3 Algebraic statistical models

3.1 Introduction

We have seen in the last section that polynomial algebra is useful in experimental design to (i) describe the design (ii) study regression polynomial models which are identifiable with a particular design. A basic point is that we have described the model in an algebraic way. But we were only interested in the mean models and except for imposing the standard regression assumptions there was almost no distribution theory.

Apart from polynomial regression the first time we see algebra in a statistics course is in probability. Thus two events A and B in a probability space are independent if

$$P(A \cap B) = P(A)P(B),$$

and in general:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If we have two binary (Bernoulli) random variables (X_1, X_2) with joint distribution $\{p_{ij},\ i = 0, 1\}$. Then independence is equivalent to

$$p_{ij} = p_{i\cdot}p_{\cdot j} \tag{5}$$

(We use the notation $p_{i\cdot} = \sum_j p_{ij}$ etc.) It is straightward to check that (5) is equivalent to

$$p_{00}p_{11} - p_{10}p_{01} = 0,$$

We also see some algebra embedded in the Bernoulli distribution itself: if $X = \{0, 1\}$ we have

$$p_x = (1 - p)^{1-x}p^x,\ \ x = 0, 1$$

Here, interestingly, the variable x appears in the exponent. This is a case of a *power product*.

By an *algebraic statistical model* we mean a model which can be defined algebraically. Typically we have a basic model and the statistical model of interest is a sub-model. Thus the independence model in (5) is a sub-model of the basic model

$$\{p_{ij},\ p_{ij} \geq 0,\ i, j = 0, 1,\ \sum_{ij} p_{ij} = 1\}.$$

A important distinction is between models defined *explicitly* and models defined *implicitly*. An explicit model is typically defined in terms of *parameters*, whereas in an implicit model some of the parameters are eliminated.

Consider a straight line regression model with design $D = \{0, 1, 2\}$ and means given by

$$\eta_x = \mathrm{E}(Y_x) = \theta_0 + \theta_1 x,\ x \in D$$

Here we have two parameters (θ_0, θ_1). But an equivalent implicit model is:

$$\eta_0 - 2\eta_1 + \eta_2 = 0,$$

which is also algebraic. We can obtain the explicit model by eliminating the parameters (we have three equations in two variables).

Typically, the basic model is described by an algebraic variety and the model of interest defines a sub-variety. In the regression case both varieties are linear subspaces. In more complex cases they may be algebraic varieties and, importantly, we may need to use the elimination methods of section (1.5), to move from the explicit to the implicit representation.

3.2 Generalised Linear Models (GLM) and toric ideals

Let $Y = (Y^{(1)}, \ldots, Y^{(d)})$ be jointly distributed random variables with support $\mathcal{N} = N_{n_1} \times \cdots \times N_{n_d}$ where

$$N_j = \{0, \ldots, n_j - 1\},\ j = 1, \ldots, d.$$

We need a notation for a "point" in $\mathcal{N}$. It is important in understanding algebraic statistics that a design becomes a support for probability models.

We shall use x. This is a little risky, but at least it fits with the "design" point notation in Section 2. It also allows us to write $x = (x_1, \ldots, x_d)$ and define the distribution as simply

$$\{p_x,\ p_x \geq 0, x \in \mathcal{N}, \sum_{x \in \mathcal{N}} p_x = 1.$$

Consider an independent sample of size N from the distribution $\{p_x\}$. Here $n = \prod_{j=1}^{d} n_j$ is the number of cells. Then if the count in cell x is Y_x the full set of "counts" $\{Y_x\}$ has the multinomial distribution

$$P(Y_x = y_x) = \frac{N!}{\prod_x y_x!} \prod_x p^{y_x},$$

and note that $\sum Y_x = N$. We obtain the same result if we assume that the count in each cell each Y_x is an independent Poisson random variable with mean $\mu_x = Np_x$ and then condition on N.

Remark. Be careful of notation: n is the number of cells, which is the sample size for the independent Poisson random variables, while, to repeat $N = \sum Y_x$ is the total of all the Poisson counts.

The Poisson formulation is a little easier to handle and this will be our starting point for the log-linear formulation. Thus, let us set up a Poisson log-linear model as

$$\log \mu_x = \sum_{j=0}^{k-1} \theta_j f_j(x),\ x \in \mathcal{N} \tag{6}$$

Again, this is to be interpreted as our having $n = \prod_{j=1,\ldots,d} n_j$ independent Poisson random variables, one for each cell of the d-way table defined by $\mathcal{N}$. To make the algebra "work" we shall require the function $f_j(x)$ to have integer values (although rational may be handled).

At this point it is important to distinguish between *qualitative*, or categorical factors and *quantitative*, or possibly mixed quantitative and qualitative factors. Most work to date has been for qualitative factors.

Consider the general independence model for this case. First, define the marginal probabilities

$$P(Y_i = x_i) = p_{x_i}^{(i)} = \sum_{y:y_i=x_i} p_y$$

Then the independence model is

$$p_x = \prod_{i=1}^{d} p_{x_i}^{(i)}$$

The log linear model is:

$$\log p_x = \sum_i \log p_{x_i}^{(i)}$$

This implies that the log-linear model is *additive*, in that $\log \mu_x$ is additive over dimension, or in other words over the factors.

In the qualitative factor case can express this using indicator functions. Taking the Poisson case and using a double index:

$$\log \mu_x = \theta_0 + \sum_{j=1}^{n_i} \sum_{i=1}^{d} f_{ij}(x)\theta_{ji} \tag{7}$$

where

$$f_{ij}(x) = \begin{cases} 1 \text{ if } x_i = j \\ 0, \text{ otherwise} \end{cases}$$

It is useful to use the matrix form:

$$[\log \mu_x] = X\theta \tag{8}$$

where $[\cdot]$ means the vector indexed by x, and X is the same as the X-matrix as in the design of experiments.

As an example, in the 2^2 case with independence we have:

$$\begin{pmatrix} \log \mu_{00} \\ \log \mu_{10} \\ \log \mu_{01} \\ \log \mu_{11} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_{00} \\ \theta_{01} \\ \theta_{10} \\ \theta_{11} \end{pmatrix} \tag{9}$$

Note that for the general independence case the X-matrix is $(\sum_j n_j + 1) \times N$ and has rank $1 + \sum_{j=1}^{d}(n_j - 1) = \sum_{j=1}^{d} n_j - d + 1$.

Since the parameters are general, the model can be considered as saying that $[\log(\mu)]$ is in the range (column space) of X. Thus without loss of generality we may (linearly) reparametrize in any convenient way which preserves this space.

In the binary 2^d case there are two alternative but attractive way to parametrise to obtain a full rank X-matrix. The first is to delete the appropriate columns, preserving one parameter in each dimension and write

$$\log \mu_x = \theta_0 + \sum \theta_j x_j.$$

In the 2^2 case we have:

$$\begin{pmatrix} \log \mu_{00} \\ \log \mu_{10} \\ \log \mu_{01} \\ \log \mu_{11} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix} \tag{10}$$

The second parametrisation recodes the "levels" of the binary variables with $\{-1, 1\}$ instead of $0, 1$. This gives, in the 2^2 case,

$$\begin{pmatrix} \log \mu_{00} \\ \log \mu_{10} \\ \log \mu_{01} \\ \log \mu_{11} \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix}. \tag{11}$$

In this case the X matrix is exactly that of the 2^k factorial design of Section 2 in which $\{-1, 1\}$ is the traditional coding. Note: if we reparametrise the meaning of the parameters will change, although we have been lazy and used the same parameter θ.

3.2.1 Exponential models

We will assume from now on that we have a polynomial log-liner Poisson model or the equivalent multinomial models by writing as usual $p_x = \frac{\mu_x}{N}$ (the only change is that the constant term parameter θ_0 is different). Throughout the notes we will notes we will use p_x or μ_x depending on the situation.

For pure interpolation at integer (or rational grid points) we can always find a *saturated* exponential model of the form:

$$\log \mu_x = \sum_{\alpha \in L} \theta_\alpha x^\alpha,$$

where L is the saturated basis arising from the choice of term order. We can then use the same notation for sub models:

$$\log \mu_x = \sum_{\alpha \in L'} \theta_\alpha x^\alpha,$$

or the probability form

$$\log p_x = \sum_{\alpha \in L'} \theta_\alpha x^\alpha \tag{12}$$

The L, L' notation is important in our development.

Then in matrix terms

$$[\log \mu_x] = \sum_{\alpha \in L'} \theta_\alpha x^\alpha$$

and the matrix form

$$[\log \mu_x] = \exp(X\theta), \tag{13}$$

where $X = \{x^\alpha\}_{x \in D, \alpha \in L'}$ for some model list L'.

3.3 Power product models

There two basis ways of obtaining power product models, within the Poission/multinomial log-linear framework. The first is to take the probability

form of the exponential model and replace the exponential terms by power terms. Thus is we write

$$t^\alpha = \exp(x_\alpha), \alpha \in L'.$$

Then we obtain the form

$$p_x = \exp\left(\sum_{\alpha \in L'} \theta_\alpha x^\alpha\right) = \prod_{\alpha \in L'} t_\alpha^{\theta^\alpha},\ x \in D \tag{14}$$

Stretching notation somewhat we write this in matrix form:

$$[p(x)] = [t^X],$$

in which each row of X gives a product whose terms are controlled by the corresponding row $[x^\alpha]^T$ of X, where $x \in D$

The second way is to directly appeal to the products that appear in the factorisation of p_x, as in the independence case. If we assume that $p_x > 0$ for all $x \in D$, then the two representation are the same. But care has to be taken: they may be independence or conditional independence models with zero probabilities which are not obtainable from the exponential model: log of zero is $-\infty$.

3.4 Kernels and toric varieties

Write the log-linear model in matrix form:

$$[\log p_x] = X\theta$$

Then this is equivalent to the statement

$$[\log p_x] \in R(X),$$

where $R(X)$ is the range or column space of X. But this, in turn, is equivalent to the statement that

$$[\log p_x] \perp K(X), \tag{15}$$

where $K(X)$ is the kernel space of X.

We can construct $K(X)$ by extending X with a matrix K, so that

$$[X : K]$$

is full rank n an $X^T K = 0$. Then we take $K(X) = R(K)$, the column space of K. There are many ways to construct K and we may seek a simple forms of K, namely one with lots of zeros. We shall need K to be integer, which is always possible when X is integer, and note that X is always integer for polynomial models over an integer support. The orthogonality imposes linear conditions on the $\log p_x$ values. which can be written:

$$[\log p_x]^T K = 0.$$

In the 2^2 example, from the $\{-1, 1\}$ coding version that we easily see that $K^T = (1, -1, -1, 1)$ so we have a single linear condition

$$\log p_{00} - \log p_{10} - \log p_{01} + \log p_{11} = 0$$

Exponentiating we have have

$$\frac{p_{00}p_{11}}{p_{10}p_{01}} = 1$$

which reduces to

$$p_{00}p_{11} - p_{10}p_{01} = 0,$$

which we have already mentioned.

This method is general, for the type of models we describe here. Thus let the equations formed from a kernel K in the above manner be:

$$f_i(p) = p^{u^{(i)}} - p^{v^{(i)}},\ i = 1, \ldots, q, \tag{16}$$

where $q = \text{rank}(K) = n - \text{rank}(X)$. We call

$$I_K = \langle f_i :\ i = 1, \ldots, q \rangle,$$

the kernel ideal. We can also refer to the *lattice ideal*. Formally, this is obtained from the set of *all* integer solution solutions z to $Xz = 0$, not just the q columns K. Note that if the log-linear model contains a constant term (the first column of X is all ones) the ideal is homogeneous.

An important point is that, in general the kernel ideal is not the full toric ideal associated with the model, but we have the following.

Lemma 22 *The toric ideal associated with with an integer X-matrix X is generated by all monomial differences $p^\alpha - p^\beta$ where $X^T\alpha = X^T\beta$:*

$$I_X \ = \ \langle p^\alpha - p^\beta, X^T\alpha - X^T\beta = 0,$$

and α and β are non-negative integer exponents.

There are two equivalent ways of obtaining the toric ideal I_X.

In the first method uses *saturation* of the kernel ideal: $\langle f_i,\ i = 1, \ldots, q \rangle$. We extend the list $\{f_i,\ i = 1, \ldots, q\}$ with an addition term $v \prod_x p_x + 1$, where v is a "dummy" variable so that $v \prod_x p_x + 1 = 0$ prevents any p_x from being zero. We then take the elimination ideal, from eliminating the dummy v, according to subsection 1.5.

The second method uses the power product representation (13) and the original definition of a toric ideal as an elimination ideal from power products. We take the power product representation (14) and eliminate the parameters θ_α.

Here is an example. Let us take a 3×2 table with probabilities

$$\{p_{ij},\ i = 0, 1, 2,\ j = 0, 1\}$$

and the independence model. Then

$$X = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix}$$

We may take

$$K = \begin{pmatrix} 1 & 1 \\ -1 & -1 \\ -1 & 0 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix},$$

So that the lattice ideal is

$$I_K \ = \ \langle p_{01}p_{11} - p_{01}p_{10}, p_{00}p_{21} - p_{01}p_{20} \rangle$$

The saturation ideal is

$$\langle p_{01}p_{11} - p_{01}p_{10}, p_{00}p_{21} - p_{01}p_{20}, v\pi + 1 \rangle,$$

where $\pi = \prod_{ij} p_{ij}$. The elimination ideal, eliminating v is

$$I_X \ = \ \langle p_{01}p_{11} - p_{01}p_{10}, p_{00}p_{21} - p_{01}p_{20}, p_{11}p_{20} - p_{10}p_{21} \rangle$$

On the other hand the power product model is

$$\begin{aligned} p_{00} &= s_0 t_0 \\ p_{01} &= s_0 t_1 \\ p_{10} &= s_1 t_0 \\ p_{11} &= s_1 t_1 \\ p_{20} &= s_2 t_0 \\ p_{21} &= s_2 t_1 \end{aligned}$$

If we eliminate the s_i and t_i, using section (1.5) we obtained the same result as saturation of the kernel ideal.

3.5 Graphical models

One of the successes of algebraic statistics is to study the algebraic structure of graphical models. The term "graphical models" is a catch-all to describe certain sets of conditional independence statements. Here we shall consider the binary case. Consider three binary random variables $X_1, X_2, X_3 = \{0, 1\}$. We use the notation $X_1 \coprod X_2 | X_3$ for " X_1 is conditionally independent of X_2 given X_3". Think of the sample space as a 3-way table with

$$\text{prob}\{X_1 = i, X_2 = j, X_3 = k\} = p_{ijk}, \quad i, j, k = 0, 1$$

we can write down the following conditions form looking at the two conditional 2-way tables:

$$p_{ij0} = p_{i\cdot 0} p_{\cdot j0}, \; p_{ij1} = p_{i\cdot 1} p_{\cdot j1}, \; (i, j = 0, 1)$$

Alternatively use the log-linear form with the $[-1, 1]$ coding:

$$\log p_x = \theta_{00} + \theta_{10} x_1 + \theta_{010} x_2 + \theta_{001} x_3 + \theta_{101} x_1 x_3 + \theta_{011} x_2 x_3.$$

We see that this gives conditional independence because, on fixing $x_3 = \{-1, 1\}$, p_x factorizes over x_1, x_2. Knowledge of factorial design gives us easy access to the kernel, in this case.

An undirected graphical model is a graph $G = \{V, E\}$ with d vertices one for each variable $X_1, \ldots, X_d$ and certain edges. The required condition is that whenever we have three disjoint index sets $I, I, K \subset \mathbf{N} = \{1, \ldots, d\}$ with $I \cap J \cap J = \mathbf{N}$, with the property that there is no edge between any vertext in I and J then

$$X_I \coprod X_J | X_K.$$

A stronger property is the undirected Markov property:

$$X_i \coprod X_j | V \setminus \{i,j\},\ (i,j) \notin E$$

This says (the random variables associated with) any non-connected vertices are conditionally independent given all other vertices.

Related to a graph is a *simplicial complex*. For these we identify not just edges but triples, and higher order (abstract) *simplices*. For example

$$\{(1,2),(1,3),(1,4),(2,3),(2,4),(3,4)\}$$

can also be considered as defining $\{(1,2,3),(1,2,4),(1,3,4),(2,3,4)\}$ and the $(1,2,3,4)$. A simplicial complex is a set of simplices which contain all it subsimplices.

Definition 23 *A graph model is a graphical model in which G considered as a simplicial complex is the largest simplicial complex which can be built from G.*

Lemma 24 *A graph model is a graphical model if and only if it has no* 3*-cycle.*

Within the class of graphical models there is an important subclass of decomposable graphical models. A simplicial complex S is decomposable if either it is already a simplex or it can be decomposed $S = S_1 \cap S_2$, such that $S_1 \cap S_2$ is a simplicial complex. A graph is *chordal* if and only if any cycle of length ≥ 4 has a chord.

The following important theorem characterises the toric basis of a decomposable graphical model

Theorem 25 *The following statements about a graphical model defined by a graph G are equivalent.*

1. *The model is decomposable*
2. *The Gröbner basis is generated by "quadratics"*
3. *The MLE of the p_x are closed form rational functions of the raw count data y.*

3.6 Sufficient statistics and MLEs

We have not yet really discussed *data*. Thus let $\{Y_x\}$ be independent Poisson random variable over $x \in D$. A column of X in the polynomial log-linear models we are considering is indexed by a monomial x^α for $\alpha \in L'$, the model list. Thus the vector of generalised margins, which we shall see in a moment comprises the sufficient statistic is

$$T_\alpha = \sum_{x \in D} x^\alpha Y_x,$$

or we may use lower case y_x when we are emphasizing computations with a real sample. We can write this in matrix form:

$$T = X^T Y \tag{17}$$

The T_α can also be thought of as "generalised margins".

The Likelihood is

$$l = \prod_x = \frac{\mu_x^{y_x} \exp(-\mu_x)}{y_x!}$$

and the log-likelihood is

$$L = \sum_x \{y_x \log \mu_x - \mu_x - \log(y_x!)\}$$

Substituting in the model we have

$$\begin{aligned} L &= \sum_x \{\sum_{\alpha \in L'} \theta_\alpha x^\alpha - \exp(\sum_{\alpha \in L'} \theta_\alpha x^\alpha) - \log(y_x!)\} && (18) \\ &= \sum_{\alpha \in L'} \theta_\alpha \sum_x x^\alpha y_x - \sum_x \{\exp(\sum_{\alpha \in L'} \theta_\alpha x^\alpha) - \log(y_x!)\} && (19) \\ &= \sum_{\alpha \in L'} \theta_\alpha T_\alpha - \sum_x \{\exp(\sum_{\alpha \in L'} \theta_\alpha x^\alpha) - \log(y_x!)\}. && (20) \end{aligned}$$

We now list some standard statistical results.

1. The U statistic. From (19)

$$\begin{aligned} U_\alpha &= \frac{\partial L}{\partial \theta_\alpha} \\ &= T_\alpha - \textstyle\sum_x x^\alpha \exp(\sum_{\alpha \in L'} \theta_\alpha x^\alpha) \\ &= T_\alpha - \textstyle\sum_x x^\alpha \mu_x \end{aligned}$$

In matrix form we write:

$$U = X^T Y - X^T \mu$$

2. The maximum likelihood equations, also called the normal equations, are given by $U = 0$ or

$$X^T Y = X\mu$$

In the multinomial case we have

$$X^T \frac{Y}{N} = X^T p \tag{21}$$

We can solve these equation explicit for the θ or explicitly. That it to say we may combine (20) with the implicit constraints imposed from the toric conditions above. In that case the solution set is the intersection between a linear variety and the toric variety.

3. The Fisher information matrix is

$$\mathcal{I} = \left\{ -\frac{\partial^2 L}{\partial\theta_\alpha \partial\theta_\beta} \right\} = \left\{ \sum_x \sum_z x^\alpha z^\beta \mu_x \right\} = X^T \mathrm{diag}(\mu) X$$

4. For the saturated model ($L' = L$) the MLE of μ_x is just the raw cell count Y_x in the Poisson case and the MLE of p_x is $\frac{Y_x}{N}$.

5. When the model is saturated the maximum likelihood estimators of the raw cell means are just the raw counts $\hat{\mu}_x = y)_x$ and $\hat{p}_x = \frac{y_x}{N}$.

Instead of considering the MLE of the parameters θ_α, we can go directly for the MLE of the raw μ_x.

3.7 Moments and cumulants

For an index α and $Y = (Y_1, \ldots, Y_d)$ a set of d random variable define the (noncentral) α-moment as

$$m_\alpha = \mathrm{E}(Y^\alpha)$$

When Y has support D, as in the underlying model for the multinomial, then the fact that the support is finite yield conditions on the higher moments,

which are inherited form the conditions on the monomials x^α, give by the design basis.

If

$$g(x) = LT(g_j(x)) - \sum_{\alpha \in L} \theta_\alpha x^\alpha,$$

is a G-basis element of D we can interpret the statement $g_j(x) = 0$, namely

$$LT(g(x)) = \sum_{\alpha \in L} \theta_{\alpha \in L} x^\alpha$$

as saying that the $g_J(x)$ is obtained as *interpolating the leading term over* D, which is just the same as $x^\beta = NF(x^\beta), x \in D$. Suppose $LT(g(x)) = x^\beta$, then taking expectation we have

$$m_\beta = \mathrm{E}(Y^\beta) = \mathrm{E}(NF(Y^\beta)) = \sum \theta_{\alpha \in L} m_\alpha$$

The moment generating function (mgf) of X is

$$M_Y(s) = \mathrm{E}_X(e^{s.Y}) = \sum_{\beta \geq 0} \frac{m_\alpha s^\alpha}{s!},$$

where $s.X = \sum_{j=1}^d s_j X_j$ and if $s = (s_1, \ldots, s_d)$ then $s! = \prod_{j=1}^d s_j$. We want to find the reduced form of $M_X(s)$ which only depends on the $m_\alpha, \alpha \in L$.

Now

$$e^{s.x} = \sum_{\beta \geq 0} \frac{x^\beta s^\beta}{s!} = \sum_{\beta \geq 0} \frac{NF(Y^\beta) s^\beta}{s!},\ x \in D.$$

Substituting y for x and taking expectations we have

$$M_Y(s) = \sum_{\beta \geq 0} \frac{\mathrm{E}(NF(Y^\beta)) s^\beta}{s!}.$$

If X is the X-matrix for D the

$$NF(x^\beta) = [x^\alpha] X^{-1} [x^\beta],$$

where we have used our vector notation so that, for example, $[x^\beta]$ is the vector of values of x^β over D. Taking expectations

$$m_\beta = \mathrm{E}_Y(NF(x^\beta)) = [m_\alpha]^T X^{-1} [x^\beta]$$

We could have obtained this by noting that

$$[m_\alpha] = X^T[p_x],$$

so that

$$m_\beta = \sum_{x \in D} x^\beta p_x$$

As slightly different way of proceeding is to interpolate $e^{s.x}$ over D: to obtain

$$e^{s.y} = [y^\alpha] X^{-1} [e^{s.x}],\ y \in D$$

So that, replacing y by Y and taking expectations:

$$M_Y(s) = [m_\alpha] X^{-1} [e^{s.x}].$$

To clarify: $[m_\alpha]$ is the vector of moments for $\alpha \in L$ and $[e^{s.x}]$ is the vector of $e^{s.x}$ values for $x \in D$.

This finite representation can be extended to the cumulant generating functions

$$K_Y(s) = \log M_Y(s)$$

The above calculations are for any distribution on a finite support. Now let us specialize situation to the base distribution for a multinomial:

$$p(x) \quad = \quad \exp(\textstyle\sum_{\alpha \in L'} \theta_\alpha x^\alpha)$$

We can write this is a form which emphasizes the role of the cumulant generating function, and in a slightly more general form. First write $L_0 = L' \setminus \{0\}$ where $\{0\}$ is the null term $(0, \ldots, 0)$. That is we remove the constant term. Also we introduce a base distribution $p_0(x)$. This gives

$$p(x) = \exp\left(\sum_{\alpha \in L_0} \theta_\alpha x^\alpha - K(\theta) \right) p_0(x).$$

Now compute the moment generating function of the variables $Z = \{X^\alpha, \alpha \in L_O\}$:

$$\begin{aligned} M_Z(s) &= \mathrm{E}(\exp(\textstyle\sum_{\alpha \in L_O} s_\alpha X^\alpha)) \\ &= \exp(\textstyle\sum_{\alpha \in L_0} (s_\alpha + \theta_\alpha) x^\alpha - K(\theta)) p_0(x) \\ &= \exp(K(\theta + s)) - K(\theta). \end{aligned}$$

Note that $K(0) = 1$. Thus $K(s)$ is the cumulant generating function of Z with respect to the base distribution $p_0(x)$ and in general the cumulant generating function is

$$K(\theta + s) - K(\theta).$$

Note that when our chosen model defined by L' is a proper submodel of the saturated model L we have nesting of the moment spaces:

$$\mathcal{M}_{L'} \subset \mathcal{M}_L \subset \mathcal{M}.$$

We can obtain the estimated moments $\hat{\mathcal{M}}_{L'}$ directly from the MLE equations. This gives us also the MLE of the cell probabilities p_x. We can use the formula relating moments tom probabilities to obtain the estimates for $\mathcal{M}_L \backslash \mathcal{M}$. Then extend these to all moments via the above methods.

There is a beautiful relationship between moments and cumulants, which is best explained in the in the square-free case. Consider that the case of three random variables: Y_1, Y_2, Y_3. The formula are related to the partition lattice whose elements are

$$\begin{array}{ccccc} & & (111) & & \\ (100,011) & & (010,101) & & (001,110) \\ & & (100,010,001) & & \end{array}$$

In addition to the above analysis we have the well-known $1-1$ relationship between the corresponding moments and cumulants.

$$\begin{aligned}
\mu_{100} &= k_{100} \\
\mu_{010} &= k_{010} \\
\mu_{001} &= k_{010} \\
\mu_{110} &= k_{110} + k_{100}k_{011} \\
\mu_{101} &= k_{101} + k_{010}k_{101} \\
\mu_{110} &= k_{011} + k_{010}k_{001} \\
\mu_{111} &= k_{111} + k_{100}k_{011} + k_{010}k_{101} + k_{010}k_{001} + k_{100}k_{010}k_{001}
\end{aligned}$$

$$\begin{aligned}
k_{100} &= \mu_{100} \\
k_{010} &= \mu_{010} \\
k_{001} &= \mu_{010} \\
k_{110} &= \mu_{110} - \mu_{100}\mu_{011} \\
k_{101} &= \mu_{101} - \mu_{010}\mu_{101} \\
k_{110} &= \mu_{011} - \mu_{010}\mu_{001} \\
k_{111} &= \mu_{111} - \mu_{100}\mu_{011} - \mu_{010}\mu_{101} - \mu_{010}\mu_{001} + 2\mu_{100}\mu_{010}\mu_{001}
\end{aligned}$$

The inverse can make use of *Möbius inversion* over the partition lattice; a cheap way to do this is to use upper triangular matrices whose (total) order or rows and columns respects partial order of partitions. Higher order cumulants and moment formulae can then be obtained by "replication" of variable: eg $\mu_{2,1} = \mu_{111}$ with Y_2 being a copy of Y_1. There are also nice formulae for the moments and cumulants of polynomials of random variables.

There are many uses of moments and cumulants:

1. Central Limit Theorems: first and second order cumulants
2. Edgeworth expansion: high order cumulants
3. Asymptotics: cumulants of MLE's
4. Saddlepoint approximations
5. Non-gaussian data: "Higher order statistics" (this has become fashionable again in signal processesing).

Finally we can extend to all estimated moments $\hat{\mathcal{M}}$, by using the moment aliasing.

4 Markov bases and simulation

4.1 Introduction: exact tests

We have seen in section 3.6 that in the Poisson/multinomial models the likelihood and hence the MLE depends only on the sufficient statistics: $\{T_\alpha, \alpha \in L\}$. By the definition of sufficiency we see that the condition distribution, *under the model* of a (potential) sample given the sufficient statistic is independent of the parameters $\{\theta_\alpha,\ \alpha \in L'\}$. We can argue, therefore, that the conditional distribution under the model of any statistic V (which, by definition is a function of the data), given T, also does not depend on the parameters.

We can use this to carry out so-called *exact conditional tests*. The argument is as follows. Let V be a test statistics, for some hypothesis, such as that the model given by L' holds as opposed to the unrestricted saturated model L. Then, if possible, compute the p-value of V under the conditional distribution, given T. That is if y is the observed data compute

$$p = \text{prob}\{V(Y) \geq V(y) \mid T, \text{under } L'\}$$

Since p is a condition probability, given T, if we take expectations with respect to T its value remains unchange, and so we have computed the unconditional p-value. We might reject the hypothesis under consideration if $p \leq 0.05$.

Unfortunately, the conditional distribution of the data given the sufficient statistics is hard to compute. This difficulty is inherited by V. One solution is to set up a Markov Chain Monte Carlo (MCMC) method to simulate the distribution. It was a major breakthrough to show that this could be done using help from the algebraic methods.

If we have such an algorithm then we can attach the statistic of interest to the algorithm and estimate p by the proportion of simulated valse which are not smaller than the observed value, that is the value for our table of data.

4.2 Markov bases

It is fairly easy to see that given a two-way table of data $\{y_x\}$, with nonzero margins we can add add and subtract integers from the table in a way which keeps the margins fixed. Thus if we add 1 to entries $(i, j)$and (r, s) and subtract 1 from entries (i, s) and (j, r) then the margins stay fixed.

Definition 26 *Given a model with X-matrix X, a move is as set of integers $m(x), x \in D$ such that*

$$X^T[m(x)] = 0$$

Then it is immedaite that if we add or subtract $m(x)$ from a table y we have that the sufficient statistics does not change:

$$X^T(y + \pm[m(x)])) = X^T y \pm X^T[m(x)] = T$$

Given a model and the corresponding X-matrix we try to construct a Markov chain based on moves. The first task is to have a list of moves which, in s certain sense, is complete. We require that (i) we can travel from any data table to any other *with the same margins* and (ii) we can do this for any data table. Note that for a particular move $m(x)$ we may use $\pm m(x)$.

Definition 27 *Given any model with X-matrix X, a Markov basis for a model is a set of moves $\mathcal{M} = \{m_x\}$ such that given any two tables with y*

and y' with the same value of the sufficient statistic, namely $T(y) = T(y')$, we can find a subset of moves $\mathcal{M}' \subseteq \mathcal{M}$ to "travel" from y to y' ie such that

$$y' = y + \sum_{m \in \mathcal{M}'} \pm m_(x)$$

From the definition of a move it is clear that all moves lie in the kernel of X, but, in general the Markov basis is not given simply by the kernel. Given a move $m(x)$ we can form "binomials" by separating the positive and negative terms, $m_+(x), m_-(x)$ and writing

$$p^{m_+}(x) - p^{m_-(x)}.$$

The following is the key theorem.

Theorem 28 *A set of moves $\{\pm(m(x)\}$ is a Markov basis for a given X if the ideal generated by the binomials constructed from all the $m(x)$ is the toric ideal I_X.*

It is important to summarize the different bases:

Lattice (kernel) basis $\subset$ *Markovbasis* $\subset$ *Gröbnerbasis* The Markov basis can be taken as the generators of the Toric ideal, (which we have explained is the Kernel basis plus saturation). We could also discuss minimal Markov bases. An example of a minimal Markov basis is based on all 2×2 sub-tables of and $I \times J$ tables. If we use the notation $m_{i,j}$ for the kernel elements at the location (i, j) in the table then a typical move is

$$m_{i,j} = 1,\ m_{i,j'} = -1,\ m_{i',j} = -1,\ m_{i',j'} = 1.$$

Note that if we only need a lattice basis then we do not even need all 2×2 tables.

Higher-way tables may have longer moves, even in the complete case and incomplete tables (tables with prohibited cells) may need longer moves, even in the $I \times J$ case.

4.3 The algorithm

The purpose, then is to run a Markov Chain Monte Carlo (MCMC) with moves give by a Markov basis. Here is one version of the algorithm

1. Derive a Markov basis: $\{m_1(x), \ldots, m_k(x)\}$
2. Start by choosing an m_i uniformly at random and any table with the required margins, such as the original data table.
3. Given the chain at table g, find the set J of integers j such that $g + jm_i(x)$ is an allowable table.
4. Move by choosing a $j \in J$ with probability:

$$\prod_{x \in C_i} ((g + jm_i)!)^{-1},$$

where $C_i = \{x : m_i(x) \rangle 0\}$.

This is a connected, reversible, aperiodic Markov chain, with the required stationary distribution: the conditional distribution of the table give the (generalised) marginals. Note that this is the basic algorithm but it can be made faster by using standard "tricks" from MCMC, such as Metropolis-Hastings.

5 Monomial ideals

5.1 Basics

As we have seen monomial ideals are at the heart of this subject. In this section we want to explore this areas in more detail to investigate whether there are further application to statistics and probability. We shall restrict ourselves,mostly, to the case that L is finite, that is to the ideals of of a finite set of points.

Let $d = 2$ and consider the monomial ideal defined by a single monomial say

$$I = \langle x_1 x_2^2 \rangle.$$

We ask: what is $V(I)$, the variety of I? Now $x_1^2 x_2 = 0$ if either $x_1 = 0$ or $x_2 = 0$ or both. So $V(I)$ is the union of the x_1 and x_2 axes. Label these axes respectively as A_1, A_2.

Let us consider another example: $I = \langle x_1^3 x_2, x_1 x_2, x_1^4 \rangle$.

$$\begin{aligned} V(I) &= V(x_1^3 x_2) \cap V(x_1 x_2) \cap V(x_1^4) \\ &= (A_1 \cup A_2) \cap (A_1 \cup A_2) \cap A_1 \\ &= A_1 \end{aligned}$$

From this example something we guess the general case

Proposition 29 *In describing the variety $V(I)$ associated to a monomial ideal, it is enough to consider its square-free equivalent, obtained by reducing all powers is all monomials in the basis to unity.*

We can easily create examples where we have, say $A_1 \cup A_{23}$, where A_{23} is the x_2, x_3 axial plane. In fact, for monomial ideals we always obtain finite unions of axial (coordinate) subspaces. From this we can define the *dimension* of $V(I)$, for a monomial ideal I as the dimension of the largest of these subspaces. In the case $V(I) = A_1 \cup A_{23}$, this would be 2.It is fairly easy to see that:

$$\dim(V(I)) = d - \min d_j$$

where d_j is the dimension of a typical coordinate subspace in the decomposition.

Now let us consider the case where $\dim(V(I)) = 0$. In that case the above decomposition must be into $A_{1\ldots d} = 0$. We can also show that in this only happens if, for every $j = 1, \ldots d$ we have a monomial in I which is a simple power $x_j^{\alpha_j}$, for some integer $\alpha \geq 0$.

Now take

$$I = \langle x^{\beta^{(i)}}, \ldots, x^{\beta^{(m)}} \rangle$$

and consider the monomials I. We describe these as follows. Let $Q(\beta)$ be the non-negative quadrant with "corner" at β:

$$Q(\beta) = \{\alpha : \beta \leq \alpha\},$$

and let $M(\beta)$ be the corresponding set of monomials. Then the monomials in I are

$$\cup_{j=1}^{m} M(\beta^{(j)})$$

The structure of this set is of great interest and is the main topic of this section. Important is the set of monomial *not* in a monomial ideal I: namely x^α, $\alpha \in L$. As we have seen, L gives the monomial basis for the quotient ideal.

We can show that L has it has the following structure.

1. It may finite and this happens when $\dim(V(I)) = 0$
2. If it is not finite it will be the union of infinite strips which are the non-negative integers vectors on coordinate hyper-planes, together (possibly) with their translations (which therefore also go to infinity).

3. We may need to include translations of the origin, which is included if the ideal is proper, that is not equal to all of $k[x_1, \ldots, x_d]$. We think of the origin a a zero-dimesional coordinate hyperplane.

4. L has the order ideal property $\alpha \in L \Rightarrow \beta \in L$ for any $0 \leq \beta \leq \alpha$.

The expression of L as union of (non-negative integer) coordinate planes and shifted coordinate planes called the *Stanley decomposition*.

5.2 Hilbert series, Hilbert functions, Hilbert dimension

We are familiar with formulae which are generating functions which hold information about a mathematical object, as eg coefficients (we have already discussed moment generating functions). Sets can also have generating functions. We have seen that integers can be "held" as exponents eg $(1, 2) \rightarrow x_1 x_2^2$.

Suppose we have a set of integer vectors $\{\alpha, \alpha \in S\}$, then we can define the generating function of S as

$$G_S(x) = \sum_{\alpha \in S} x^\alpha.$$

We already know some generating functions from elementary algebra:

1. $S = \{0, 1, 2, \ldots\}$
$$G_S(x) = \frac{1}{1-x}$$

2. $S = \{0, 1, 2, \ldots n\}$
$$G_S(x) = \frac{1 - x^{n+1}}{1-x}$$

3. $S = \{x^k, x^{k+1}, \ldots\}$
$$G_S(x) = \frac{x^k}{1-x}$$

The generating function of the integer positive orthant, Z_+ is

$$\frac{1}{(1-x_1)\cdots(1-x_d)}$$

The generating function of the upper orthant (see above) $Q(\beta) = \{\alpha : \beta \leq \alpha\}$ is easily seen to be

$$G_{Q(\beta)} = \frac{x^\beta}{(1-x_1)\cdots(1-x_d)}.$$

This is just a shifted version of the formula for Z_+.

As we have said, the union of upper orthants gives all the monomials in the corresponding monomial ideal and its complement "gives" the quotient ideal. We can find the generating function of the union of upper orthants by brute force inclusion-exclusion. For just two orthants we have

$$I = \langle x^\alpha, x^\beta \rangle$$

and the generating function is

$$G_{Q(\alpha)\cup Q(\beta)} = G_{Q(\alpha)} + G_{Q(\beta)} - G_{Q(\alpha\wedge\beta)},$$

where

$$\alpha \wedge \beta = (\max(\alpha_1, \beta_1), \ldots, \max(\alpha_d, \beta_d)),$$

which corresponds to $\mathrm{LCM}(x^\alpha, x^\beta)$. The generating function of L is obtained by subtracting this from the generating function for Z_+.

Definition 30 *If I is a monomial ideal the* multigraded Hilbert series *is the generating function of the set L of exponents vectors which give the basis of the quotient ring $k[x_1, \ldots, x_n]$*

$$\tilde{H}(k/I)(x) = \sum_{\alpha\in L} x^\alpha$$

There are other well known entities which can be derived from $\tilde{H}(K/I)$. The (ordinary) *Hilbert series* is formed by setting $x_1 = \cdots = x_d = s$. We call this $H(k/I)(s) = \sum_k c_k s^k$. We may just write $H(s)$, for simplicity. An important feature of this is that the coefficient c_k in the Taylor expansion is the number x^α in the quotient basis of degree $|\alpha| = k$.

For experimental designs, that is ideals of points, we know that L is finite so that $H(s)$ is a finite polynomial, again whose coefficients count the terms of given degree. Moreover

$$H(1) = |L| = |D|.$$

A nice result from the design view point is that if the monomial ordering $\prec$ is that if L is arises from the design method in Section 2 and we have used *any* monomial ordering which respects degree ($|\alpha|\langle = |\beta| \Leftrightarrow \alpha \prec \beta$) then $H(s)$ is always the same. The *Hilbert function* is simply the values of the coefficients in the Hilbert series.

$$H_k = c_k$$

If the Hilbert series in finite, then we know that we have a finite variety and L is finite. This is exactly the case of experimental design. Conversely, we should be able to find out the dimension by seeing over how in how many dimensions L goes to infinity. But this is contain in the numerator of the (ordinary) Hilbert series. If, for example, (after taking into account any cancelations) we have $(1-s)^2$ in the denominator, then we know this must have come from some term $(1-x_i)(1-x_j),;$
$;i \neq j$ in the multigraded Hilbert series, which in turn indicates that we have the x_i, x_j integer orthant in L. Now,

$$\frac{1}{(1-s)^2} = 1 + 2s + 3s^2 + 4s^3 + \ldots$$

This shows that the rate of growth of the Hilbert function; the numerator has no affect on the rate, only the coefficient. Question: find a formula for the dimension using the Hilbert function.

5.3 Resolution

We sketch with a brief example the idea of resolution and its connection to the Hilbert function.

Consider the ideal

$$I = \langle x_1^4, x_2^4, x_3^4, x_1^3x_2^2x_3, x_1x_2^3x_3^2, x_1^2x_2x_3^3\rangle,$$

We are interested in the structure of the lower (integer) boundary of the set of all the monomials in I. By use of inclusion-exclusion (LCM, $\wedge$) we obtain additional terms on this boundary. In this case and just putting in the α exponents we obtain, including the generators

$$400, 040, 004, 321, 132, 213,$$
$$404, 214, 134, 044, 413, 233, 142, 323, 332, 421, 341, 440,$$
$$414, 234, 144, 423, 333, 342, 441$$

In some cases we can obtain the maximal elements (the third line above) from the irreducible components. In this case we have

$$I = \langle x_1^4, x_2, x_3^4\rangle \cap \langle x_1^2, x_2^3, x_3^4\rangle \cap \langle x_1, x_2^4, x_3^4\rangle \cap \langle x_1^4, x_2^2, x_3^3\rangle \\ \cap\langle x_1^3, x_2^3, x_3^3\rangle \cap \langle x_1^3, x_2^4, x_3^2\rangle \cap \langle x_1^4, x_2^4, x_3^1\rangle$$

The Hilbert series is obtained by taking alternating sums on the numerator at each "level" of intersection. In this example:

$$H(k/I)(x) = \frac{1 - (x_1^4 + \cdots + x_1^2x_2x_3^3) + (x_1^4x_3^4 + \cdots x_1^4x_2^4) - (x_1^4x_2x_3^4 + \cdots + x_1^4x_2^4x_3)}{(1-x_1)(1-x_2)(1-x_3)}.$$

If we remove the "1 and reverse the signs we obtain the generating function for the monomials in original monomials ideal given by the multi-graded Hilbert series:

$$H(I)(x) = \frac{(x_1^4 + \cdots + x_1^2x_2x3^3) - (x_1^4x_3^4 + \cdots x_1^4x_2^4) + (x_1^4x_2x_3^4 + \cdots + x_1^4, x_2^4, x_3)}{(1-x_1)(1-x_2)(1-x_3)}$$

This decomposition into alternating sums is related to the *resolution* of the ideal. Such resolutions are not unique and there is a special class called a *free resolution* and within this class a unique *minimal free resolution*. Although we have not defined minimal it means, roughly, that we have the fewest number of terms at each level. A useful fact (see next section) is that it gives inclusion-exclusion formulae with low complexity, but for the same union. In the above case we could have continued taking inclusion-exclusion with alternating signs ending with $x_1^4x_2^4x_3^4$. This is called the *Taylor resolution* and usually gives redundant extra terms.

There is an "easy" case where we know immediately that we have a minimal free resolution. This is when the "points" (generators) are in generic position: no point has a coordinate the same as for another point. In case the structure is called the *Scarf complex*.

5.4 Application to reliability and inclusion-exclusion

A multi-state system is defined here as a system of n components whose states are described by real variables $Y = (Y_1, \ldots, Y_n)$, which can be in one of a set of states which we define as the n-dimensional non-negative integer grid $\mathcal{Y} = \mathbb{N}^n$. There is a distinguished subset, $\mathcal{F} \subset \mathcal{Y}$, called the *failure set*, with the interpretation that if $Y \in \mathcal{F}$ the system is said to fail. A

member of $\mathcal{F}$ is called a *cut*. Let $\leq$ be the usual multivariate inequality $y \leq z \Leftrightarrow y_i \leq z_i,\ i = 1, \ldots, n$ and let $y \langle z$ when $y \leq z$ and $y_i \langle z_i$ for at least one $i = 1, \ldots, n$. Also define $x \vee y = (\max(x_1, y_1), \ldots, \max(x_n, y_n))$. Then we call the system *coherent* if

$$y \in \mathcal{F},\ y \leq z \Rightarrow z \in \mathcal{F} \tag{22}$$

Coherency is the principle that if a system has failed and the components move to a worse (higher) state value then the system remains failed.

In reliability, Y is a random variable, which summarises the consequence of internal degradation or external shock to the system liable to increase the values of states, although by repair one can also decrease the value.

We want to evaluate or bound the probability of failure $P(\mathcal{F}) = \text{prob}\{Y \in \mathcal{F}\}$. We will be concerned, not so much with the dependence of $P(\mathcal{F})$ on the distribution of Y, but rather with the set $\mathcal{F}$. For any set $U \subseteq \mathcal{Y}$ we define the indicator

$$I_U(y) = \begin{cases} 1 & \text{if } y \in U \\ 0 & \text{otherwise} \end{cases}.$$

Then $P(\mathcal{F}) = \mathrm{E}\left(I_{\mathcal{F}}(\mathcal{Y})\right)$ and identities and bounds on indicator functions give identities and bounds on $P(\mathcal{F})$, whatever the distribution of Y

The first step in the algebraization of coherent systems is to encode a point $\alpha = (\alpha_1, \ldots \alpha_n) \in \mathcal{Y}$ by a monomial $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, where $x = (x_1, \ldots, x_n)$ is a vector of variables. The shape of the failure set which gives the coherency property coherence property, when coded is coded into a set of monomials which defines a monomial ideal $Id_{\mathcal{F}}$ in $k[x_1, \ldots, x_n]$ (where k is a field of characteristic 0).

$$Id_{\mathcal{F}} = \langle x^\alpha : \alpha \in \mathcal{F} \rangle$$

which is equivalent to the Ideal Property

$$x^\alpha \in Id_{\mathcal{F}},\ \alpha \leq \beta \ \Rightarrow\ x^\beta \in Id_{\mathcal{F}}.$$

Conversely, any monomial ideal gives a failure set, under coherency. The minimal generating sets for the monomial ideal $Id_{\mathcal{F}}$ can be identified with the set, $\mathcal{F}^*$, of *minimal cuts*, in the reliability context. Moreover, α is a minimal cut if and only if $\alpha \in \mathcal{F},\ \beta \langle \alpha \Rightarrow \beta \notin \mathcal{F}$ and moreover $Id_{\mathcal{F}} = \langle x^\alpha \mid \alpha \in \mathcal{F}^* \rangle$.

As an example, consider just two minimal cuts, $\mathcal{F}^* = \{\beta, \gamma\}$. Then the failure ideal is $Id_{\mathcal{F}} = \langle x^\beta, x^\gamma \rangle$, and the generating function of the associated monomial set is

$$\mathcal{F}(x) = \frac{x^\alpha + x^\beta - \mathrm{lcm}(x^\alpha, x^\beta)}{\prod_{i=1}^n (1 - x_i)} = \{\alpha\}(x) + \{\beta\}(x) - \{\alpha \vee \beta\}(x). \qquad (23)$$

This represents inclusion-exclusion for the failure set of the relevant upper orthants in the original system $\mathcal{Y}$:

$$I_{Q(\alpha)\cup Q(\beta)} = I_{Q(\alpha)}(y) + I_{Q(\beta)}(y) - I_{Q(\alpha)\cap Q(\beta)}(y) = I_{Q(\alpha)}(y) + I_{Q(\beta)}(y) - I_{Q(\alpha)}(y) I_{Q(\beta)}(y),$$

where $Q(\alpha) = \{\gamma | \alpha \leq \gamma\}$, etc are the orthants. Note that if we omit the last term on the right hand side we obtain an upper bound to the indicator function which gives the elementary Bonferroni bound: $\mathrm{prob}(Q(a) \cup Q(b)) \leq \mathrm{prob}(Q(a)) + \mathrm{prob}(Q(b))$.

A little care is needed with regard to probability statements. For a particular α, $P(\alpha)$ is interpreted as $\mathrm{Prob}\{Y = \alpha\}$. This means that α occurs and no other α'. Whereas, $P(Q(\alpha))$ is the probability that α occurs *and any worse event*. In terms of cuts this means distinguishing the probability of exactly a particular cut (and nothing else) and the probability of the totality of all outcomes which include that cut in the sense of being at least as bad in terms of the $\leq$ ordering. In the binary cases, discussed below, where individual components may fail, this means $Q(\alpha)$ is all cuts which simply *include* the components indicated by α. In that case $P(Q(\alpha))$ is the *marginal* probability of the latter components being cut.

Let us take a simple network. This is directed graph $G(V, E)$ with nodes and edges with a "flow" from an input node, to an output node. We label the edges $e_i,\ 1, \ldots, d$. With each edge we associate a variable x_i. By a monomial, say $x_1 x_5 x_7$ we mean that edges e_1, e_5, e_7 have failed. A bad set of edges represented again by a monomial, is one which disconnects the flow: is there is no connected path using unfailed edges. Clearly if we have a cut and and a further edge fails then we still have cut. This is an example of a a *coherent system*: if you car has broken done and some other part fails, the car still does not work. We see that the structure of the full failure set a monomial ideal, or more precisely the square free monomials n such an ideal. Indeed, we can talk about *minimal cuts*, which correspond precisely to minimal generators.

We now summarize the application of Hilbert series and the multigraded Betti numbers to reliability.

1. If the failure set has minimal cuts indexed by monomial ideals the Hilbert series $H(I)(x)$ give an inclusion exclusion formula.

2. The minimal free resolution give the expression of the Hilbert function in terms of the multi-graded Betti numbers. These are the coefficients of each term at each level of the Hilbert series.

3. Key point: truncation of the Hilbert function at particular depths gives respectively upper and lower bounds for the indicator function for the failure event.

4. The free resolutions gives the tightest bounds relative to any other resolution

5. The Taylor resolution corrresponds to the classical Bonferroni bounds which, therefore, are typically not as strict as the minimal free resolution.

As an example consider the sequential k-out-of-n-system: at least one string of k consecutive 1's in a binary string of length k.

The Betti numbers are $\beta_0 = 5$, $\beta_1 = 7$, $\beta_2 = 4$, $\beta_3 = 1$. The multigraded Hilbert series is:

$$\begin{aligned}(R/\bar{I}_{2,6};) = &\frac{1-(xy+yz+zt+tu+uv)}{(1-x)(1-y)(1-z)(1-t)(1-u)(1-v)}\\ &+\frac{(xyuv+yzuv+tuv+xytu+ztu+yzt+xyz)}{(1-x)(1-y)(1-z)(1-t)(1-u)(1-v)}\\ &-\frac{(xytuv+yztuv+xyzuv+xyztu)}{(1-x)(1-y)(1-z)(1-t)(1-u)(1-v)}\\ &+\frac{(xyztuv)}{(1-x)(1-y)(1-z)(1-t)(1-u)(1-v)},\end{aligned}$$

Here are some notes

1. Algorithms for finding the minimal free resolution

2. Algorithms give rise to recurrence relationships for multigraded Betti numbers

3. Can use these to obtain exact formulae and bounds

4. Application to other areas: eg scan statistics

5.5 Application to 2^k factorials designs

Let $k = 4$ and take the defining contrast subgroup

$$\{I, ABCD\}$$

We take as our design D the block given by

$$\{a^2 = 1, b^2 = 1, c^2 = 1, d^2 = 1, abcd = 1\}$$

. It consists of the points:

$$(1, 1, 1), (1, -1, -1), (-1, 1, -1), (-1, -1, 1)$$

It is common to write out the aliased terms as rows in a table, with an $=$ sign. In this case we obtain

$$\begin{array}{lcl} I & = & ABCD \\ A & = & BCD \\ B & = & ACD \\ C & = & ABD \\ D & = & ABC \\ AC & = & BD \\ AB & = & CD \\ AD & = & BC \end{array}$$

Thus, in the second row, $A = BCD$ means $a \sim bc$ in that $a = bcd$ on the design, or $a - bcd \in I(D)$.

We are now in a position to investigate the effect of generating models using the algebraic method. We state some simple results.

1. Each row is an alias class. Thus, in the second row, $A = BCD$ means $a \sim bcd$ in that $a = bcd$ on the design, or $a - bcd \in I(D)$.

2. For any monomial ordering, the models selects the lowest term in the alias table wrt the ordering

3. The Normal Form of any monomial is a monomial

4. G-basis elements only have two terms

5. There can be at most one square-free leading term in each alias class: therefore number of leading terms is bounded by $\leq 2^{k-r} + k$.

For the standard term ordering `DegRevLex` in CoCoA, the design identifies the model with terms

$$1, a, b, c, d, ad, bd, cd$$

, see Figure 1(a). We can confirm that each term is associated with one row of the alias table.

For the standard term ordering in CoCoA(`DegRevLex`), the design identifies the model with terms $1, a, b, c, d, ad, bd, cd$, see Figure 1(a). For this design, its algebraic fan comprises twelve models which belong to only three classes according to the simplicial structure of the model (up to permutations of variables). Representative models of each class are shown in Figure 1, and each class is identified from 1(a) to 1(c). Models of the classes 1(a) and 1(b) have the same Hilbert series $HS(s)$, while 1(c) has different Hilbert series. A Schlegel diagram of the state polytope of the design ideal is shown in Figure 3.

Table 1 summarizes the results for each class, where in each class, the model is a simplicial complex Δ; Betti numbers were computed for the Stanley-Reisner ideal I_Δ (squarefree ideal of simplicial non terms in the model), for its Artinian closure $\bar{I}_\Delta = I_\Delta + \langle a^2, b^2, \ldots \rangle$ and for the quotient ring $R/\bar{I}_\Delta$

Class	1(a)	1(b)	1(c)
$HS(s)$	$1 + 4s + 3s^2$		$1 + 3s + 3s^2 + s^3$
$\beta(\bar{I}_\Delta)$	0 1 2 3 2 : 7 8 3 3 : 6 8 3 Total 7 14 11 3	0 1 2 3 2 : 7 9 4 1 3 : 1 7 9 3 Total 8 16 13 4	0 1 2 3 1 : 1 2 : 3 3 3 : 3 3 4 : 1 1 Total 4 6 4 1
$\beta(R/\bar{I}_\Delta)$	0 1 2 3 4 0 : 1 1 : 7 8 3 2 : 6 8 3 Total 1 7 14 11 3	0 1 2 3 4 0 : 1 1 : 7 9 4 1 2 : 1 7 9 3 Total 1 8 16 13 4	0 1 2 3 4 0 : 1 1 1 : 3 3 2 : 3 3 3 : 1 1 Total 1 4 6 4 1
$\beta(I_\Delta)$	0 1 2 : 3 2 Total 3 2	0 1 2 2 : 3 3 1 3 : 1 1 Total 4 4 1	0 1 : 1 Total 1

Table 1: Betti numbers for simplicial models of Example 5.5.

Consider the fractional factorial design with generator $I = ABCD = CDEF$. Its algebraic fan consists of 132 models which belong to six simplicial

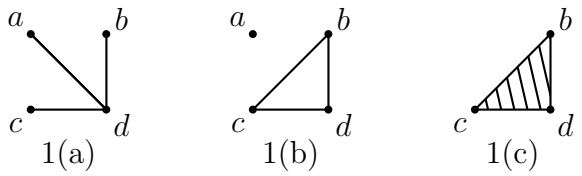

Figure 1: Representatives of simplicial models identified by the design in Example 5.5.

classes (up to permutation of variables). Representatives of each class are depicted in Figure 2. Classes 2(b)and 2(c) have the same Hilbert series, while classes 2(d), 2(e) and 2(f) share a common Hilbert series. Table 2 shows the Betti numbers for $\bar{I}_\Delta$ and for I_Δ for each class. See Figure 3 for a Schlegel diagram of the state polytope of the design ideal.

Class	$HS(s)$	$\beta(\bar{I}_\Delta)$	$\beta(I_\Delta)$
2(a)	$1+4s+6s^2+4s^3+s^4$	0 1 2 3 4 5 1 : 2 1 2 : 4 8 4 3 : 6 12 6 4 : 4 8 4 5 : 1 2 1 Total 6 15 20 15 6 1	0 1 1 : 2 1 Total 2 1
2(b)	$1+5s+7s^2+3s^3$	0 1 2 3 4 5 1 : 1 2 : 8 17 13 5 1 3 : 1 15 32 25 8 1 4 : 1 8 16 12 3 Total 10 33 53 46 21 4	0 1 2 3 1 : 1 2 : 3 6 4 1 3 : 1 2 1 Total 5 8 5 1
2(c)		0 1 2 3 4 5 1 : 1 2 : 8 16 11 3 3 : 13 29 22 6 4 : 6 14 11 3 Total 9 29 46 39 17 3	0 1 2 1 : 1 2 : 3 5 2 Total 4 5 2
2(d)	$1+6s+7s^2+2s^3$	0 1 2 3 4 5 2 : 14 32 33 20 7 1 3 : 2 21 46 40 15 2 4 : 3 8 7 2 Total 16 53 82 68 29 5	0 1 2 3 4 2 : 8 16 14 6 1 3 : 2 5 4 1 Total 10 21 18 7 1
2(e)		0 1 2 3 4 5 2 : 14 31 31 18 6 1 3 : 1 19 44 39 15 2 4 : 3 8 7 2 Total 15 50 78 65 28 5	0 1 2 3 4 2 : 8 15 12 5 1 3 : 1 3 3 1 Total 9 18 15 6 1
2(f)		0 1 2 3 4 5 2 : 14 30 28 12 2 3 : 16 38 35 14 2 4 : 3 8 7 2 Total 14 46 69 55 23 4	0 1 2 3 2 : 8 14 9 2 Total 8 14 9 2

Table 2: Betti numbers for simplicial models of Example 5.5.

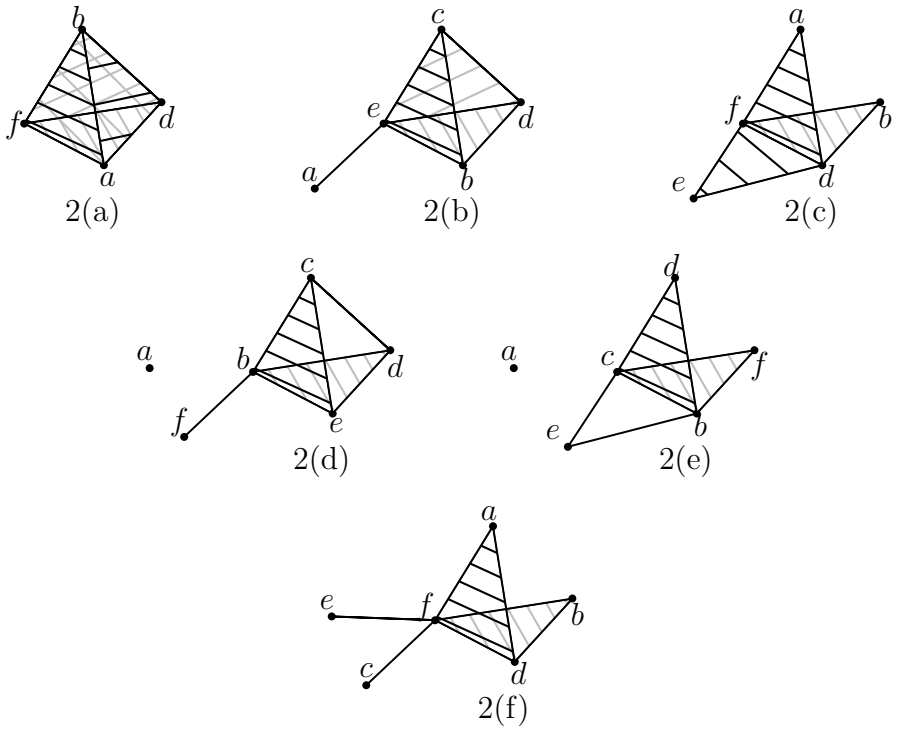

Figure 2: Representatives of simplicial models identified by the design in Example 5.5.

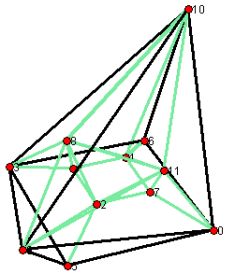

Figure 3: Schlegel diagram of the state polytope of the design ideals of Examples 5.5 (left) and 5.5 (right).