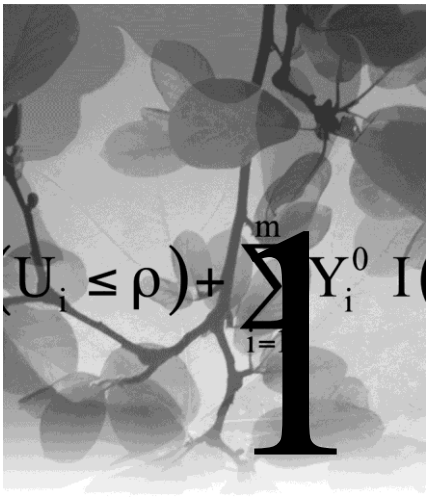


Overdispersion

1.1 A Simple Introductory Example 1
1.2 Overdispersion 3
1.3 A Common Characterization of Overdispersion 4
1.4 An Extended Definition of Overdispersion 6
1.5 A Simulated Example 7
1.6 The Exponential Family 11
Appendix 1.1: Hierarchical Models to be Covered in this Book 14

1.1 A Simple Introductory Example

The anticonvulsant phenytoin is known to be teratogenic on the prenatal development of inbred mice. Hartsfield et al. (1990) conducted a completely randomized experiment with 81 pregnant mice to investigate the possible synergistic effect of phenytoin (PHT) and trichloropropene oxide (TCPO). The study was set up as a 2x2 factorial; PHT and TCPO were the main factors. PHT was administered at the levels 0 mg/kg and 60 mg/kg, and TCPO was administered at 0 mg/kg and 100 mg/kg. The presence or absence of ossification in the phalanges on both the right and left forepaws on each of the fetuses is considered a measure of the teratogenic effect of the substances under study. In some phalanges the ossification did not occur at all; in some others, the ossification was almost complete. We selected the middle third phalanges because they yield around a 50% chance of ossification in the control groups. To simplify this part of our study further, we have analyzed only the response on the left middle third phalanx. Table 1.1 depicts the number of fetuses showing ossification in the middle third phalanx of the left forepaw out of the total number of fetuses in 11 mice that had been treated with both substances PHT and TCPO. This example is used here in a simple form to introduce the concept of overdispersion and will be revisited in greater depth in subsequent chapters.

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

Table 1.1: Ossification Data for 11 Mice Exposed to Phenytoin (PHT) and Trichloropropene Oxide (TCPO)

Mouse (j)	1	2	3	4	5	6	7	8	9	10	11	Total
Number of fetuses showing ossification (t_j)	2	0	1	7	0	0	0	0	6	1	1	18
Total number of fetuses (m_j)	2	7	8	8	10	4	6	7	6	6	7	71

If t_j and m_j represent, respectively, the number of fetuses showing ossification and the total number of fetuses of the j-th mouse, $j = 1,2,...,11$, then the estimated probability of observing ossification in any middle third phalanx of the left forepaw can be calculated as

$$\hat{\pi} = \frac{\sum_{j=1}^{11} t_j}{\sum_{j=1}^{11} m_j}$$

which, in this example, turns out to be $\frac{18}{71} = 0.2535$. If the t_j's were truly distributed as Binomial random variables with parameters (π, m_j), where π is the underlying true probability of ossification, an estimator of the variance of $\hat{\pi}$, given that mice are independent, would be

$$\hat{\text{Var}}(\hat{\pi}) = \frac{\hat{\pi}(1-\hat{\pi})}{\sum_{j=1}^{11} m_j}$$

The value of this estimated variance is 0.0027. Note that the above estimator is not valid if the Binomial model does not hold. Based on sampling theory (see Cochran 1977, p. 155), a consistent estimator, which does not depend on the binomial assumption, is

$$\tilde{\text{Var}}(\hat{\pi}) = \frac{n\sum_{j=1}^{n}(t_j - m_j\hat{\pi})^2}{\left(\sum_{j=1}^{n} m_j\right)^2 (n-1)}$$

which turns out to be 0.0142. The remarkable difference between the model-based estimate and the sampling-theory estimate leads us to the following question: Are these data really binomially distributed with a common probability of success π? An easy way to answer this question is by computing the usual Pearson's Chi-square

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

$$\chi^2 = \sum_{j=1}^{11} \frac{\left(t_j - m_j\hat{\pi}\right)^2}{m_j\hat{\pi}\left(1-\hat{\pi}\right)}.$$

Based on theoretical grounds, the value of this statistic divided by its degrees of freedom (10) is considered a Goodness-of-fit (GOF) statistic for the binomial-type of data. In this illustrative example, the value of the GOF statistic is 5.28. Under the Binomial model its value should be approximately 1. This means that in this case, the estimated variance of $\hat{\pi}$, previously reported as 0.0027, should be approximately 5.28 times larger than the value expected under the Binomial model. Note that 0.0142 divided by 0.0027 is 5.26. Thus, the data in this example exhibit variability larger than that permitted under the binomial assumption. This phenomenon is known as **overdispersion** or **extra variation**. This phenomenon also occurs under the Multinomial and Poisson distributions. In this example, the correct value of the estimated variance of $\hat{\pi}$ should be close to $(0.0027)(5.28) = 0.014$.

1.2 Overdispersion

Consider a study where experimental units are clusters, and where each elemental unit within the cluster is classified into two mutually exclusive categories. If we assume elemental units are independent with a common probability of success, the Binomial model is appropriate for analyzing the data. Clustering introduces a lack of independence among elemental units. The probability of success may vary from one elemental unit to another within a cluster. Consequently, the data usually exhibit larger variances than the variance permitted by the Binomial model. This phenomenon, as previously stated, is known as **overdispersion** or **extra variation**.

Overdispersion might be caused by other factors. Omitted variables in the model can cause overdispersion. That is, the response probability is in fact a function of several covariates, but the model contains a subset of them. This lack of fit manifests itself as overdispersion. Some of the diagnostic plots presented later are motivated by this observation.

Overdispersion is also observed on multinomial responses (on count data) analyzed under a Multinomial (Poisson) model when the residual variation obtained after fitting the Multinomial (Poisson) model may be greater than that what can be attributed to the sampling variation assumed by the model. McCullagh and Nelder (1989, pp. 124-125) point out

> "Overdispersion is not uncommon in practice. In fact, some would maintain that overdispersion is the norm in practice and nominal dispersion the exception."

The phenomenon of **underdispersion** occurs less frequently. Cox and Snell (1988, p. 107) point out that "underdispersion is sometimes observed, although this is uncommon." It might take place when there is a competition for a positive response. For instance, in a case where a person is tasting sweet drinks, he or she may feel the second drink is not so sweet—even if it indeed is. While underdispersion does occur, this phenomenon will not be treated here.

As discussed in Stigler (1986), examples of overdispersion arising from a clustered population can be traced in the literature as far back as the year 1876. Examples of Binomial and/or Multinomial models with overdispersion may arise (a) in a teratological study of a genetic trait which is passed on with a certain probability to offspring of the same mother; (b) in reproductive

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

toxicity experiments where chemicals are administered to litters and responses are measured on individual fetuses; (c) in experiments where individuals receiving treatments are allowed to interact with each other via support groups, as in stop-smoking programs or stop-drinking programs; (d) in household surveys where respondents within a household (or a neighborhood) may be strongly influenced by one or two individuals.

There are many papers that contain real examples which discuss and illustrate the overdispersion phenomenon and its consequences on erroneous inferences due to inflated Type I error rates. Some relevant examples in the areas of teratology and toxicology are given in Williams (1975, 1982), Williams et al. (1988), Kupper and Haseman (1978), Haseman and Kupper (1979), Kupper et al. (1986), Shirley and Hickling (1981), Paul (1982), Pack (1986), Lefkopoulou et al. (1989), Ryan (1992), Liang and McCullagh (1993), Carr and Portier (1993), Donner et al. (1994), Morel and Koehler (1995), and Morel and Neerchal (1997). Other important examples of overdispersion are found in Rosner (1984) and Donner (1989) in ophthalmology, Altham (1976) and Cohen (1976) in studies of hospitalized sibling pairs, Crowder (1978) in the analysis of seed germination, Heckman and Willis (1977) in connection with a panel study, Efron (1986) in an epidemiology example involving toxoplasmosis in 34 cities of El Salvador, and Moore (1987) in the modeling of chromosome aberration in survivors of the atomic bomb of Hiroshima.

Early remarks regarding overdispersion under the Poisson model can be found in Student (1919). Examples of Poisson models with overdispersion can be found (a) in the analysis of counts and rates of longitudinal studies, and (b) in behavioral studies and in studies of number of accidents where there is inter-subject variability. Real examples of overdispersion under the Poisson model are found in the analysis of number of tumors in rats, Gail et al. (1980) and Lawless (1987); in a longitudinal clinical trial aimed to study the number of epileptic seizures, Breslow and Clayton (1993); and on the frequency of recurrence of bladder cancer, Dean and Balshaw (1997). An example with archaeological data is used by Efron (1992) to discuss the various patterns of overdispersion relative to Poisson variability; in another study, Zeger and Edelstein (1989) assess the efficacy of Vitamin C in reducing child mortality in Indonesia.

1.3 A Common Characterization of Overdispersion

The phenomenon of binomial outcomes with overdispersion can be characterized in terms of the first two moments. Suppose that T is a Binomial random variable representing the number of successes out of m trials. Then,

$$E(T) = m\pi, \tag{1.1}$$

where π represents the probability of success of each Bernoulli trial. Furthermore, if the Bernoulli trials $Y_1, Y_2, ..., Y_m$ that encompass T satisfy the assumptions of the Binomial model, then $Var(T) = m\pi(1-\pi)$. On the other hand, if they are positively correlated, then the usual assumption of independence in the Binomial model is violated. Let us denote the correlation between two Bernoulli trials by ρ^2, $0 \le \rho^2 \le 1$. We use ρ^2 to indicate that the correlation is non-negative. Then, it can be shown that

$$Var(T) = m\pi(1-\pi)\{1+\rho^2(m-1)\}. \tag{1.2}$$

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

Note that while the expected value of T matches the Binomial model, its variance given in equation (1.2) is more than the variance modeled by a Binomial model because the factor $\{1+\rho^2(m-1)\}$ is greater than one. A random variable T with mean and variance, as in equations (1.1) and (1.2), is said to have overdispersion or extra variation relative to the Binomial model. This simple structure of overdispersion has been well-recognized among sampling practitioners. See, for instance, equation 8.1 in Hansen et al. (1953), where the variance under one-stage cluster sampling is expressed as a function of the variance under simple random sampling. Similar observations can be found in the classical books on sampling by Kish (1965) and Cochran (1977). Among sampling practitioners, the parameter ρ^2 represents the **intra-cluster** correlation and the factor $\{1+\rho^2(m-1)\}$ is the **design effect**. The case "positively correlated" corresponds to overdispersion. As previously stated, the occurrence of underdispersion is less common and corresponds to a negative correlation among the Bernoulli outcomes that comprise T.

Similarly, the definition of overdispersion can be extended to the case of Multinomial distribution. Consider a cluster consisting of m elemental units where the elemental units are being classified into (d+1) categories. Let **T** denote the d-dimensional vector of counts, corresponding to the first d categories, whose i-th component represents the number of elemental units classified into the i-th category. With clustered data, the vector **T**, which is usually modeled by a Multinomial distribution, often exhibit higher variability in the data than that allowed by the Multinomial model. In order to accommodate this overdispersion, the following moment structure is usually hypothesized as

$$E(\mathbf{T}) = m\boldsymbol{\pi} \tag{1.3}$$

and

$$Var(\mathbf{T}) = m\{Diag(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^t\}\{1+\rho^2(m-1)\}, \tag{1.4}$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2, ..., \pi_d)^t$ is a vector of probabilities, $0 < \pi_i < 1, \sum_{i=1}^{d} \pi_i \le 1$, and $\rho\left(\text{or } \rho^2\right)$ is the overdispersion parameter. If the multinomial-type variable **T** has mean vector and variance-covariance matrix as in equations (1.3) and (1.4), the variable **T** is said to have overdispersion relative to the Multinomial distribution. Note that when $\rho = 0$ the variance estimators in equations (1.2) and (1.4) reduce to that of the Binomial/Multinomial model. Several issues such as maximum likelihood estimation and Quasi-likelihood Models under the Multinomial distribution will be covered in Chapter 7.

Recall that for a Poisson random variable Y, $E(Y) = Var(Y) = \mu, \ \mu > 0$. Poisson random variable is a count of the number of occurrences of an event in a given time interval. As in the binomial case, there are underlying assumptions on the process for deriving the properties of a Poisson random variable. When these assumptions are not met, overdispersion may result. When Y is a Poisson-type of random variable, Y is said to have overdispersion if its first two moments are of the form

$$E(Y) = \mu, \ \mu > 0 \tag{1.5}$$

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

and

$$\text{Var}(Y) = \mu + \kappa\mu^2 = \mu(1+\kappa\mu),\ \kappa > 0. \tag{1.6}$$

The parameter κ is known as the **dispersion parameter**. Observe that when $\kappa = 0$, variance in equation (1.6) becomes that of the Poisson model. Dean (1992) provides an additional form of overdispersion by considering (1.6) as $\text{Var}(Y) = \mu(1+\kappa)$.

The consequences of ignoring overdispersion are well known. First of all, the estimated standard errors of naïve estimates (naïve in the sense of ignoring the clustered nature of the data) are smaller than they should be. Using them in testing of hypotheses result in inflated Type I Error Rates, i.e., false positive rates are larger than nominal ones. Hence, erroneous inferences ensue.

1.4 An Extended Definition of Overdispersion

There are many causes of overdispersion. In the case of the Binomial/Multinomial distribution, overdispersion results when either the assumption of independence (within elemental units) or common probability of success is violated. In the case of the Poisson distribution, overdispersion is observed when the rate of the events varies from subject to subject; such rates are assumed to be distributed as a Gamma random variable. Overdispersion also arises in a natural way in clustered data. It is even found in the case of longitudinal data with repeated measurements. Observed counts on subjects usually exhibit overdispersion. For instance, think of a clinical study of osteoporotic patients at high risk for vertebral fractures. Suppose that the number of incidents of vertebral fractures is measured during a given period. The same patient may suffer multiple occurrences and therefore make several contributions to the count, which leads to the violation of the "independent occurrences" assumption underlying the Poisson distribution. Therefore, the counts are not truly arising from a Poisson distribution and thus exhibit overdispersion.

The developments in the last 20 years or so on mixture distributions and **Generalized Linear Mixed Models (GLMM)** have brought overdispersion to full maturity. Overdispersion has surpassed its own boundaries: it is a much broader topic than the common characterization we provided in Section 1.2. Mixtures of Multinomial distributions, such as the Dirichlet-multinomial (Mosimann, 1962) and the Random-clumped Multinomial distribution proposed by Morel and Nagaraj (1993) and Neerchal and Morel (1998), have overdispersion relative to the multinomial distribution with the mean vector and variance-covariance matrix, as in equations (1.3) and (1.4). The Negative-binomial distribution with mean and variance, as depicted in equations (1.5) and (1.6), exhibits overdispersion relative to the Poisson distribution. A Binomial distribution with random effects properly defined has overdispersion relative to the Binomial model. Additionally, a Beta-binomial distribution with random effects has overdispersion relative to the Beta-binomial, and the Beta-binomial itself has overdispersion relative to the Binomial distribution. The list goes on and on. Appendix 1.1 in this chapter shows some hierarchical diagrams of overdispersion that we intend to cover in this book.

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

1.5 A Simulated Example

A simple way of simulating correlated Bernoulli data is as follows. Let $Y, Y_1^0, ..., Y_m^0$ be independent and identically distributed Bernoulli (π) random variables. For each i, i=1,...,m, define Y_i as Y with probability ρ, or as Y_i^0 with probability $(1 - \rho), 0 \le \rho \le 1$. In other words, each Y_i can be represented by the mixture

$$Y_i = Y\ I(U_i \le \rho)\ +\ Y_i^0\ I(U_i > \rho), \tag{1.7}$$

where the U_i are independent uniform (0,1) random variables, and I(.) denotes an indicator function. In this model for clumped data,

$$E(Y_i) = \pi,\ Var(Y_i) = \pi(1-\pi) \text{ and } Corr(Y_1, Y_2) = \rho^2.$$

The programming statements for generating data under model (1.7) are provided next for m=5, $\pi = 0.6$ and $\rho^2 = 0.3$. Note that n=20,000 clusters are being generated, each consisting of a string of m=5 correlated binary outcomes.

Program 1.1: Correlated Bernoulli Outcomes

```
/*
       Generation of correlated Bernoulli outcomes
       For Y(1),Y(2),...,Y(m)
       E(Y(i))          = Pi
       Var(Y(i))        = Pi*(1-Pi)
       Corr(Y(i),Y(i')) = Rho*Rho
*/

data correlated_bernoullis;
       n    = 20000; *--- Number of clusters;
       m    = 5;     *--- Number of elemental units within the cluster;
       pi   = 0.6;   *--- Probability of success of each elemental unit;
       rho2 = 0.3;   *--- Intra-cluster correlation;
       seed = 16670;
       rho  = sqrt( rho2 );
       do subjid = 1 to n;
              yy = 0;      *--- Variable yy plays the role of y of eq. (1.7);
              u  = uniform( seed );
              if u < pi then yy = 1;
              do i=1 to m;
                     y = 0;
                     u = uniform( seed );
                     if u < rho then y = yy;
                     else do;
                     uu = uniform( seed );
                     if uu < pi then y = 1;
                     end;
                     output;
              end;
       end;
       keep subjid i y;
run;
```

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

```
proc transpose data=correlated_bernoullis out=new;
       by subjid;
       id i;
       var y;
run;

ods html;
proc corr data=new;
       var _1 - _5;
run;
ods html close;
   proc transpose data=correlated_bernoullis out=new;
       by subjid;
       id i;
       var y;
   run;

   ods html;
   proc corr data=new;
       var _1 - _5;
   run;
   ods html close;
```

Output 1.1 indicates that the means of each Y_i is (as expected) approximately $\pi = 0.6$. The estimated standard deviations are about $\sqrt{\pi(1-\pi)} = \sqrt{(0.6)(0.4)} = 0.4849$. The estimated correlations among elemental units are about $\rho^2 = 0.3$. Note that the correlation structure of the repeated measurements is **exchangeable**. That is, $\text{Corr}(Y_i, Y_t) = \rho^2, \quad i \neq t, \text{ for } i, t = 1, 2, ..., m.$

Output 1.1: Results for Correlated Bernoulli Outcomes

The CORR Procedure

5 Variables:	_1 _2 _3 _4 _5

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
_1	20000	0.59995	0.48992	11999	0	1.00000
_2	20000	0.60305	0.48928	12061	0	1.00000
_3	20000	0.59855	0.49020	11971	0	1.00000
_4	20000	0.60065	0.48978	12013	0	1.00000
_5	20000	0.60140	0.48962	12028	0	1.00000

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

Pearson Correlation Coefficients, N = 20000 Prob > |r| under H0: Rho=0					
	_1	**_2**	**_3**	**_4**	**_5**
_1	1.00000	0.29579 <.0001	0.31689 <.0001	0.30128 <.0001	0.29763 <.0001
_2	0.29579 <.0001	1.00000	0.28872 <.0001	0.29704 <.0001	0.27709 <.0001
_3	0.31689 <.0001	0.28872 <.0001	1.00000	0.32044 <.0001	0.30346 <.0001
_4	0.30128 <.0001	0.29704 <.0001	0.32044 <.0001	1.00000	0.29617 <.0001
_5	0.29763 <.0001	0.27709 <.0001	0.30346 <.0001	0.29617 <.0001	1.00000

If we let $T = \sum_{i=1}^{m} Y_i$, Y_i's defined as in equation (1.7), it can be shown $E(T) = m\pi$ and $Var(T) = m\pi(1-\pi)\{1+\rho^2(m-1)\}$. Therefore, the binomial-type random variable T derived from the mixture model (1.7) exhibits overdispersion relative to the Binomial distribution. Thus, for m=5, $\pi = 0.6$ and $\rho^2 = 0.3$, the target values for the mean and the variance in the simulation are $E(T) = m\pi = (5)(0.6) = 3.00$ and $Var(T) = m\pi(1-\pi)\{1+\rho^2(m-1)\} = (5)(0.6)(0.4)(2.2) = 2.64$, which can be verified by adding the following statements to Program 1.1. We obtain 3.004 and 2.636 as empirical estimates of the mean and variance, respectively, in our simulation using 20,000 replications.

Program 1.1 Correlated Bernoulli Outcomes (Continued)

```
*--- Program 1.1 Continued;
proc means data=correlated_bernoullis sum noprint;
       class subjid;
       var y;
       output out=out1 sum=t;
run;

data out1;
       set out1;
       if _type_ = 1;
       drop _type _freq_;
run;
```

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

```
ods html;
proc means data=out1 mean var maxdec=2;
      var t;
run;
ods html close;
```

Recall that one of the consequences of overdispersion in the data is inflated Type I Errors. Program 1.1 can be easily extended to a simulation to demonstrate this undesirable effect. To this end, consider a sample of n=20 clusters each of size m=5. If the elemental units within the clusters are treated as independent, then we will have a total of 100 Bernoulli trials with the same probability of success π. We need to test the null $H_0 : \pi = 0.6$ versus the alternative hypothesis $H_1 : \pi > 0.6$. For the Binomial distribution, the Z-test at $\alpha = 0.05$ is to reject the null when $Z = \dfrac{\hat{\pi} - 0.6}{\sqrt{\hat{\pi}(1-\hat{\pi})/100}} > 1.6449$. An estimate of the true Type I Error can be obtained by performing this test on a series of overdispersed data sets generated by Program 1.1, and then by computing the percentage of rejections. Program 1.2 provides the necessary statements to perform this simulation.

Program 1.2: Estimating Actual Type I Error Rate for Data Created in Program 1.1

```
*--- Estimating actual Type I Error rate for overdispersed data;
data correlated_bernoullis; *--- This data set was created in Program 1.1;
      set correlated_bernoullis;
      reps = ceil( subjid / 20 ); *--- Creates 1000 reps of n=20 clusters of size m=5 each;
run;

proc means data=correlated_bernoullis mean noprint;
      class reps;
      var y;
      output out=test mean=Pi_hat;
run;

data test;
set test;
      if _type_ = 1;
      z     = (Pi_hat - 0.6) / sqrt( Pi_hat*(1-Pi_hat) / 100 );
      P_val = 0;
      if z > 1.6449 then P_val = 1;
run;

title "Monte Carlo Estimate of Type I Error Rate";
proc means data=test n mean maxdec=3;
      var P_val;
run;
```

For $\rho^2 = 0.3$, the estimated Type I Error rate turns out to be 0.160, about three times the nominal level of 0.05. We can now rerun Programs 1.1 and Program 1.2, appropriately concatenated for other values of ρ^2. The following table provides a summary for three runs using the same seed as above for convenience.

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

Table 1.2: Inflation of the Actual Type I Error Rate at Nominal Level $\alpha = 0.05$

ρ^2 (Overdispersion)	Actual Type I Error Rate
0.3	0.160
0.5	0.197
0.7	0.210

We hope that this simple example has convinced the reader that study of overdispersion is very important from the point of view of making accurate statistical inferences in practical situations. Many methods of taking overdispersion into account in several commonly used models are discussed in this book.

1.6 The Exponential Family

Several of the main topics to be covered in this book, namely Quasi-likelihood Functions, Generalized Linear Models (GLM), Generalized Estimating Equations (GEE), and Generalized Linear Mixed Models (GLMM), are motivated by the generality of the many statistical methods applicable to all members of the exponential family. Therefore, we begin with a brief introduction to the exponential family of densities by outlining some of their basic properties. The exponential family is a class of density functions sharing certain characteristics and properties. Among the members of this family are the Normal, Bernoulli, Binomial, Multinomial, Poisson, Negative-binomial, Beta, and Gamma distributions. This family of distributions plays a very important role in the classical statistical inference topics. Following the literature, we use the canonical parameter and canonical link function set up for our discussion, but all examples will be in the familiar notation, which may be or not be canonical.

A density function in this family can be written as

$$f(y) = \exp\left\{\frac{y\theta - b(\theta)}{\phi} + c(y, f(\phi))\right\}, \tag{1.8}$$

where $b(.)$ and $c(.)$ are some specific functions. The parameter θ is a real-valued **location parameter** and is known as the **natural (or canonical) parameter**. The function $b(\theta)$ is referred to as the **cumulant** and the parameter ϕ as the **dispersion (or scale) parameter**. We assume that $b(\theta)$ is continuously twice differentiable. In the case of the Normal distribution, the parameter ϕ corresponds to the variance σ^2 and is functionally unrelated with the mean. On the other

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

hand, in other distributions such as Bernoulli, Binomial, and Poisson, mean and variance are given by a single parameter and consequently, in the likelihood function (1.8) the dispersion parameter is not present in the sense that $\phi \equiv 1$. Therefore, an additional parameter, $\phi > 1$, is included to accommodate data with overdispersion. More information about this issue will follow in the next chapter.

If Y is a random variable whose density function belongs to the exponential family, then it can be shown that the mean and variance of Y are related to the cumulant function as follows

$$E(Y) = \mu = \frac{\partial b(\theta)}{\partial \theta} \tag{1.9}$$

and

$$Var(Y) = \phi \frac{\partial^2 b(\theta)}{\partial \theta^2}, \tag{1.10}$$

where $\frac{\partial b(\theta)}{\partial \theta}$ and $\frac{\partial^2 b(\theta)}{\partial \theta^2}$ denote the first and second derivatives of the function $b(\theta)$ with respect to θ.

Note that in equation (1.9) the parameter θ is expressed as a function of μ. This relationship is known as the **canonical link function** or as the **natural link function**. We will denote the canonical link function as $g(\mu) = \theta$. Obviously, $\mu = g^{-1}(\theta)$ since $b(.)$ is an invertible function with smooth first and second derivatives. The function $g^{-1}(.)$, which is nothing but the first derivative of the function $b(.)$, is known as the **mean function**. It seems as if we were going around in circles; however, it is important to distinguish between the link and the mean functions.

The variance of Y is the product of two terms, the dispersion parameter ϕ and the second derivative of $b(\theta)$. The latter is known as the **variance function** and depends on the mean μ through the first derivative of $b(\theta)$. Thus, the variance function is a function of μ and will be denoted by $v(\mu)$. Hence, equation (1.10) can be written as

$$Var(Y) = \phi v(\mu). \tag{1.11}$$

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

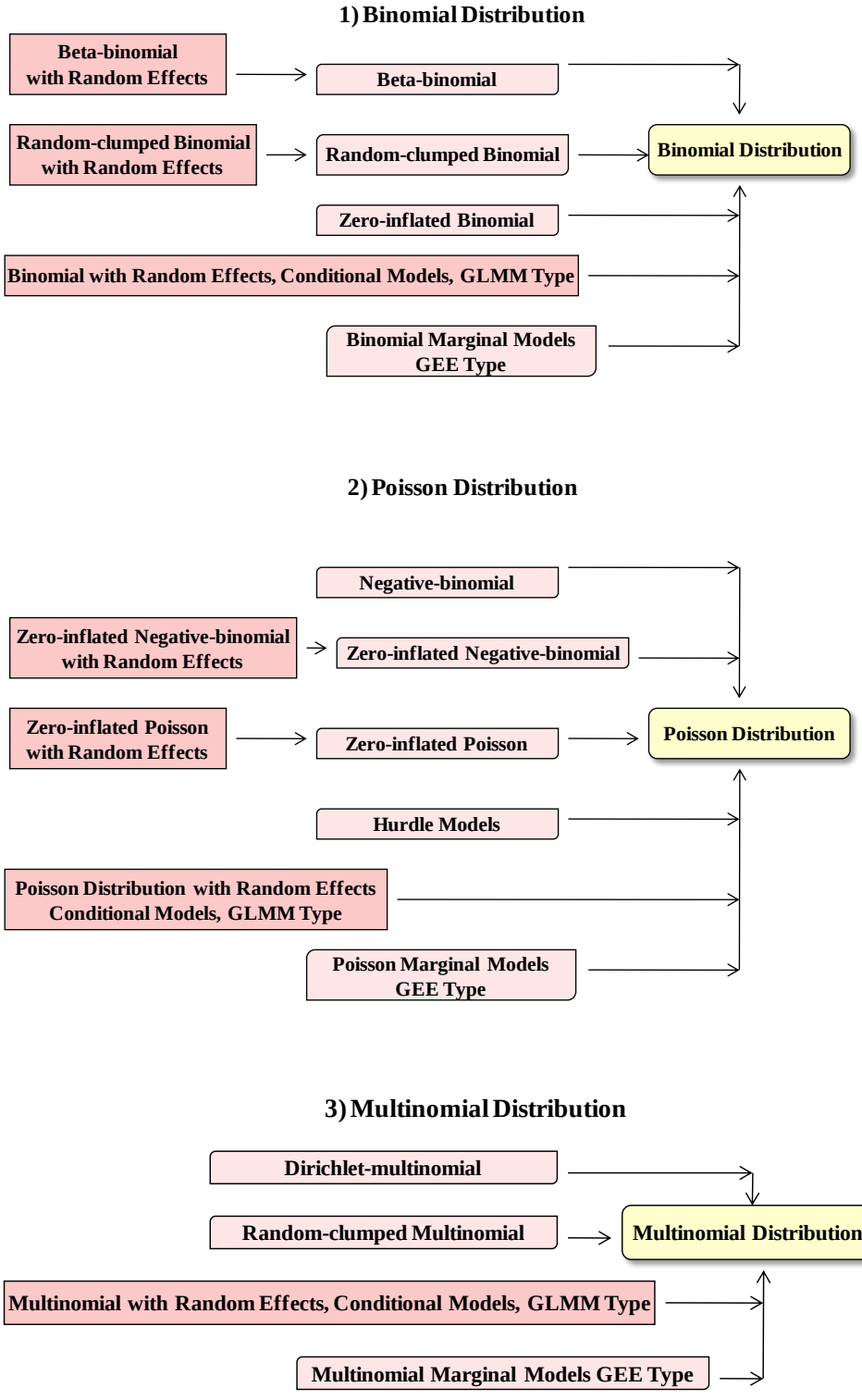
Table 1.3 shows the main features of the Bernoulli and Poisson distributions as members of the exponential family. Note that for the Bernoulli distribution, the mean and canonical link functions are the logistic and logit functions. They are also the same for the Binomial distribution (not depicted in Table 1.3). For the Poisson, the exponential and logarithmic functions are, respectively, the associated mean and link functions.

Table 1.3: Main Features of the Bernoulli and Poisson Distributions

	Distribution	
Feature	**Bernoulli**	**Poisson**
Probability function (usual form)	$\pi^{y}(1-\pi)^{1-y},\ y=0,1,\ 0<\pi<1$	$\frac{e^{-\lambda}\lambda^{y}}{y!},\ y=0,1,2,...,\ \lambda>0$
Cumulant $b(\theta)$	$\log(1+e^{\theta})$	$\exp(\theta)$
Mean function	$\pi=\frac{e^{\theta}}{1+e^{\theta}}$ (logistic function)	$\lambda=\exp(\theta)$ (exponential function)
Canonical link function	$\log\left(\frac{\pi}{1-\pi}\right)=\theta$ (logit link)	$\log(\lambda)=\theta$ (logarithm link)
Dispersion parameter ϕ	1	1
Variance function $v(\mu)$	$\pi(1-\pi)$	λ
Probability function (canonical form)	$\exp\left\{y\ln\left(\frac{\pi}{1-\pi}\right)+\ln(1-\pi)\right\}$	$\exp\{y\ln(\lambda)-\lambda-\ln(y!)\}$

Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.

Appendix 1.1: Hierarchical Models to be Covered in this Book



Morel, Jorge, and Nagaraj Neerchal. Overdispersion Models in SAS®. Copyright © 2011, SAS Institute Inc., Cary, North Carolina, USA. ALL RIGHTS RESERVED. For additional SAS resources, visit support.sas.com/publishing.