000335 (-.45)	.000229 (.009)	.000230 (.004)
6	.000014	.000255 (-17.21)	.000013 (.07)	-.000258 (19.42)	.000104 (-6.43)	.000014 (0.00)	.000014 (0.00)

where $\pi_{\lambda_1}(q)$ is the probability generating function for a Poisson random variable with mean λ_1. Now expand the exponential in a power series. Again consider only the first order series for now. This leads to the following:

$$\pi_x(q) \approx \pi_{\lambda_1}(q)[1 + (q-1)(q-\hat{q})m'(\hat{q})].$$

Now consider the following:

$$\sum_{x=0}^{\infty} p_{\lambda_1}(x) q^x (q-1)(q-\hat{q})m'(\hat{q}) = m'(\hat{q}) \sum_{x=0}^{\infty} p_{\lambda_1}(x) q^x (q^2 - (\hat{q}+1)q + \hat{q}).$$

If the approximation is at the point x, this suggests the following approximation:

$$\begin{aligned} f(x) &\approx p_{\lambda_1}(x) + m'(\hat{q})(\hat{q}p_{\lambda_1}(x) - (\hat{q}+1)p_{\lambda_1}(x-1) + p_{\lambda_1}(x-2)) \\ &= p_{\lambda_1}(x) + m'(\hat{q})(\hat{q}\nabla p_{\lambda_1}(x) - \nabla p_{\lambda_1}(x-1)). \end{aligned}$$

Now $m'(\hat{q})$ may be found in the following way:

$$m'(\hat{q}) = \left(\frac{k_x(\hat{q})}{\hat{q}-1}\right)'$$

$$
\begin{aligned}
&= \frac{(\hat{q}-1)k_x'(\hat{q}) - k_x(\hat{q})}{(\hat{q}-1)^2} \\
&= \frac{x/\hat{q} - \lambda_1}{\hat{q}-1}.
\end{aligned}
$$

Thus the approximation for $f(x)$ is as follows:

$$
f(x) \approx p_{\lambda_1}(x) + \left[\frac{x/\hat{q} - \lambda_1}{\hat{q}-1}\right](\hat{q}\nabla p_{\lambda_1}(x) - \nabla p_{\lambda_1}(x-1)). \tag{6.6}
$$

To obtain approximations for tail probabilities, sum both sides of the expression to obtain the following:

$$
P(X \le y) \approx R_{\lambda_1}(y) + \left[\frac{y/\hat{q} - \lambda_1}{\hat{q}-1}\right](\hat{q}p_{\lambda_1}(y) - p_{\lambda_1}(y-1)) \tag{6.7}
$$

$$
P(X \ge y) \approx R^*_{\lambda_1}(y) + \left[\frac{y/\hat{q} - \lambda_1}{\hat{q}-1}\right](p_{\lambda_1}(y-2) - \hat{q}p_{\lambda_1}(y-1)). \tag{6.8}
$$

Notice that the first term of the approximations are the results obtained from tilting with the constraints, presented in section 5.2. Thus this may be viewed as incorporating an expansion with the tilting with constraints. The resulting approximations are simple. To determine how the methods works, they will be used to approximate the binomial distribution in the following section.

6.2.1 Example: Binomial

To use the methods in the previous section, $\hat{q}$ and $k_x(\hat{q})$ are needed. For the binomial distribution, $\hat{q} = \frac{(1-p)y}{p(n-y)}$ and $k_x(\hat{q}) = n\ln(1-p+p\hat{q})$. Table 6.4 displays the results for approximating binomial right tail probabilities with $n = 50$ and $p = .1$. In the table, TilExpCon is being used to denote tilting and expanding with a constraint.

From table 6.4 it appears that the method in equation(6.8) leads to very accurate results. It is more accurate than the combination method using one correction term, and often performs better than the combination method using two correction terms. It should also be noted that tilting and expanding with a constraint leads to an approximation that is easier to implement than the previous combination methods.

Table 6.5 displays the results for approximating binomial left tail probabilities with $n = 50$ and $p = .4$. This table reveals that the approximation in equation(6.8) is very accurate. However it is not as accurate as the combination methods presented earlier. As mentioned

Table 6.4: Approximating the Right Tail of a Binomial Distribution with $n = 50$ and $p = .1$ by Combining Tilting and Expansion using the Constraint

y	Binomial	Poisson	Comb1	Comb2	TilExpCon
6	.383877	.384039 (-.0004)	.382906 (.003)	.383796 (.0002)	.383409 (.001)
8	.122145	.133372 (-.09)	.121227 (.008)	.122036 (.0009)	.122047 (.0008)
10	.024538	.031828 (-.30)	.024195 (.014)	.024484 (.0022)	.024581 (-.002)
12	.003220	.005453 (-.69)	.003150 (.022)	.003206 (.004)	.003239 (-.006)
14	.000285	.000698 (-1.45)	.000276 (.032)	.000283 (.007)	.000288 (-.01)
16	.000017	.000069 (-3.06)	.000017 (0.00)	.000017 (0.00)	.000017 (0.00)
18	$7.60 * 10^{-7}$	$5.42 * 10^{-6}$ (-6.13)	$7.18 * 10^{-7}$ (.055)	$7.47 * 10^{-7}$ (.017)	$7.50 * 10^{-7}$ (.01)

earlier, the combination methods are very accurate for binomial left tail probabilities due to the fact that the probability of success for the conjugate density is decreasing. Thus the constraint being imposed by the methods in this section is not as useful. However, the approximation in equation(6.8) is of a simpler form than the combination methods.

6.3 Tilting and Expansion for Finite Populations

To combine tilting and expansion for finite population settings, begin with equation(5.3). Instead of approximating with a Poisson distribution, use the expansion of equation(4.2). For simplicity, the approximations will be presented using the expansion with only one correction term. This leads to the following approximation:

$$\left(\frac{Q^{(W-x)}}{W^{(W-x)}}\right)\frac{f(x)}{\varphi_x(Q)} \approx \left[p_{x^*}(x-W+Q) + \frac{k_Q^{(2)}}{2}\nabla^2 p_{x^*}(x-W+Q)\right]$$

where $k_Q^{(2)}$ is the second factorial cumulant from the conjugate density. Now solve the expression for $f(x)$ to obtain the following approximation:

$$f(x) = \varphi_X(Q)\frac{W!(Q-W+x)!}{x!Q!}\left[p(x-W+Q) + \frac{k_Q^{(2)}}{2}\nabla^2 p(x-W+Q)\right].$$

Table 6.5: Approximating the Left Tail of a Binomial Distribution with $n = 50$ and $p = .4$ by Combining Tilting and Expansion using the Constraint

y	Binomial	Poisson	Comb1	Comb2	TilExpCon
18	.335613	.381422 (-.13)	.330462 (.02)	.33434 (.004)	.337789 (-.006)
16	.156091	.221074 (-.42)	.152980 (.02)	.155391 (.004)	.159537 (-.02)
14	.053955	.104864 (-.94)	.052915 (.02)	.053743 (.004)	.056155 (-.04)
12	.013251	.039012 (-1.94)	.013038 (.02)	.013212 (.003)	.014041 (-.06)
10	.002197	.010812 (-3.92)	.002171 (.01)	.002193 (.002)	.002364 (-.08)
8	.000231	.002087 (-8.03)	.000229 (.009)	.000230 (.004)	.000251 (-.09)
6	.000014	.000255 (-17.21)	.000014 (0.00)	.000014 (0.00)	.000015 (-.07)

To form the approximations for tail probabilities, again sum both sides of the previous equations. Summing the right side of the expression involves using summation by parts. The details of this are presented in the Appendix in Section B.3. Taking the sum leads to the following approximations:

$$P(X \le y) \approx \varphi_X(Q)\frac{W!}{Q!}\left[(y^*)^{Q-W}R_{y^*}(y) + \frac{k_Q^{(2)}}{2}L_2(Q)\right] \tag{6.9}$$

$$P(X \ge y) \approx \varphi_X(Q)\frac{W!}{Q!}\left[(y^*)^{Q-W}R^*_{y^*}(y) + \frac{k_Q^{(2)}}{2}L^*_2(Q)\right] \tag{6.10}$$

where $L_2(Q)$ and $L_2^*(Q)$ are defined as follows:

$$\begin{aligned} L_2(Q) &= (W-Q+1)^{(2)}(y^*)^{Q-W-2}R_{y^*}(y+2) + \frac{(y^*+1)!}{(y+1)!}\left[\nabla p(y^*) + \frac{(W-Q)}{(y+2)}p(y^*)\right] \\ &- (Q-W)[\nabla p(Q^*) + (W-Q)p(Q^*)] \end{aligned}$$

$$L_2^*(Q) = (W-Q+1)^{(2)}(y^*)^{Q-W-2}R^*_{y^*}(y+2) - \frac{y^*!}{y!}\left[\nabla p(y^*-1) + \frac{(W-Q)}{(y+1)}p(y^*-1)\right]$$

with $Q^* = Q - W - 1$, $R_{y^*}(y+2) = \sum_{x=0}^{y} p(x+2)$, and $R^*_{y^*}(y+2) = \sum_{x=y}^{\infty} p(x+2)$.

Note that the form of the approximations here is similar to the saddlepoint approximations presented in Daniels (1987). The important thing to notice is that the approximations to tail probabilities have a closed form. Thus if there is a way to easily obtain Poisson probabilities, then approximations will be quick and easy to perform. Section B.3 of the appendix shows the form of the approximations for using more than one correction term.

6.3.1 Example: Hypergeometric

The hypergeometric distribution will be used here to illustrate the use of combining tilting and expansion for distributions associated with finite populations. Section 4.3.3 illustrated the use of the expansion to approximate hypergeometric probabilities. Section 5.3 illustrated the use of tilting. Table 6.6 displays the results for approximating right tail probabilities for a hypergeometric with $W = 20$, $B = 30$, and $n = 15$. In the table Tilt/Exp1 refers to using an expansion with one correction term along with tilting. Tilt/Exp2 refers to using an expansion with two correction terms. From the table it may be seen that combining tilting and expanding continues to lead to highly accurate results. These methods continue to do well for points in the tails of the distribution.

Table 6.6: Approximating Right Tail Probabilities of a Hypergeometric Distribution with $W = 20$, $B = 30$, and $n = 15$ by Combining Tilting and Expansion

y	Hypergeometric	Poisson	Tilt	Tilt/Exp1	Tilt/Exp2
7	.373808	.393697 (-.04)	.309841 (.18)	.361887 (.032)	.376551 (-.007)
8	.172229	.25602 (-.49)	.137661 (.20)	.164329 (.046)	.172242 (-.00008)
9	.058293	.152736 (-1.62)	.045854 (.21)	.054737 (.061)	.057346 (.016)
10	.013985	.083924 (-5.00)	.012689 (.09)	.013127 (.061)	.013707 (.020)
11	.002288	.042621 (-18.12)	.001976 (.11)	.002252 (.016)	.002321 (-.014)
12	.000243	.020092 (-81.78)	.000204 (.16)	.000231 (.049)	.000237 (.025)

Table 6.7 presents the results for approximating left tail probabilities with $W = 75$, $B = 100$, and $n = 25$. The notation in this table is the same as that in the previous table. From this table it may also be seen that combining tilting and expanding may lead to highly accurate results.

Table 6.7: Approximating Left Tail Probabilities of a Hypergeometric Distribution with $W = 75$, $B = 100$, and $n = 25$ by Combining Tilting and Expansion

y	Hypergeometric	Poisson	Tilt	Tilt/Exp1	Tilt/Exp2
9	.300313	.372304 (-.24)	.22868 (.24)	.268629 (.106)	.280929 (.065)
8	.167092	.258345 (-.55)	.121977 (.27)	.148076 (.114)	.157001 (.060)
7	.078667	.162619 (-1.07)	.055816 (.29)	.069225 (.120)	.074019 (.059)
6	.030699	.091143 (-1.97)	.021348 (.30)	.026889 (.124)	.028660 (.066)
5	.009696	.044446 (-3.58)	.006416 (.34)	.008482 (.125)	.008515 (.122)

6.4 Generalized Approximations

6.4.1 Generalizing Tilting and Expansion with a Constraint

In certain applications neither of the generating functions may have a simple form. It is also possible to have a generating function that is useful, but the value of the tilting parameter (either q or Q) is difficult to obtain. For these cases, it may still be possible to use the methods presented previously by approximating the necessary terms.

Consider using the approximations presented in equations (6.8) and (6.8). The only terms needed to perform the approximations are $\hat{q}$ and $k_x(\hat{q})$. If $k_x(q)$ is unknown, then it may be approximated in the following way:

$$k_x(q) \approx (q-1)k^{(1)} + \frac{1}{2}(q-1)^2 k^{(2)} + \frac{1}{6}(q-1)^3 k^{(3)} + \frac{1}{24}(q-1)^4 k^{(4)}. \tag{6.11}$$

Recall that $\hat{q}$ is found by solving $qk'_x(q) = y$. This could be accomplished by simply solving the polynomial, but this has been shown to lead to problems (see Wang (1992) for example). Consider a different approach to the problem. Let $h(q) = qk'_x(q)$. Now let $g(x) = h^{-1}(x)$. Notice that $h(1)$ is the mean of the distributed being approximated, call it μ. Then $g(\mu) = 1$. So $g(x)$ may be approximated in the following way:

$$g(x) \approx 1 + (x-\mu)g'(\mu) + \frac{1}{2}(x-\mu)^2 g''(\mu) + \frac{1}{6}(x-\mu)^3 g'''(\mu). \tag{6.12}$$

Now

$$\begin{aligned} g'(\mu) &= (h^{-1}(\mu))' = \frac{1}{h'(h^{-1}(\mu))} \\ &= \left. \frac{1}{k'_x(q) + qk''_x(q)} \right|_{q=1} \\ &= \frac{1}{k^{(1)} + k^{(2)}}. \end{aligned}$$

$g''(\mu)$ and $g'''(\mu)$ may be found in a similar way to be the following:

$$\begin{aligned} g''(\mu) &= \frac{-(2k^{(2)} + k^{(3)})}{(k^{(1)} + k^{(2)})^3} \\ g'''(\mu) &= \frac{-(3k^{(3)} + k^{(4)})}{(k^{(1)} + k^{(2)})^4} + \frac{3(2k^{(2)} + k^{(3)})^2}{(k^{(1)} + k^{(2)})^5}. \end{aligned}$$

Now solving $qk'_x(q) = y$ leads to solving $h(q) = y$ which implies that $q = h^{-1}(y)$, or $q = g(y)$. Thus $\hat{q}$ may be approximated by plugging y into equation (6.12). Approximating $\hat{q}$ in this way only requires the first four factorial cumulants, which may be obtained from the first four factorial moments.

6.4.2 Example: Binomial

The binomial distribution will be used to determine the impact of estimating $\hat{q}$ and $k_x(\hat{q})$. $\hat{q}$ will be estimated by the technique in equation (6.12) while $k_x(\hat{q})$ will be estimated using equation (6.11). Thus there are two separate approximations being used to obtain the parts needed to approximate the probabilities.

Table 6.8 presents the results for approximating right tail probabilities for a binomial with $n = 50$ and $p = .1$. TilExpCon still refers to the approximation due to tilting and expanding with the constraint. GenApprox refers to the generalized approximation. In the table it appears that the generalized approximation does very well. In a few cases it performs better than the approximation without approximating the necessary components. This happens in very few examples. In most other cases the generalized approximation is slightly less accurate. However the loss is usually small, and thus this method may be useful in cases where the terms must be approximated. More examples will be presented later.

6.4.3 Generalizing Tilting and Expansion for Finite Populations

In most finite population cases, if the factorial moments are easy to obtain, then the finite population probability generating function will also be easy to obtain. However in some

Table 6.8: Approximating the Right Tail of a Binomial Distribution with $n = 50$ and $p = .1$ using the Generalized Approximation

y	Binomial	Poisson	TilExpCon	GenApprox
6	.383877	.384039 (-.0004)	.383409 (.001)	.383408 (.001)
8	.122145	.133372 (-.09)	.122047 (.0008)	.122027 (.001)
10	.024538	.031828 (-.30)	.024581 (-.002)	.024553 (-.0006)
12	.003220	.005453 (-.69)	.003239 (-.006)	.003225 (-.002)
14	.000285	.000698 (-1.45)	.000288 (-.01)	.000284 (.004)
16	.000017	.000069 (-3.06)	.000017 (0.00)	.000017 (0.00)

cases the factorial moments may only be estimated as opposed to being obtained exactly. In these cases it may be necessary to have a means of obtaining the necessary pieces for the approximations presented in section 6.3.

Recall that solving for Q involved solving the following expression:

$$\frac{Q\varphi_x(Q-1)}{\varphi_x(Q)} + W - Q = y.$$

In the calculations for right tail probabilities that have been observed, Q is less than W. So now let $Q = W - k$ for some k. Then finding the value for Q will involve finding the appropriate value for k. Solving the above expression now leads to finding k to solve the following expression:

$$\frac{(W-k)\varphi_x(W-k-1)}{\varphi_x(W-k)} + k = y.$$

Now let $m^{(k)}$ denote the k^{th} factorial moment. Using the relationship between the finite population probability generating function and the factorial moments leads to the following:

$$\begin{aligned}
\varphi_x(W-k) &= \frac{m^{(k)}}{W^{(k)}} \\
\varphi_x(W-k-1) &= \frac{m^{(k+1)}}{W^{(k+1)}}.
\end{aligned}$$

Now

$$\begin{aligned}
\frac{(W-k)\varphi_x(W-K-1)}{\varphi_x(W-k)}+k &= \frac{(W-k)m^{(k+1)}W^{(k)}}{W^{(k+1)}m^{(k)}}+k \\
&= \frac{(W-k)m^{(k+1)}W!(W-k-1)!}{m^{(k)}(W-k)!W!}+k \\
&= \frac{m^{(k+1)}}{m^{(k)}}+k.
\end{aligned}$$

Solving for Q now involves finding the value of k that solves the following expression:

$$\frac{m^{(k+1)}}{m^{(k)}}+k=y.$$

Thus if higher factorial moments may be approximated, then this may be used to find the value of Q needed for the right tail approximations presented in section 6.3. The other term needed for the approximation is $\varphi_x(Q)$. Recall that this may be found by $\varphi_x(Q) = \varphi_x(W-k) = \frac{m^{(k)}}{W^{(k)}}$.

Other methods for generalizing these approximations is currently being investigated. The need for such methods has not been fully determined for this setting.

Chapter 7

Examples

The goal of the previous chapters has been to develop methods that may be used to improve Poisson approximations. Throughout those chapters, simple examples have been used to illustrate the use and accuracy of the methods. The goal of this chapter is to illustrate the use of the methods for more complicated examples.

7.1 Sum of Bernoulli Random Variables

The sum of Bernoulli random variables with the same probability of success is simply the binomial distribution. This setting has been treated throughout to illustrate the use of the methods. Approximating binomial probabilities is not very useful since the exact probabilities are not difficult to obtain. However approximating probabilities for the sum of Bernoulli random variables with different probabilities of success is much more useful since obtaining exact probabilities is more difficult.

Barbour and Jensen(1989) discuss using the combination of exponential tilting and the Edgeworth-Like expansion to obtain approximations. However in their work they approximate the point probabilities individually and sum them to obtain tail probabilities. Thus their method is tedious and may also be time consuming. The value for the tilting parameter, q, must also be determined numerically.

Here the method of tilting and expansion with a constraint will be used to approximate tail probabilities. Since q must be determined numerically, the method presented in section 6.4.1 will be used to approximate it. Since the exact cumulant generating function is not too difficult, it will not be approximated here. The first four factorial cumulants are as follows:

$$k^{(1)} = \sum_i p_i$$

$$
\begin{aligned}
k^{(2)} &= -(\sum_i p_i^2) \\
k^{(3)} &= 2(\sum_i p_i^3) \\
k^{(4)} &= -6(\sum_i p_i^4).
\end{aligned}
$$

The factorial cumulant generating function is $\sum_i \ln(1 - p_i + qp_i)$. Now consider approximating right tail probabilities for the sum of twenty Bernoulli random variables where $p_i = (.5)(.8)^{i-1}$. The results are presented in table 7.1. In the table Gen. tilt is using the generalized approach for tilting with constraints while Gen. Tilt/Exp. refers to using the generalized approach for tilting and expansion with a constraint.

Table 7.1: Approximating Right Tail Probabilities for a Sum of 20 Bernoulli Random Variables with $p_i = (.5)(.8)^{i-1}$

y	Exact	Poisson	Gen. Tilt	Gen. Tilt/Exp.
1	.9474	.9155 (.034)	.9365 (.012)	.9475 (-.0001)
2	.7573	.7067 (.067)	.7235 (.042)	.7536 (.005)
3	.4651	.4487 (.035)	.4218 (.093)	.4592 (.013)
4	.2107	.2362 (-.121)	.1756 (.167)	.2055 (.025)
5	.0698	.1050 (-.504)	.0514 (.264)	.0665 (.047)
6	.0170	.0401 (-1.358)	.0106 (.376)	.0154 (.094)
7	.00306	.01339 (-3.38)	.00155 (.493)	.00256 (.163)

From table 7.1 it may be seen that the generalized approximations are better than the usual Poisson approximation. In this example the relative errors stay below ten percent until the points move far enough into the tails. Even then there is a large improvement.

7.2 Compound Poisson

Let $Y = \sum_{i=1}^{N} X_i$ where X_i has a Poisson distribution with mean λ and N has a Poisson distribution with mean μ. Then Y has a compound Poisson distribution. The compound

Poisson distribution, and other compound distributions are discussed in Douglas(1980). The density of Y is as follows:

$$f(y) = \sum_{N=1}^{\infty} \left(\frac{e^{-\mu} \mu^N}{N!} \right) \left(\frac{e^{-N\lambda} (N\lambda)^y}{y!} \right) + e^{-\mu} I_{(y=0)}.$$

$I_{(y=0)}$ is an indicator function which equals one if f(0) is being computed, and zero otherwise. Thus it is difficult to obtain point probabilities and even more difficult to obtain tail probabilities.

The probability generating function for a compound Poisson distribution is fairly easy to obtain. It is as follows:

$$\begin{aligned} \pi_y(t) &= \exp[\mu \exp(t\lambda - \lambda) - \mu] \\ &= \pi_N(\pi_{X_i}(t)). \end{aligned}$$

This suggests that the combination of exponential tilting and the expansion method with or without the constraint may be used here. Solving for q does not lead to a simple formula here. One possibility is to use the method suggested in section 6.4. In the examples presented below, q will be solved for numerically.

Another possibility for approximating the distribution is to use the density based approach. All that is required are two or three point probabilities. This is not a simple task since the density does not have a closed form. However they may be obtained by finding convergence in the infinite sum. The density based method is beneficial in that only two or three point probabilities are required as opposed to all of the point probabilities needed for the exact calculation.

Table 7.2 presents the results for approximating the compound Poisson with $\lambda = .2$ and $\mu = 10$ with the combination method of chapter 6 and the density based method of chapter 4. Here the column labels are as before, except now CPoi will be used to refer to the compound Poisson and Constr refers to using tilting and expansion with the constraint.

The table shows results similar to the ones seen previously. The density based methods are very accurate, but they are also the most difficult to perform. The tilting and expansion with the constraint is the simplest method to use, but its accuracy diminishes a little sooner for points in the tails of the distribution.

In the previous example λ was chosen to be less than one. It is well known that this is the setting that the Poisson approximation is most appropriate, see Douglas(1980). Table 7.3 displays the results for approximating a compound Poisson with $\mu = 1$ and $\lambda = 2$.

It can easily be seen from the table that the combination methods break down when λ

Table 7.2: Comparing Density Based Methods and Combination Methods for the Right Tail of a Compound Poisson Distribution with $\lambda = .2$ and $\mu = 10$

y	CPoi.	Poisson	Comb1	Comb2	Constr.	Den0	Den1
4	.159016	.142877 (.101)	.154439 (.029)	.159869 (-.005)	.157216 (.011)	.149383 (.061)	.159327 (-.002)
5	.069101	.052653 (.238)	.066072 (.044)	.069834 (-.011)	.068971 (.002)	.065885 (.047)	.069193 (-.001)
6	.027120	.016564 (.389)	.025508 (.059)	.027591 (-.017)	.027701 (.021)	.026114 (.037)	.027147 (-.001)
7	.009770	.004534 (.536)	.009027 (.076)	.010001 (-.024)	.010349 (-.059)	.009464 (.031)	.009768 (.0002)
8	.003260	.001097 (.663)	.002964 (.091)	.003372 (-.034)	.003633 (-.114)	.003176 (.026)	.003262 (-.0006)
9	.001020	.000237 (.768)	.000911 (.107)	.001065 (-.044)	.001206 (-.182)	.000997 (.023)	.001020 (0.00)
10	.000301	.000046 (.847)	.000026 (.123)	.000318 (-.056)	.000380 (-.262)	.000295 (.020)	.000301 (0.00)

is greater than one. To see why this is the case, consider the factorial cumulants for the compound Poisson:

$$
\begin{aligned}
k^{(1)} &= \mu\lambda \\
k^{(2)} &= \mu\lambda^2 \\
k^{(3)} &= \mu\lambda^3 \\
&\vdots \\
k^{(n)} &= \mu\lambda^n.
\end{aligned}
$$

Thus if λ is greater than one, then the factorial cumulants are increasing instead of decreasing. The combination methods rely on the fact that the factorial cumulants of the distribution being approximated are decreasing. Recall that the factorial cumulants for a Poisson random variable are all zero after the first one. The density based approach still gives at least reasonable estimates. Thus the density based methods may be useful in some situations where a Poisson approximation does not seem reasonable.

The compound Poisson example reveals much about the nature of the approximation methods presented in previous chapters. Other compound distributions are currently being investigated as well as some examples of their uses.

Table 7.3: Comparing Density Based Methods and Combination Methods for the Right Tail of a Compound Poisson Distribution with $\lambda = 2$ and $\mu = 1$

y	CPoi.	Poisson	Comb1	Comb2	Constr	Den0	Den1	Den2
4	.227117	.142877 (.371)	.072019 (.683)	.468602 (-1.06)	.214644 (.055)	.159105 (.299)	.190754 (.160)	.219358 (.034)
6	.095444	.016564 (.826)	-.002965 (1.03)	.280870 (-1.94)	.179694 (-.883)	.073094 (.234)	.090737 (.049)	.098899 (-.036)
8	.036828	.001097 (.970)	-.012353 (1.34)	.139209 (-2.78)	.101274 (-1.75)	.030318 (.177)	.035677 (.031)	.035930 (.024)
10	.013126	.000046 (.996)	-.007832 (1.60)	.060079 (-3.58)	.040654 (-2.10)	.011235 (.137)	.012788 (.026)	.013009 (.009)
12	.004384	$1.36 * 10^{-6}$ (1.00)	-.003605 (1.82)	.023329 (-4.32)	.012687 (-1.89)	.003860 (.120)	.004308 (.017)	.004367 (.004)

7.3 Occupancy Problems

Occupancy problems have been discussed in many places, and thoroughly in Johnson and Kotz(1977). They are examples of distributions associated with finite populations, and thus the methods associated with finite populations will be the main focus of this section.

7.3.1 Classical Occupancy: Number of Empty Boxes

The classical occupancy problem has received a fair amount of attention in various types of literature. Chakraborty(1993) presents an application in genetics, while Mertz and Davies(1968) discusses applications for predator-prey models.

The setting considered is as follows: Suppose there are W boxes lined up in a row. N balls are then thrown at the W boxes. Each ball is equally likely to land in each box. The random variable of interest here is the number of empty boxes after all N balls have been thrown. Computing exact probabilities may be time consuming, especially for tail probabilities. The problem will get worse as N and W get large. The density function is as follows:

$$f(x) = \binom{W}{x} \sum_{j=0}^{W-x} (-1)^j \binom{W-x}{j} \left(1 - \frac{j+x}{W}\right)^N$$

Many variations of the occupancy problem have been discussed. Menon and Prasad (1985) discuss the probability generating function for a variation of the occupancy distribution. The probability generating function for the occupancy distribution may be obtained from their work. Unfortunately it is not simple enough for use in exponential tilting.

The finite population probability generating function for this setting may be found analytically, and is as follows:

$$\begin{aligned}
\varphi_x(Q) &= E\left(\frac{Q^{(W-x)}}{W^{(W-x)}}\right) \\
&= \sum_x \left(\frac{Q!x!}{W!(Q-W+x)!}\right)\binom{W}{x}\sum_{j=0}^{W-x}(-1)^j\binom{W-x}{j}\left(1-\frac{j+x}{W}\right)^N \\
&= \sum_x \binom{Q}{x-W+Q}\sum_{j=0}^{W-x}(-1)^j\binom{W-x}{j}\left(1-\frac{j+x}{W}\right)^N \\
&= \left(\frac{Q}{W}\right)^N\sum_x \binom{Q}{x-W+Q}\sum_{j=0}^{W-x}(-1)^j\binom{W-x}{j}\left(1-\frac{j+x-W+Q}{Q}\right)^N \\
&= \left(\frac{Q}{W}\right)^N.
\end{aligned}$$

Thus the finite population probability generating function again has a very simple form. The value of Q needed to perform tilting in this example must be determined numerically. However, given the simplicity of the generating function, this is not too difficult a task.

Table 7.4 presents the results approximating right tail probabilities for the Occupancy distribution with $W = 200$ and $N = 650$. In the table Tilting refers to tilting using the finite population probability generating function. Tilt/Expand refers to the approximation due to using the tilting and the Edgeworth-like expansion with one correction term.

From table 7.4 it may again be seen that the tilting and expansion combination leads to accurate results. The approximation not only produces accurate results, but does so within a very short period of time.

7.3.2 Classical Occupancy: Number of Boxes with k Balls

A more complicated calculation for the classical occupancy problem is to examine the number of boxes with $k > 0$ balls. Here the joint distribution for the number of boxes the different possible number of balls is known, but the marginal distribution of the number of boxes with a specified k number of balls can not be expressed in simple terms. Let n_j denote the number of boxes with j balls. Then the joint distribution for n_j with $\{j = 0, 1, 2, \cdots, N\}$ is as follows:

$$f(n_0, n_1, \cdots, n_N) = \frac{W!N!}{W^N \prod_{j=0}^{N}(j!)^{n_j} n_j!}.$$

Table 7.4: Approximations to the Right Tail of an Occupancy Distribution with $W = 200$ and $N = 650$

y	Exact	Poisson	Tilting	Tilt/Expand
9	.359682	.364530 (-.013)	.320174 (.11)	.335888 (.07)
10	.230923	.245956 (-.065)	.210855 (.09)	.222419 (.04)
11	.135289	.154751 (-.14)	.126212 (.07)	.133551 (.013)
12	.072365	.090974 (-.26)	.068851 (.05)	.072971 (-.008)
13	.035389	.050094 (-.42)	.031816 (.10)	.034092 (.04)
14	.015853	.025906 (-.63)	.014371 (.09)	.015385 (.03)

The finite population probability generating function for the number of boxes with k balls may be found analytically using the joint the distribution in the following way:

$$\begin{aligned}
\varphi_x(Q) &= E\left(\frac{Q^{(W-x)}}{W^{(W-x)}}\right) \\
&= \sum_{alln_j}\left(\frac{Q!x!}{W!(Q-W+x)!}\right)\frac{W!N!}{W^N\prod_{j=0}^{N}(j!)^{n_j}n_j!} \\
&= \sum_{alln_j}\frac{Q!N!}{W^N\prod_{j=0,j\neq k}^{N}(j!)^{n_j}n_j!(k!)^x(x-W+Q)!} \\
&= \frac{N!Q^{N-k(W-Q)}}{[N-k(W-Q)]!W^N(k!)^{W-Q}}\sum_{alln_j}\frac{Q!(N-k(W-Q))!}{Q^{n-k(W-Q)}\prod_{j=0,j\neq k}^{N}(j!)^{n_j}n_j!(k!)^{x-W+Q}(x-W+Q)!} \\
&= \frac{N!Q^{N-k(W-Q)}}{[N-k(W-Q)]!W^N(k!)^{W-Q}}.
\end{aligned}$$

The value of Q to be used in tilting must be determined numerically. The factorial moments may be obtained easily from the generating function. Table 7.5 displays the results for approximating some right tail probabilities for the distribution of the number of boxes with $k = 1, 2,$and 3 balls given $W = 25$ and $N = 100$. Here the results for using the tilting and expansion with two correction terms is presented as well. The exact values were obtained using the method in equation(3.1).

From table 7.5 it may be seen that the methods produce accurate results. It appears that in

Table 7.5: Approximations to the Right Tail of an Occupancy Distribution with $W = 25$, $N = 100$ and $k = 1, 2,$or 3

y	Exact	Poisson	Tilting	Tilt/Expand1	Tilt/Expand2
k=1					
3	.245461	.257980 (-.051)	.215693 (.121)	.237614 (.032)	.240241 (.021)
4	.074542	.101950 (-.368)	.067023 (.101)	.072230 (.031)	.072907 (.022)
5	.015545	.033401 (-1.15)	.014403 (.073)	.015094 (.029)	.015317 (.015)
k=2					
5	.281554	.298249 (-.059)	.247689 (.120)	.270704 (.039)	.274143 (.026)
7	.041099	.075312 (-.832)	.035274 (.142)	.038309 (.068)	.038763 (.057)
8	.010659	.031842 (-1.99)	.009361 (.122)	.009959 (.066)	.010039 (.058)
k=3					
7	.203893	.228133 (-.119)	.177329 (.130)	.193411 (.051)	.195690 (.040)
9	.037541	.063828 (-.700)	.032420 (.136)	.035350 (.058)	.035769 (.047)
10	.012410	.029474 (-1.375)	.010732 (.135)	.011516 (.072)	.011616 (.064)

most of the cases presented, the tilting and expansion with one correction term is sufficient. There appears to be little gained by using and additional correction term. Perhaps the most appealing feature of the approximations is that they lead to simple and accurate results for a setting where the exact probabilities are difficult to obtain.

7.3.3 Occupancy with Leaks

Consider extending the classical occupancy problem in the following way: there are still N balls and W boxes, once the ball lands in the box, it has probability p of staying in the box, or $1 - p$ of leaking out of the box. Interest now returns to the number of empty boxes. This distribution does not have a probability generating function that may be written in terms of simple functions. The finite population probability generating function is difficult to obtain analytically. However, recall that for the classical occupancy problem:

$$\varphi_x(Q) = \left(\frac{Q}{W}\right)^N .$$

Notice that if we start with W boxes, and mark Q of them as being special, then this is simply the probability that no "nonspecial" box gets a ball. For the current setting of allowing leaks, this would lead to the following:

$$\varphi_x(Q) = \left(1 - \frac{p(W-Q)}{W}\right)^N .$$

Using this to compute factorial moments would lead to the following:

$$E(X^{(j)}) = W^{(j)}\varphi_x(W-j) = W^{(j)}\left(1 - \frac{pj}{W}\right)^N$$

which is the same as was obtained in Johnson and Kotz(1977) through different methods.

The value of Q needed to perform tilting must again be found numerically. However this is not difficult since the generating function has a simple form. Table 7.6 presents the results for approximating the distribution with $W = 200$, $N = 700$, and $p = .9$. The notation in this table is the same as that used in the previous table. From this table it appears that the methods for improving the Poisson approximation are very effective.

7.3.4 Extended Occupancy Problem

The occupancy problem may also be extended in the following way: there are still W boxes, but now there are N sets of m balls that are thrown into the boxes. For each of the N trials, the m balls must go into different boxes. However two or balls from different sets may occupy the same box. This setting is discussed in Johnson and Kotz (1977). The density function is as follows:

$$f(x) = \binom{W}{x} \sum_{j=0}^{W-x} (-1)^j \binom{W-x}{j} \left(\frac{\binom{W-x-j}{m}}{\binom{W}{m}} \right)^N .$$

The finite population probability generating function may be determined analytically, and is as follows:

Table 7.6: Approximations to the Right Tail of an Occupancy Distribution with Leaks with $W = 200,\ N = 650$, and $p = .9$

y	Exact	Poisson	Tilting	Tilt/Expand
10	.341559	.348288 (-.020)	.314502 (.079)	.328860 (.037)
11	.221906	.237711 (-.071)	.209088 (.057)	.219838 (.009)
12	.132592	.152168 (-.148)	.127532 (.038)	.134530 (-.015)
13	.072890	.091505 (-.255)	.065023 (.108)	.069386 (.048)
14	.036906	.051796 (-.403)	.033324 (.097)	.035556 (.037)
15	.017238	.027659 (-.605)	.015739 (.087)	.016783 (.026)
16	.007440	.013966 (-.877)	.006870 (.077)	.007318 (.016)

$$\begin{aligned}
\varphi_x(Q) &= E\left(\frac{Q^{(W-x)}}{W^{(W-x)}}\right) \\
&= \sum_x \left(\frac{Q!x!}{W!(Q-W+x)!}\right)\binom{W}{x}\sum_{j=0}^{W-x}(-1)^j\binom{W-x}{j}\left(\frac{\binom{W-x-j}{m}}{\binom{W}{m}}\right)^N \\
&= \sum_x \binom{Q}{x-W+Q}\sum_{j=0}^{W-x}(-1)^j\binom{W-x}{j}\left(\frac{\binom{W-x-j}{m}}{\binom{W}{m}}\right)^N \\
&= \frac{Q!(W-m)!}{W!(Q-m)!}\sum_x \binom{Q}{x-W+Q}\sum_{j=0}^{W-x}(-1)^j\binom{W-x}{j}\left(\frac{\binom{W-x-j}{m}}{\binom{Q}{m}}\right)^N \\
&= \left(\frac{Q!(W-m)!}{W!(Q-m)!}\right)^N = \left(\frac{Q^{(m)}}{W^{(m)}}\right)^N.
\end{aligned}$$

The value of Q needed for tilting must again be determined numerically. Table 7.7 presents the results for $W = 25$, $N = 10$, and $m = 4, 5,$and 6. It may be seen that the methods for improving Poisson approximations prove to be effective for this setting as well as the others.

Table 7.7: Approximations to the Right Tail of an Extended Occupancy Distribution with $W = 25$, $N = 10$ and $m = 4, 5$, or 6

y	Exact	Poisson	Tilting	Tilt/Expand1
m=4				
6	.218735	.275462 (-.259)	.179312 (.180)	.209169 (.044)
7	.075879	.152973 (-1.02)	.061357 (.191)	.072283 (.047)
8	.018568	.076460 (-3.12)	.015937 (.142)	.018132 (.023)
m=5				
4	.255736	.282462 (-.105)	.222523 (.130)	.249705 (.024)
5	.084890	.134773 (-.588)	.075474 (.111)	.088159 (-.039)
6	.019294	.055483 (-1.88)	.016962 (.121)	.019063 (.012)
7	.002962	.020009 (-5.76)	.002676 (.097)	.002910 (.018)
m=6				
3	.197289	.218510 (-.108)	.174431 (.116)	.192179 (.026)
4	.049851	.079812 (-.601)	.046166 (.074)	.049724 (.003)
5	.008144	.024083 (-1.96)	.008104 (.005)	.008549 (-.049)

Chapter 8

Future Work

There are some unanswered questions and other research ideas that are a result of this work. The goal of this chapter is to present some of the ideas that are currently being investigated, or may be investigated in the future.

8.1 Binomial Approximations

The main focus of this work has been improving Poisson approximations. However there are certain instances where binomial approximations may be more appropriate. Some of the methods for improving Poisson approximations center around the fact that the log of the probability generating function, or the factorial cumulant generating function, is linear for the Poisson distribution. The form of the probability generating function for the binomial distribution is not that simple. However the finite population probability generating function for the binomial distribution is simple. It is derived in the following way:

$$\begin{aligned}
\varphi_x(Q:n) &= E\left(\frac{Q^{(n-x)}}{n^{(n-x)}}\right) \\
&= \sum_x \frac{Q!x!}{n!(Q-n+x)!}\binom{n}{x}p^x(1-p)^{n-x} \\
&= \sum_x \binom{Q}{x-n+Q}p^x(1-p)^{n-x} \\
&= p^{n-Q}\sum_x \binom{Q}{x-n+Q}p^{x-n+Q}(1-p)^{n-x} \\
&= p^{n-Q}.
\end{aligned}$$

Notice that $\ln(\varphi_x(Q:n)) = (n-Q)\ln(p)$, which is linear in Q. Now consider expanding $\ln(\varphi_x(Q:n))$ for any distribution about $Q = n$, using Newton's backward difference formula.

$$\begin{aligned}
\ln(\varphi_x(Q:n)) &\approx (Q-n)\nabla\ln(\varphi_x(n:n)) \\
&= (Q-n)\ln\left(\frac{\varphi_x(n:n)}{\varphi_x(n-1:n)}\right) \\
&= (n-Q)\ln\left(\frac{\varphi_x(n-1:n)}{\varphi_x(n:n)}\right) \\
&= (n-Q)\ln(\varphi_x(n-1:n)).
\end{aligned}$$

This suggests using a binomial approximation with $p = \varphi_x(n-1:n)$. As an example, consider the hypergeometric distribution. The generating function is as follows:

$$\varphi_x(Q:n) = \frac{W!(W+B-n+Q)!}{(W-n+Q)!(W+B)!}.$$

Note that n is being tilted here, as opposed to W. Now notice the following:

$$\varphi_x(n-1:n) = \frac{W!(W+B-n+n-1)!}{(W-n+n-1)!(W+B)!} = \frac{W}{W+B}.$$

This suggests approximating a hypergeometric with a binomial with $p = \frac{W}{W+B}$, which is the usual form of the approximation.

Since the natural log of the generating function is linear for a binomial, it seems possible that methods analogous to the ones used to improve Poisson approximations, may also be used to improve binomial approximations. As a simple illustration, consider tilting with constraints. The idea here will be to let $L(Q)$ denote an approximation to $\ln(\varphi_x(Q:n))$. Then the two constraints will be as follows:

1. $L(n) = \ln(\varphi_x(n:n)) = 0$.
2. $L(\hat{Q}) = \ln(\varphi_x(\hat{Q}:n))$

where $\hat{Q}$ denotes the solution to the following:

$$\frac{Q\varphi_x(Q-1:n)}{\varphi_x(Q)} = y-n+Q.$$

It may be easily verified that $L(Q) = \frac{(Q-n)\ln(\varphi_x(\hat{Q}:n))}{(\hat{Q}-n)} = \frac{(n-Q)[-\ln(\varphi_x(\hat{Q}:n))]}{(\hat{Q}-n)}$. This suggests using a binomial approximation with $p = \exp\left[\frac{-\ln(\varphi_x(\hat{Q}:n))}{(\hat{Q}-n)}\right]$.

Table 8.1 presents some results for approximating point probabilities for a hypergeometric with $W = 20$, $B = 22$, and $n = 10$. The table presents the exact probabilities, the results for the usual binomial approximation, and the results for tilting with the constraints. The results from the table seem to suggest that methods analogous to the ones used to improve Poisson approximations may also be used to improve binomial approximations.

Table 8.1: Approximating Point Probabilities for a Hypergeometric with $W = 20$, $B = 22$, and $n = 10$

y	Exact	Binomial	Tilt w/Const.
6	.192688	.184330 (.043)	.195228 (-.013)
7	.081132	.095756 (-.18)	.084819 (-.045)
8	.019776	.032644 (-.651)	.020213 (-.022)
9	.002511	.006595 (-1.626)	.002608 (-.039)

Very little has been done on improving binomial probabilities. This work is currently being investigated. Improving negative binomial approximations is also currently being pursued, since the methods or techniques will probably be very similar to the ones needed for binomial approximations.

8.2 Categorical Data Analysis

8.2.1 Introduction

Statisticians have become interested in exact procedures, or at least better approximate procedures for the analysis of categorical data. There has been, and continues to be some concern about the usual chi-square procedures. Haberman(1988) gives a warning about the use of chi-squared statistics when the expected cell counts are small. This paper shows that under some circumstances it is possible to have a test in which the power of the test is less than the size. Haldane(1940) derives the mean and variance of the chi-square test statistic when sample sizes are small. Haberman(1977) discusses situations in which chi-square procedures are still valid, despite small sample expected cell counts. Cox and Plackett(1980) discusses some methods for improving the asymptotic procedures. Among the methods they propose is enumerating all the tables with the observed marginal totals. They also propose a simulated random sampling procedure for cases where exact enumeration is impossible. Gail and Mantel(1977) also discusses exact enumeration. They also discuss a normal approximation to the number of tables. However, their work only involves two-way tables. Stumpf and Steyn(1986) discuss the exact distributions associated with three-way tables. Agresti(1992) surveys methods for exact inference in contingency tables.

All of the papers mentioned above involve improving or developing a test for the fit of a particular model. Very little attention has been given to cell residuals. Cell residuals are used for similar reasons as residuals in other forms of analysis. Cell residuals are useful for determining whether or not a cell may be labeled as an outlier. It is possible for omnibus tests to yield significant results because of a small subset of outlier cells. Cell residuals are often useful for suggesting another model when one fails to fit well. If there is some pattern exhibited by the cells that don't fit well, then this may suggest an additional term for the model. Thus it appears that accurate methods are needed for determining how unusual a cell is for a given model.

The next section will present methods that are currently available for analyzing cell residuals. Section 8.2.3 will propose using the methods developed in the previous chapters. Much of the material for this chapter is currently being developed. The work here will also be based on conditional inference. For a discussion of conditional inference and it's advantages, see Agresti(1992) and Reid(1995).

8.2.2 Current Methods

Much of the information presented here may be found in Agresti(1990). In contingency tables, the residuals are usually defined as $n_i - \hat{m}_i$. These residuals are not very useful since a difference of fixed size is more important for smaller samples. What is often used are the standardized residuals. The standardized residuals are defined as:

$$e_i = \frac{n_i - \hat{m}_i}{\hat{m}_i^{1/2}}.$$

These residuals are important since the sum of these squared makes up the Pearson chi-squared test statistic. So analyzing these is essentially determining which cells contribute the most to the chi-square test statistic. The standardized residuals are asymptotically normal with mean 0. Agresti(1990) argues that comparing these to the standard normal distribution often leads a conservative indication of lack of fit.

The adjusted residuals are the standardized residuals divided by their estimated standard errors. These residuals take simple forms for two-way tables and some three-way tables. Fuchs and Kenett(1980) suggest using these residuals to form a test for outlier cells. Haberman(1973) suggests plotting these to highlight unusual cells. The problem with these residuals is that their form is no longer simple for more complicated models. Also, their distribution is only asymptotically normal.

Another method for analyzing individual cells involves cell deletion. The idea here is to remove the cell from the table. Re-fit a model to the remaining cells. Then the effect of removing the cell is measured using something like the Pearson's chi-square or the generalized likelihood ratio test statistic. The difference between the statistic for all cells and the statistic computed after the cell is removed is usually compared to a chi-square with one degree of freedom. Simonoff(1988) suggests using a backwards stepping procedure with the deleted residuals to determine which cells are outliers.

Stumpf and Steyn(1986) shows that the conditional distribution of a cell is asymptotically Poisson. This suggests comparing the cell to a Poisson distribution as opposed to the normal distribution. However, conditioning on the marginal totals leads to simple forms for the factorial moments. These may be used to obtain the conditional mean and variance of the cell. The cell may be standardized using this mean and variance, and then compared to a standard normal distribution. These methods are asymptotic, and only valid if the marginal totals corresponding to the cell are large.

Another approach is to enumerate all the possible tables to obtain information about the cells as well as omnibus tests. The problem with this approach is that it is only feasible for small tables. Even though there has been remarkable advances in computing power recently, it is still not enough to handle most situations.

8.2.3 Proposed Methods

Most of the methods currently being used tend to have problems if all or some of the marginal totals are not large. The distributions associated with the residuals are all asymptotic. However, a method that does not require asymptotic techniques for obtaining the distribution or tail probabilities for the cells or residuals is often needed. The methods proposed here will be for the conditional distributions of the cells. The discussion will first center around the cases that have relatively simple distributions. Then some comments will be made regarding the case where the distributions are not simple. These are the cases where a technique such as iterative proportional fitting is needed to obtain maximum likelihood estimates for the cells.

For the cases where the distribution takes a relatively simple form, the methods presented in section 6.3 should be applicable, since it is possible to obtain the generating function $\varphi_x(Q)$. Once this is obtained, it may be used to perform tilting for the distribution of a cell. Since the factorial moments for the cell may be obtained using the conditional distribution, then it seems reasonable to suggest combining the tilting method with the expansion method to obtain accurate results.

The cases where the distribution does not take a simple form appears to involve a little more effort. The methods being proposed here will involve some algorithm like iterative proportional fitting. Some work will need to be done to determine the accuracy of this technique for obtaining estimates of the mean. Factorial moments may also be estimated using this technique. Notice that for the cases where the distribution takes a simple form, the factorial moments may be computed in the following way:

1. Compute the mean for a given cell.
2. Reduce the observed marginal totals corresponding to the cell by one.
3. Compute the mean for the cell in the adjusted table.
4. Return to step two and continue if necessary.

Once all the necessary means are computed, then they are multiplied together to obtain the estimate for the factorial moments. The procedure above works for the cases where the distribution takes a simple form. It seems reasonable to use this in the cases where the distribution does not take a simple form. Only now the mean at each step would be computed using iterative proportional fitting. Terrell(Personal Communication, 1997) gives a probability argument for the procedure above. Some work needs to be done to verify that this will give accurate estimates. Assuming that it does, this will enable the use of the expansion method as well as computing probabilities from factorial moments.

A reasonable approach to tilting may be possible using the methods of section 6.4.3. The idea here is to consider $Q = W - k$ where k is a positive integer. Find k through trial and error. That is, start subtracting numbers from the cell of interest until $Q = W - k$ is found such that the mean of the cell for the new table plus k is equal to the value of interest, say y, or at least as close to y as possible. Once Q is found, $\varphi_x(Q)$ may be computed using the following relationship:

$$m^{(k)} = w^{(k)}\varphi_x(W - k).$$

Once these pieces have been computed, then the tilting may be performed. Thus the methods proposed in this work should apply to all models and all dimensions.

8.3 Other Future Work

Other areas of future work include error analysis. Very little has been done as far as developing the theory necessary for error analysis in cases involving Poisson or normal approximations. An upper bound for the relative errors would be very useful. The issue is currently dealt with by examining the order of the error. However this is inadequate since the approximations often perform well, even when the arguments based on order would suggest that the methods would not work well. The classic examples of this are based on normal approximations where the order is based on the sample size. The methods often continue to perform well, even when the sample size is one.

Another area of future work is to continue to examine other methods of improving approximations. Although many techniques have been presented in this work, there still may be others.

Chapter 9

Conclusion

The Poisson distribution has been studied by many statisticians. It is the limiting distribution for many other distributions, thus suggesting that it may be used to approximate them. The most appealing feature of Poisson approximations is their simplicity, all that is required is the mean of the distribution being approximated. Unfortunately Poisson approximations often prove to be inadequate. However, it has been shown in this work that if some more information is available, then it is possible to improve the usual Poisson approximations.

Many distributions have been approximated throughout this work. For all of these examples it has been possible, though often difficult, to obtain exact probabilities. In practice, these methods are probably most useful in cases where it is too difficult to obtain exact probabilities. In some cases it may be possible to work with an easier version of the distribution to determine which technique is the most accurate. This was the approach commonly used for the examples in this work. Unfortunately this is not always possible or even desirable.

When determining which method to use, the first step should be to determine what information is available. If the probability generating function and moment generating function are available, then some form of tilting and expansion should be used. If it is not obvious which distribution to use, then a method similar to that of Jin and Robinson (1999) may be used. That is, form the conjugate distribution: if it is best approximated by a Poisson distribution, then use the methods presented in this work. If it is best approximated by a normal distribution, then the methods of Daniels (1987) should be used. If other continuous distributions are more appropriate, then the methods of Wood, Booth, and Butler (1993) should be used. Using other discrete distributions, like the binomial or negative binomial is the subject of future research. One method for determining whether or not the Poisson is appropriate is to examine the factorial cumulants. If they are small and decreasing, then the methods of this work should be appropriate. An improved Poisson approximation should be used if approximating a non-negative integer valued distribution when appropriate since it does not involve a continuity correction and it uses the factorial cumulants as opposed to the usual cumulants. The factorial cumulants are slightly easier to obtain. In the cases where the Poisson approximation is appropriate, then the tilting and expansion with the constraint is probably the easiest method to use.

If the distribution being approximated is associated with a finite population, then the tilting and expansion using the finite population probability generating function may be the only option. Note that often in this case there is no known way of improving the normal approximations. For most approximations, using this technique with only one correction term in the expansion will be sufficient. If a Poisson approximation seems to be less appropriate, then a second correction term may be needed. For cases where the Poisson approximation is inappropriate is an area for possible future work. It is possible that an improved binomial approximation may be useful here.

If generating functions are too difficult to work with, then the methods of section 6.4 should be used. Analogous methods for normal approximations have been studied by Wang (1992) and Terrell (personal communication). Determining which distribution to use is more difficult here since the conjugate densities resulting from tilting would be more difficult to assess. Here the distribution most appropriate for the original distribution is probably the one that should be used.

If the generating functions and moments are difficult to obtain, but the density is obtainable and has a shape similar to that of a Poisson distribution, then the density based approach should be used. This method is most appropriate for cases where the density is known, but taking the sum is difficult or time consuming. Density based approaches based on other distributions is another area of possible future research.

For most non-negative integer valued distributions, at least some of the information needed to use these methods is obtainable. How to handle cases where none of the information is available is currently unknown. However, as has been shown in this work, if some of the information is available, it may be used to lead to drastic improvements over the usual Poisson approximations.

Appendix A

Mathematical Preliminaries

This paper presents results that were derived using techniques that may not be familiar to many statisticians. Thus some of these methods are presented here to assist in the understanding of the derivations in this work.

A.1 Backward and Forward Differencing

Backward differencing is analogous to differentiating in the continuous setting. The first backward difference is defined as $\nabla p(x) = p(x) - p(x-1)$. The k^{th} backward difference may be computed as follows:

$$\nabla^k p(x) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} p(x-j). \tag{A.1}$$

These may be computed recursively as $\nabla^k p(x) = \nabla(\nabla^{k-1} p(x))$. A fact regarding backward differencing that proves to be useful is the following:

$$\sum_{x=y}^{n} \nabla^k p(x) = \nabla^{k-1} p(n) - \nabla^{k-1} p(y-1). \tag{A.2}$$

Forward differencing is very similar to backward differencing. Forward differencing is discussed in Davis and Rabinowitz(1975). The first forward difference is defined as $\triangle p(x) = p(x+1) - p(x)$. The k^{th} forward difference may be computed as follows:

$$\triangle^k p(x) = \sum_{j=0}^{k} (-1)^{k-j} \binom{k}{j} p(x+j). \tag{A.3}$$

These may be computed recursively as $\triangle^k p(x) = \triangle(\triangle^{k-1} p(x))$

A.2 Summation by Parts

Summation by parts is analogous to integration by parts. It is presented briefly in Goldberg (1976). Let $\{a_n\}_{n=1}^{\infty}$ and $\{b_n\}_{n=1}^{\infty}$ be two sequences. Let $S_n = \sum_1^n a_n$. Then, for integers n and m:

$$\sum_{k=m}^{n} a_k b_k = S_n b_n - S_{m-1} b_m + \sum_{j=m}^{n-1} S_j (b_j - b_{j+1}). \tag{A.4}$$

Another way of presenting summation by parts is to let $S_n = \sum_{j=n}^{\infty}$. Then, for each integer m:

$$\sum_{k=m}^{\infty} a_k b_k = S_m b_m + \sum_{j=m+1}^{\infty} S_j (b_j - b_{j-1}). \tag{A.5}$$

This form for summation by parts is useful for the derivation of the density based tail probability approximation. For a proof of summation by parts, see Goldberg (1976).

A.3 Newton's Backward Difference Formula and Forward Difference Formula

Newton's backward difference formula is analogous to a Taylor Series. The Taylor series is a very useful tool when a complicated function needs to be approximated, especially if the integral of the complicated function is needed. Likewise, Newton's backward difference formula is useful when the sum of a complicated function is needed. This is used in the derivation of the density based tail probability approximation. To present Newton's backward difference formula, consider a function $h(x)$ that needs to be expanded around some point y. Then the expansion is as follows:

$$\begin{aligned} h(x) = h(y) + (x-y)\nabla h(y) + \binom{x-y+1}{2} \nabla^2 h(y) + \cdots \\ + \binom{x-y+n-1}{n} \nabla^n h(y) + \cdots . \end{aligned} \tag{A.6}$$

The derivation of Newton's backward difference formula is similar to that of the Taylor series. Begin with $h(x) = h(y) + \sum_{t=y+1}^{x} \nabla h(t)$, and then proceed by repeatedly using summation by parts.

The forward difference formula is discussed in Davis and Rabinowitz(1975). The formula will be expressed as expanding a function about zero since that is how it is used in this work. The expansion is as follows:

$$f(x) = \sum_{k=0}^{\infty} \frac{\triangle^k f(0) x^{(k)}}{k!}. \tag{A.7}$$

Appendix B

Derivations

B.1 Derivation of Tail probabilities for Exponential Tilting / Expansion Combination

A detailed derivation for the right tail will be given here. Then the modifications needed for deriving the left tail formula will be presented.

B.1.1 Right Tail Approximation

Start with equation (6.1):

$$f(x) = \pi_x(q)q^{-x}\left(p(x) + \frac{k_q^{(2)}}{2}\nabla^2 p(x) + \left[-\frac{k_q^{(3)}}{6}\nabla^3 p(x) + \frac{(k_q^{(2)})^2}{8}\nabla^4 p(x)\right] + \cdots\right).$$

Then sum both sides of the equation for x going from y to ∞. The first term on the right hand side is just the result from exponential tilting:

$$\sum_{x=y}^{\infty} q^{-x}p(x) = e^{-y(1-1/q)}R^*_{y/q}(x).$$

$\sum_{x=y}^{\infty} q^{-x}\nabla^2 p(x)$ can be computed using summation by parts. The result after performing summation by parts one time is as follows:

$$\sum_{x=y}^{\infty} q^{-x}\nabla^2 p(x) = -q^{-y}\nabla p(y-1) + \left(\frac{q-1}{q}\right)\sum_{x=y}^{\infty} q^{-x}\nabla p(x).$$

Perform summation by parts again on the last term to obtain:

$$\left(\frac{q-1}{q}\right)\sum_{x=y}^{\infty} q^{-x}\nabla p(x) = -q^{-y}\left(\frac{q-1}{q}\right)p(y-1) + \left(\frac{q-1}{q}\right)^2\sum_{x=y}^{\infty} q^{-x}p(x).$$

The last term in this expression is exactly the same as what was obtained in exponential tilting alone, and it was obtained above. If the first correction term was the only one of interest, then this information would be sufficient. To obtain more terms, notice that in general:

$$\sum_{x=y}^{\infty} q^{-x}\nabla^k p(x) = -q^{-y}\nabla^{k-1}p(y-1) + \left(\frac{q-1}{q}\right)\sum_{x=y}^{\infty} q^{-x}\nabla^{k-1}p(x).$$

Continue inductively to obtain:

$$\begin{aligned}\sum_{x=y}^{\infty} q^{-x}\nabla^k p(x) &= -q^{-y}\Bigg(\nabla^{k-1}p(y-1) + \left(\frac{q-1}{q}\right)\nabla^{k-2}p(y-1) + \\ &\left(\frac{q-1}{q}\right)^2\nabla^{k-3}p(y-1) + \cdots + \\ &\left(\frac{q-1}{q}\right)^{k-1}p(y-1)\Bigg) + \left(\frac{q-1}{q}\right)^k e^{-y(1-1/q)}R^*_{y/q}(x).\end{aligned}$$

Now consider the following observation:

$$\begin{aligned} e^{y(1-1/q)}(1/q)^y &= e^y e^{-y/q}(1/q)^y y^y y^{-y} = e^y e^{-y/q}(y/q)^y y^{-y} \\ &= e^y y^{-y} e^{-y/q}(y/q)^y 1/y!y! = ((y!)y^{-y}e^y)P_{y/q}(y)\end{aligned}$$

where $P_{y/q}(y)$ is the probability that a Poisson random variable with mean y/q is equal to y. Now define $L^*_k(q)$ in the following way:

$$L^*_k(q) = \left(\frac{q-1}{q}\right)^k R^*_{y/q}(y) - \left(\frac{y!e^y}{y^y}\right)P_{y/q}(y)[\nabla^{k-1}p(y-1) + \left(\frac{q-1}{q}\right)\nabla^{k-2}p(y-1) + \cdots + \left(\frac{q-1}{q}\right)^{k-1}p(y-1)].$$

This leads to:

$$P(X \geq y) = \pi_x(q)e^{-(y-y/q)}(R^*_{y/q}(y) + \frac{k_q^{(2)}}{2}L_2^*(q) - \frac{k_q^{(3)}}{6}L_3^*(q) + \frac{(k_q^{(2)})^2}{8}L_4^*(q) + \cdots).$$

B.1.2 Adjustments to Approximate Left Tail Probabilities

The derivation of the approximation for the left tail probability requires a few minor adjustments to the previous section. The formula for $\sum_{x=0}^{y} q^{-x}\nabla^k p(x)$ becomes:

$$\sum_{x=0}^{y} q^{-x}\nabla^k p(x) = q^{-y-1}\nabla^{k-1}p(y) + \left(\frac{q-1}{q}\right)\sum_{x=0}^{y} q^{-x}\nabla^{k-1}p(x).$$

Continue inductively to obtain:

$$\begin{aligned}\sum_{x=0}^{y} q^{-x}\nabla^k p(x) \quad = \quad & q^{-y-1}\left(\nabla^{k-1}p(y) + \left(\frac{q-1}{q}\right)\nabla^{k-2}p(y) + \right. \\ & \left(\frac{q-1}{q}\right)^2 \nabla^{k-3}p(y) + \cdots + \\ & \left.\left(\frac{q-1}{q}\right)^{k-1} p(y)\right) + \left(\frac{q-1}{q}\right)^k e^{-y(1-1/q)}R_{y/q}(x).\end{aligned}$$

Using these minor adjustments and proceeding as above leads to the results of equation (6.2).

B.2 Density Based Tail Probability Approximation

A more detailed derivation of the formulas in chapter 4.3 will be given here. The detailed derivation of the right tail probability approximation formula will be given first. Then details regarding the adjustments for the left tail probability will be given.

B.2.1 Right Tail Approximation

First consider the derivation of the L functions given in equation (4.9). Define $L_k^*(y)$ to be $\sum_{x=y}^{\infty}(x-y)^k p(x)$ where $p(x)$ is the density function for a Poisson random variable. Consider first $L_1^*(y)$:

$$
\begin{aligned}
L_1^*(y) &= \sum_{x=y}^{\infty}(x-y)p(x) = \sum_{x=y}^{\infty}[(x-\lambda)-(y-\lambda)]p(x) \\
&= -zR_\lambda^*(y) + \sum_{x=y}^{\infty}(x-\lambda)p(x) \\
&= -zR_\lambda^*(y) + \sum_{x=y}^{\infty}\lambda[p(x-1)-p(x)] \\
&= -zR_\lambda^*(y) + \lambda p(y-1)
\end{aligned}
$$

where $z = y - \lambda$. Now consider $L_2^*(y)$, for simplicity, let $z^* = x - \lambda$:

$$
\begin{aligned}
L_2^*(y) &= \sum_{x=y}^{\infty}(x-y)^2 p(x) = \sum_{x=y}^{\infty}(z^*-z)^2 p(x) \\
&= \sum_{x=y}^{\infty}(z^*-z)(z^*-z)p(x) = -zL_1^*(y) + \sum_{x=y}^{\infty}(z^*-z)zp(x).
\end{aligned}
$$

Now use summation by parts on the last term. The form of summation by parts used here is that of equation (A.5). To match the formula, let $a_x = z^*p(x)$ and $b_x = (z^* - z) = (x - y)$. Notice that $b_y = (y - y) = 0$. So the last term becomes:

$$
\begin{aligned}
\sum_{x=y}^{\infty}(z^*-z)z^*p(x) &= \sum_{x=y+1}^{\infty}\lambda p(x-1)[(x-y)-(x-1-y)] \\
&= \sum_{x=y}^{\infty}\lambda p(x)[(x+1-y)-(x-y)] \\
&= \lambda R_\lambda^*(y).
\end{aligned}
$$

Putting this together with the above terms yields $L_2^*(y) = \lambda R_\lambda^*(y) - zL_1^*(y)$. Now consider $L_3^*(y)$:

$$
\begin{aligned}
L_3^*(y) &= \sum_{x=y}^{\infty}(x-y)^3 p(x) = \sum_{x=y}^{\infty}(z^*-z)^3 p(x) \\
&= \sum_{x=y}^{\infty}(z^*-z)(z^*-z)^2 p(x) = -zL_2^*(y) + \sum_{x=y}^{\infty}(z^*-z)^2 z^* p(x).
\end{aligned}
$$

Now proceed as above by using summation by parts. The last part of this equation becomes:

$$\begin{aligned}
\sum_{x=y}^{\infty}(z^*-z)^2 z^* p(x) &= \sum_{x=y+1}^{\infty}\lambda p(x-1)[(x-y)^2-(x-1-y)^2] \\
&= \sum_{x=y}^{\infty}\lambda p(x)[(x+1-y)^2-(x-y)^2] \\
&= \sum_{x=y}^{\infty}\lambda p(x)[2(x-y)+1] \\
&= 2\lambda L_1^*(y)+\lambda R_\lambda^*(y).
\end{aligned}$$

Putting all the terms together yields $L_3^*(y) = -zL_2^*(y) + 2\lambda L_1^*(y) + \lambda R_\lambda^*(y)$. Using the same techniques, it can be shown that $L_4^*(y) = -zL_3^*(y) + 3\lambda L_2^*(y) + 3\lambda L_1^*(y) + \lambda R_\lambda^*(y)$. Other L^* functions can be derived recursively using the previous L^* functions and the binomial coefficients. These L^* functions will be useful later.

The form of the approximation being considered here is $\sum_{x=y}^{\infty} f(x) = \sum_{x=y}^{\infty} p(x)g(x)$. Now let $g(x) = e^{h(x)}$. This will prevent g(x) from taking negative values. So $f(x) = p(x)e^{h(x)}$. This implies that $h(x) = \ln(f(x)) - \ln(p(x))$. Consider taking the backward difference of $h(x)$. This is necessary in order to expand it using Newton's Backward difference formula. The first backward difference of $h(x)$ is as follows:

$$\begin{aligned}
\nabla h(x) &= h(x)-h(x-1) \\
&= \ln(f(x))-\ln(p(x))-\ln(f(x-1))+\ln(p(x-1)) \\
&= \nabla\ln(f(x))-\nabla\ln(p(x)).
\end{aligned}$$

Now $-\nabla\ln(p(x))$ can be simplified as follows:

$$\begin{aligned}
-\nabla\ln(p(x)) &= \ln(p(x-1))-\ln(p(x)) \\
&= \ln\left(\frac{p(x-1)}{p(x)}\right) \\
&= \ln\left(\frac{x}{\lambda}\right) \\
&= \ln(x)-\ln(\lambda).
\end{aligned}$$

Now the first backward difference of h(x) becomes $\nabla h(x) = \nabla\ln(f(x)) + \ln(x) - \ln(\lambda)$. The second backward difference can be found by taking the first backward difference of the first backward difference. This yields $\nabla^2 h(x) = \nabla^2\ln(f(x)) + \nabla\ln(x)$. In general, the k^{th} backward difference of h(x) is $\nabla^k h(x) = \nabla^k\ln(f(x)) + \nabla^{k-1}\ln(x)$.

Notice that λ can be chosen to make $\nabla h(y)$ equal to zero. Consider the following:

$$\nabla h(y) = \ln(f(y)) - \ln(f(y-1)) + \ln(y) - \ln(\lambda) = 0.$$

Solving for $\ln(\lambda)$ yields:

$$\ln(\lambda) = \ln\left(\frac{yf(y)}{f(y-1)}\right).$$

So λ becomes:

$$\lambda = \frac{yf(y)}{f(y-1)}.$$

Working with this value of λ will lead to a simpler version of the tail probability approximation.

Now back to $f(x) = p(x)e^{h(x)}$. Expanding $h(x)$ about the point $x = y$ using Newton's backward difference formula leads to:

$$f(x) = p(x)\exp\left[h(y) + (x-y)\nabla h(y) + \begin{pmatrix} x-y+1 \\ 2 \end{pmatrix} \nabla^2 h(y) + \cdots\right].$$

Now expand the exponential in the usual power series. Recall that $\nabla h(y)$ is now equal to zero. The resulting form is as follows:

$$\begin{aligned} f(x) &= p(x)g(y)\left(\left[1 + \begin{pmatrix} x-y+1 \\ 2 \end{pmatrix} \nabla^2 h(y) + \begin{pmatrix} x-y+2 \\ 3 \end{pmatrix} \nabla^3 h(y) + \cdots\right]\right. \\ &\left. +1/2\left[\begin{pmatrix} x-y+1 \\ 2 \end{pmatrix}^2 (\nabla^2 h(y))^2 + \cdots\right] + \cdots\right). \end{aligned}$$

The issue of grouping the terms will be addressed in the next section. The terms needed for the approximation with two correction terms are given above. That is, the largest correction terms are presented in the previous equation.

In order to obtain the approximation to the tail probability, both sides of the previous equation need to be summed over the appropriate region. The expression will be summed term by term. This yields the following:

$$\sum_{x=y}^{\infty} f(x) = g(y)\left[R(y) + \nabla^2 h(y) \sum_{x=y}^{\infty} \binom{x-y+1}{2} p(x) + \left[\nabla^3 h(y) \sum_{x=y}^{\infty} \binom{x-y+2}{3} p(x) + (1/2)(\nabla^2 h(y))^2 \sum_{x=y}^{\infty} \binom{x-y+1}{2}^2 p(x)\right] + \cdots\right.$$

Consider part of the first correction term:

$$\begin{aligned}\sum_{x=y}^{\infty} \binom{x-y+1}{2} p(x) &= (1/2) \sum_{x=y}^{\infty} (x-y)(x-y+1)p(x) \\ &= (1/2)\Big(\sum_{x=y}^{\infty} (x-y)^2 p(x) + \sum_{x=y}^{\infty} (x-y)p(x)\Big) \\ &= (1/2)(L_2^*(y) + L_1^*(y)).\end{aligned}$$

Similarly it can be shown that:

$$\begin{aligned}\sum_{x=y}^{\infty} \binom{x-y+2}{3} p(x) &= (1/6)(L_3^*(y) + 3L_2^*(y) + 2L_1^*(y)) \\ \sum_{x=y}^{\infty} \binom{x-y+1}{2}^2 p(x) &= (1/4)(L_4^*(y) + 2L_3^*(y) + L_2^*(y)).\end{aligned}$$

Using this information and making a substitution for the backward differences of h(y) leads to the final approximation. The final form of the approximation with two correction terms is the following:

$$\begin{aligned}P(X \geq y) &= (f(y)/p(y))[R_\lambda^*(y) + (1/2)(\nabla^2 \ln(f(y)) + \nabla \ln(y))(L_2^*(y) + L_1^*(y)) \\ &+ [(1/6)(\nabla^3 \ln(f(y)) + \nabla^2 \ln(y))(L_3^*(y) + 3L_2^*(y) + 2L_1^*(y)) \\ &+ (1/8)(\nabla^2 \ln(f(y)) + \nabla \ln(y))^2 (L_4^*(y) + 2L_3^*(y) + L_2^*(y))] \\ &+ \cdots].\end{aligned}$$

B.2.2 Adjustments for Approximating Left Tail Probabilities

The only adjustments needed to approximate left tail probabilities involve changing the L^* functions. Let $L_k(y) = \sum_{x=0}^{y} (x-y)^k p(x)$. First consider $L_1(y)$:

$$
\begin{aligned}
L_1(y) &= \sum_{x=0}^{y}(x-y)p(x) = \sum_{x=0}^{y}[(x-\lambda)-(y-\lambda)]p(x) \\
&= -zR_\lambda(y) + \sum_{x=0}^{y}(x-\lambda)p(x) \\
&= -zR_\lambda(y) + \sum_{x=0}^{y}\lambda[p(x-1)-p(x)] \\
&= -zR_\lambda(y) - \lambda p(y)
\end{aligned}
$$

Now consider $L_2(y)$:

$$
\begin{aligned}
L_2(y) &= \sum_{x=0}^{y}(x-y)^2 p(x) = \sum_{x=0}^{y}(z^*-z)^2 p(x) \\
&= \sum_{x=0}^{y}(z^*-z)(z^*-z)p(x) = -zL_1(y) + \sum_{x=0}^{y}(z^*-z)zp(x).
\end{aligned}
$$

Now use summation by parts on the last term. The form of summation by parts used here is that of equation (A.4). To match the formula, let $a_x = z^* p(x)$ and $b_x = (z^*-z) = (x-y)$. So the last term becomes:

$$
\begin{aligned}
\sum_{x=0}^{y}(z^*-z)z^* p(x) &= \sum_{x=0}^{y-1} -\lambda p(x)[(x-y)-(x+1-y)] \\
&= \lambda R_\lambda(y-1).
\end{aligned}
$$

Thus $L_2(y) = \lambda R(y-1) - zL_1(y)$. Before computing $L_3(y)$, define $L_1(y-1)$ to be equal to $\sum_{x=0}^{y-1}(x-y)p(x)$. It is easy to show that $L_1(y-1) = -\lambda p(y-1) - zR_\lambda(y-1)$. Now consider $L_3(y)$:

$$
\begin{aligned}
L_3(y) &= \sum_{x=0}^{y}(x-y)^3 p(x) = \sum_{x=0}^{y}(z^*-z)^3 p(x) \\
&= \sum_{x=0}^{y}(z^*-z)(z^*-z)^2)p(x) = -zL_2(y) + \sum_{x=0}^{y}(z^*-z)^2 zp(x).
\end{aligned}
$$

Now proceed as above by using summation by parts. The last part of this equation becomes:

$$\begin{aligned}\sum_{x=0}^{y}(z^*-z)^2 z^* p(x) &= \sum_{x=0}^{y-1} -\lambda p(x)[(x-y)^2-(x+1-y)^2] \\ &= \sum_{x=0}^{y-1} \lambda p(x)[(x+1-y)^2-(x-y)^2] \\ &= \sum_{x=0}^{y-1} \lambda p(x)[2(x-y)+1] \\ &= 2\lambda L_1(y-1)+\lambda R_\lambda(y-1).\end{aligned}$$

Thus $L_3(y) = -zL_2(y) + 2\lambda L_1(y-1) + \lambda R_\lambda(y-1)$. $L_4(y)$ can be derived in a similar fashion. The end result is $L_4(y) = -zL_3(y) + 3\lambda L_2(y-1) + 3\lambda L_1(y-1) + \lambda R(y-1)$, where $L_2(y-1) = \lambda R(y-1) - zL_1(y-1)$. Replacing the L^* functions in the right tail approximation with the respective L functions will produce the left tail approximation.

$$\begin{aligned}P(X \le y) &= (f(y)/p(y))[R_\lambda(y) + (1/2)(\nabla^2 \ln(f(y)) + \nabla \ln(y))(L_2(y) + L_1(y)) \\ &+ [(1/6)(\nabla^3 \ln(f(y)) + \nabla^2 \ln(y))(L_3(y) + 3L_2(y) + 2L_1(y)) \\ &+ (1/8)(\nabla^2 \ln(f(y)) + \nabla \ln(y))^2(L_4(y) + 2L_3(y) + L_2(y))] \\ &+ \cdots].\end{aligned}$$

B.2.3 Size of the Correction Terms

The size of the correction terms is used to determine which terms should be grouped together. It also determines which terms are added or left out. To determine the size or order of the terms, the case of approximating the binomial distribution will be considered here. In the tilting / expansion approximation, the cumulants were used to determine the size of the terms. The analogous thing here is to examine the size of the $\nabla^k h(y)$ or the $\nabla^k \ln(f(y)) + \nabla^{k-1} \ln(y)$ terms. For the binomial distribution, $\nabla h(y) = \ln(n-y+1) + \ln(p/(1-p)) - \ln(\lambda)$.

Now consider $\nabla^2 h(y)$:

$$\begin{aligned}\nabla^2 h(y) &= \nabla \ln(n-y+1) = \ln\left(\frac{n-x+1}{n-x+2}\right) \\ &= \ln\left(1 - \frac{1}{n-x+2}\right).\end{aligned}$$

If this expanded in a Taylor Series, then it is easy to see that $O(\nabla^2 h(y)) = 1/(n-x)$. Now consider $\nabla^3 h(y)$ and $\nabla^4 h(y)$:

$$\begin{aligned}\nabla^3 h(y) &= \ln\left(1 - \frac{1}{(n-x+2)^2}\right)\\ \nabla^4 h(y) &= \ln\left(1 - \frac{2(n-x)+5}{(n-x+2)^3(n-x+4)}\right).\end{aligned}$$

If these to are expanded in a Taylor Series, then it is easy to see that $O(\nabla^3 h(y)) = 1/(n-x)^2$ and $O(\nabla^4 h(y)) = 1/(n-x)^3$. Higher orders of differencing will produce terms at least as small as the fourth backward difference. These are the only ones needed to produce the approximation with two correction terms.

B.3 Derivation of Tail Probabilities for Tilting with $\varphi_x(Q)$ / Expansion Combination

B.3.1 Right Tail Probabilities

The derivation here will be very similar to the derivation considered in section B.1. Start with the following approximation to a point probability:

$$\begin{aligned}f(x) &= \varphi_x(Q)\frac{W!(x-W+Q)!}{x!Q!}\left[p(x-W+Q) + \frac{k_Q^{(2)}}{2}\nabla^2 p(x-W+Q)\right.\\ &+ \left.\left(-\frac{k_Q^{(3)}}{6}\nabla^3 p(x-W+Q) + \frac{(k_Q^{(2)})^2}{8}\nabla^4 p(x-W+Q)\right) + \cdots\right].\end{aligned}$$

The mean for the Poisson densities is $x-W+Q$. Here $k_Q^{(i)}$ is the i^{th} factorial cumulant for the transformed density. From this point on, let $x^* = x-W+Q$ and $y^* = y-W+Q$. Now sum both sides of this equation for x going from y to ∞. The first term on the right hand side is just the result for exponential tilting:

$$\sum_{x=y}^{\infty}\varphi_x(Q)\frac{W!(x-W+Q)!}{x!Q!}p(x-W+Q) = \frac{\varphi_X(Q)W!(y^*)^{Q-W}}{Q!}R_{y^*}^*(y).$$

Now $\sum_{x=y}^{\infty}\left[\frac{(x-W+Q)!}{x!}\right]\nabla^2 p(x-W+Q)$ can be computed using summation by parts. To use summation by parts, notice that if we let $b_x = \frac{(x-W+Q)!}{x!}$ then

$$\begin{aligned} b_x - b_{x+1} &= \frac{(x-W+Q)!}{x!} - \frac{(x+1-W+Q)!}{(x+1)!} \\ &= \frac{(x-W+Q)!}{x!}\left(1 - \frac{x+1-W+Q}{x+1}\right) \\ &= \frac{(x-W+Q)!(W-Q)}{(x+1)!}. \end{aligned}$$

Using this, the following is obtained by performing summation by parts one time:

$$\sum_{x=y}^{\infty}\left[\frac{(x-W+Q)!}{x!}\right]\nabla^2 p(x-W+Q) = -\frac{y^*!}{y!}\nabla p(y^*-1)+(W-Q)\sum_{x=y}^{\infty}\frac{(x-W+Q)!}{(x+1)!}\nabla p(x-W+Q).$$

Now perform summation by parts again on the last term. This time notice that if $b_x = \frac{(x-W+Q)!}{(x+1)!}$, then

$$\begin{aligned} b_x - b_{x+1} &= \frac{(x-W+Q)!}{(x+1)!} - \frac{(x+1-W+Q)!}{(x+2)!} \\ &= \frac{(x-W+Q)!}{(x+1)!}\left(1 - \frac{x+1-W+Q}{x+2}\right) \\ &= \frac{(x-W+Q)!(W-Q+1)}{(x+2)!}. \end{aligned}$$

Using this leads to the following results for summation by parts:

$$(W-Q)\sum_{x=y}^{\infty}\frac{(x-W+Q)!}{(x+1)!}\nabla p(x-W+Q) =$$
$$-(W-Q)\frac{y^*!}{(y+1)!}p(y^*-1) + (W-Q)(W-Q+1)\sum_{x=y}^{\infty}\frac{(x-W+Q)!}{(x+2)!}p(x-W+Q).$$

Now the last term of the previous expression may be expressed as follows:

$$\sum_{x=y}^{\infty}\frac{(x-W+Q)!}{(x+2)!}p(x-W+Q) = (y^*)^{Q-W-2}R^*_{y^*}(y+2).$$

Thus the first order approximation due to combining the tilting and expanding may be expressed as:

$$P(X \geq y) \approx \varphi_X(Q)\frac{W!}{Q!}\left[(y^*)^{Q-W}R^*_{y^*}(y) + \frac{k_Q^{(2)}}{2}L_2^*(Q)\right]$$

where $R^*_{y^*}(y+k) = \sum_{x=y}^{\infty} p(x+k)$ and

$$L_2^*(Q) = (W-Q+1)^{(2)}(y^*)^{Q-W-2}R^*_{y^*}(y+2) - \frac{y^*!}{y!}\left[\nabla p(y^*-1) + \frac{(W-Q)}{(y+1)}p(y^*-1)\right].$$

To obtain higher order terms for the approximation, proceed in the following way:

$$\begin{aligned} L_k^*(Q) &= \sum_{x=y}^{\infty} \frac{(x-W+Q)!}{x!}\nabla^k p(x-W+Q) \\ &= -\frac{y^*!}{y!}\nabla^{k-1}p(y^*-1) + (W-Q)\sum_{x=y}^{\infty}\frac{(x-W+Q)!}{(x+1)!}\nabla^{k-1}p(x-W+Q). \end{aligned}$$

Continue inductively to obtain the following:

$$\begin{aligned} L_k^*(Q) &= (W-Q+k-1)^{(k)}(y^*)^{Q-W-k}R^*_{y^*}(y+k) \\ &- \frac{y^*!}{y!}\left[\nabla^{k-1}p(y^*-1) + \frac{(W-Q)}{(y+1)}\nabla^{k-2}p(y^*-1) + \cdots\right. \\ &+ \left.\frac{(W-Q+k-2)^{(k-1)}}{(y+k-1)^{(k-1)}}p(y^*-1)\right]. \end{aligned}$$

Thus the right tail approximation becomes

$$\begin{aligned} P(X \geq y) &= \varphi_X(Q)\frac{W!}{Q!}\left[(y^*)^{Q-W}R^*_{y^*}(y) + \frac{k_Q^{(2)}}{2}L_2^*(Q)\right. \\ &\quad - \left.\frac{k_Q^{(3)}}{6}L_3^*(Q) + \frac{(k_Q^{(2)})^2}{8}L_4^*(Q) + \cdots\right]. \end{aligned}$$

B.3.2 Adjustments to Approximate Left Tail Probabilities

The derivation of the approximation for the left tail probability requires a few minor adjustments to the previous section. The formula for $\sum_{x=0}^{y}\frac{(x-W+Q)!}{x!}\nabla^k p(x-W+Q)$ becomes:

$$\sum_{x=0}^{y} \frac{(x-W+Q)!}{x!} \nabla^k p(x-W+Q) = \frac{(y^*+1)!}{(y+1)!} \nabla^{k-1} p(y^*) - (Q-W)\nabla^{k-1} p(Q-W-1) + (W-Q)\sum_{x=0}^{y} \frac{(x-W+Q)!}{(x+1)!} \nabla^{k-1} p(x-W+Q).$$

Continue inductively to obtain:

$$\begin{aligned} L_k(Q) &= (W-Q+k-1)^{(k)} (y^*)^{Q-W-k} R_{y^*}(y+k) \\ &+ \frac{(y^*+1)!}{(y+1)!} \left[\nabla^{k-1} p(y^*) + \frac{(W-Q)}{(y+2)} \nabla^{k-2} p(y^*) + \cdots \right. \\ &+ \left. \frac{(W-Q+k-2)^{(k-1)}}{(y+k)^{(k-1)}} p(y^*) \right] \\ &- (Q-W)[\nabla^{k-1} p(Q-W-1) + (W-Q)\nabla^{k-2} p(Q-W-1) \\ &+ \frac{(W-Q)(W-Q+1)}{2} \nabla^{k-3} p(Q-W-1) + \cdots \\ &+ \frac{(W-Q+k-2)^{(k-1)}}{(k-1)!} p(Q-W-1)] \end{aligned}$$

where $R_{y^*}(y+k) = \sum_{x=0}^{y} p(x+k)$. Thus the L-functions are slightly more complicated than the L^*-functions. Using these L-functions leads to the following approximation for left tail probabilities:

$$\begin{aligned} P(X \leq y) &= \varphi_X(Q) \frac{W!}{Q!} \left[(y^*)^{Q-W} R_{y^*}(y) + \frac{k_Q^{(2)}}{2} L_2(Q) \right. \\ &- \left. \frac{k_Q^{(3)}}{6} L_3(Q) + \frac{(k_Q^{(2)})^2}{8} L_4(Q) + \cdots \right]. \end{aligned}$$

Bibliography

[1] Agresti, A. (1990). *Categorical Data Analysis.* Wiley. New York.

[2] Agresti, A. (1992). Survey of exact inference for contingency tables (with discussion). *Statist. Sci.* **7**. 131-177.

[3] Baglivo, J., Olivier, D., and Pagano, M. (1988). Methods for the analysis of contingency tables with large and small cell counts. *Journal of the American Statist. Assoc.* **83**. 1006-1013.

[4] Barbour, A., Holst, L., and Janson, S. (1992). *Poisson Approximation.* Oxford University Press. New York.

[5] Barbour, A. and Jensen, J. (1989). Local and tail approximations near the Poisson limit. *Scand. J. Statist.* **16**. 75-87.

[6] Barndorff-Nielson, O.E., and Cox, D.R. (1979). Edgeworth and saddlepoint approximations with statistical applications. *J. Roy. Statist. Soc. Ser. B.* **41**. 279-312.

[7] Bedrick, E., and Hill, J. (1992). An empirical assessment of saddlepoint approximations for testing a logistic regression parameter. *Biometrics.* **48**. 529-544.

[8] Chakraborty, R. (1993). A class of population genetic questions formulated as the generalized occupancy problem. *Genetics.* **134**. 953-958.

[9] Cox, M.A.A., and Plackett, R.L. (1980). Small samples in contingency tables. *Biometrika.* **67**. 1-13.

[10] Daniels, H.E. (1954). Saddlepoint approximations in statistics. *Annals of Mathematical Statistics.* **25**. 631-650.

[11] Daniels, H.E. (1987). Tail probability approximations. *International Statistical Review.* **55**. 37-48.

[12] Davis, P. and Rabinowitz, P. (1975). *Methods of Numerical Integration.* Academic Press Inc. New York.

[13] Douglas, J.B. (1980). *Analysis with Standard Contagious Distributions.* International Co-operative Publishing House. Maryland.

[14] Easton, G. and Ronchetti, E. (1986). General saddlepoint approximations with applications to L statistics. *Journal of the American Statist. Assoc.* **81**. 420-430.

[15] Esscher, F. (1932). The probability function in the collective theory of risk. *Skand. Aktuarietidskr.* **15**. 175-195.

[16] Field, C.A., and Ronchetti, E. (1990). *Small Sample Asymptotics.* ISM Lecture Notes-Monograph Series. Vol.13. Hayward, CA: Institute of Mathematical Statistics.

[17] Fuchs, C., and Kenett, R. (1980). A test for detecting outlying cells in the multinomial distribution and two-way contingency tables. *Journal of the American Statist. Assoc.* **75**. 395-398.

[18] Gail, M., and Mantel, N. (1977). Counting the number or $r \times c$ contingency tables with fixed margins. *Journal of the American Statist. Assoc.* **72**. 859-862.

[19] Goldberg, R. (1976). *Methods of Real Analysis.* Wiley. New York.

[20] Haberman, S. (1973). The analysis of residuals in cross-classified tables. *Biometrics.* **29**. 205-220.

[21] Haberman, S. (1977). Log-linear models and frequency tables with small expected cell counts. *The Annals of Statistics.* **5**. 1148-1169.

[22] Haberman, S. (1988). A warning on the use of chi-squared statistics with frequency tables with small expected cell counts. *Journal of the American Statist. Assoc.* **83**. 555-560.

[23] Haldane, J.B.S. (1940). The mean and variance of χ^2, when used as a test of homogeneity, when expectations are small. *Biometrika.* **31**. 346-355.

[24] Hall, P. (1992). *The Bootstrap and Edgeworth Expansion.* Springer-Verlag. New York.

[25] Jensen, J. (1995). *Saddlepoint Approximations.* Oxford University Press. New York.

[26] Jin, R. and Robinson, J. (1999). Saddlepoint approximation near the endpoints of the support. *Stat. and Prob. Letters.* **45**. 295-303.

[27] Johnson, N.L. and Kotz, S. (1977). *Urn Models and Their Applications.* John Wiley. New York.

[28] Lugannani, R. and Rice, S. (1980) Saddlepoint approximations for the distribution of the sum of independent random variables. *Advances in Applied Probability.* **12**. 475-490.

[29] Menon, V. and Prasad, B. (1985). The probability generating function of empty cell variable in a randomized occupancy problem. *Comm. Statist.-Theory and Methods.* **14**. 2287-2292.

[30] Mertz, D. and Davies, R. (1968) Cannibalism of the pupal stage by adult flour beetles: An experiment and a stochastic model. *Biometrics.* **24**. 247-275.

[31] Reid, N. (1988). Saddlepoint methods and statistical inference. *Statist. Sci.* **3**. 213-238.

[32] Reid, N. (1995). The roles of conditioning in inference. *Statist. Sci.* **10**. 138-199.

[33] Simonoff, J. (1988). Detecting outlying cells in two-way contingency tables via backwards-stepping. *Technometrics.* **30**. 339-345.

[34] Skovgaard, I.M. (1987). Saddlepoint expansions for conditional distributions. *Journal of Applied Probability.* **24**. 875-887.

[35] Statulevicius, V., and Aleskeviciene, A. (1993). On large deviations in the Poisson approximation. *Theory of Probability and it's Applications.* 385-393.

[36] Stumpf, R.H., and Steyn, H.S. (1986). Exact distributions associated with an I×J×K contingency table. *Comm. Statist.-Theory and Methods.* **15**. 1889-1904.

[37] Terrell, G. (1999). *Mathematical Statistics, A Unified Introduction.* Springer. New York.

[38] Wang, S. (1992). General saddlepoint approximations in the bootstrap. *Stat. and Prob. Letters.* **13**. 61-66.

[39] Wood, A.T.A., Booth, J.G., and Butler, R.W. (1993). Saddlepoint approximations to the CDF for some statistics with non-normal limit distributions. *Journal of the American Statist. Assoc.* **88**. 680-686.

Appendix C

Vita

Steven Kathman was born on September 9, 1974 in Daytona Beach, Florida. After living in Daytona Beach for several years, Steven and his family moved to Winston-Salem, North Carolina. There he graduated from Parkland high school in 1992. After graduating high school he attended the University of North Carolina in Greensboro for one year, and then transferred to Wake Forest University. In December 1995 Steven completed his Bachelor of Science in Mathematics at Wake Forest. He graduated Cum Laude with Honors. In August 1996 Steven enrolled in the Masters program at Virginia Polytechnic Institute and State University, commonly referred to as Virginia Tech. In December 1997 he completed his Master of Science in Statistics. He continued in graduate school until December 1999 where completed the requirements for a Ph.D. in statistics. While attending graduate school, Steven received two awards: the Boyd Harshbarger award for superior scholarship during the first year of graduate studies, and the Jesse Arnold award for superior teaching by a graduate student. During the final year of graduate school, Steven served as a visiting assistant professor in the department of statistics at Virginia Tech. Upon completion of his Ph.D., he will be joining the Agency for Toxic Substances and Disease Registry as a Mathematical Statistician. The position is located in Atlanta, Georgia. Although Steven has enjoyed his time at Virginia Tech, he is looking forward to starting his career.