1. Modelos Loglineales tablas de 2 entradas

Los modelos loglineales para tablas de 2 entradas, describen las asociaciones entre dos variables **discretas** digamos X y Y. El modelo loglineal nos dice cuan grande es el conteo de la celda dependiendo de los niveles de las dos variables categóricas.

Las frecuencias de la tabla pueden calcularse como $\mu_{ij} = n\pi_{ij}$ y si consideramos que las variables X y Y son **independientes** se tiene que $\pi_{ij} = \pi_{i\bullet}\pi_{\bullet j}$ entonces

$$\mu_{ij} = n\ \pi_{i\bullet}\pi_{\bullet j}$$

.

Si sacamos logaritmo de ambos lados de la igualdad tendremos

$$\log(\mu_{ij}) = \log(n) + \log(\pi_{i\bullet}) + \log(\pi_{\bullet j})$$

el modelo podemos verlo como:

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$$

Es importante hacer notar que en el conjunto de los términos $\{\lambda_i^X\}$hay un término redundante y sólo $I-1$ de ellos son desconocidos. Análogamente para el conjunto $\{\lambda_j^Y\}$ hay un término redundante y solo hay $J-1$ parámetros desconocidos. Puede haber distintas parametrizaciones y no hay un único conjunto de parámetros. Deben imponerse diferentes condiciones para obtener los parámetros.

Una manera de imponer condiciones por ejemplo es hacer $\lambda_1^X = \lambda_1^Y = 0$, otra es $\sum_i \lambda_i^X = \sum_j \lambda_j^Y = 0$.

1.1. Interpretación

En el caso de tener un modelo de independencia en una tabla de Ix2, para la fila i el logit es

$$\text{logit}(P(Y=1|X=i)) = \log(\frac{P(Y=1|X=i)}{P(Y=2|X=i)}) = \log(\frac{\mu_{i1}}{\mu_{i2}})$$

$$= \log(\mu_{i1}) - \log(\mu_{i2}) = \lambda + \lambda_i^X + \lambda_1^Y - \lambda - \lambda_i^X - \lambda_2^Y = \lambda_1^Y - \lambda_2^Y$$

Este último término NO depende de i, así que es el mismo para todos los niveles de i. Entonces $\exp(\lambda_1^Y - \lambda_2^Y)$ son los momios de la columna 1 vs la columna 2.

Esto puede generalizarse para $J > 2$, la diferencia de los parámetros para una variable $(\lambda_k^Y - \lambda_m^Y)$están relacionados con el logaritmo de los momios de una categoría con otra.

Ahora si **no hay independencia** entre X y Y se debe ajustar el modelo saturado:

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

las λ_{ij}^{XY} son los términos de asociación y representan la interacción entre X y Y.

NOTA

Las restricciones de estimabilidad pueden cambiarse, así en vez de que todas las $\lambda_{Ij} = \lambda_{iJ} = 0$ se puede hacer que $\sum_i \lambda_{ij}^{XY} = \sum_j \lambda_{ij}^{XY} = 0$ para toda i y j. Cada paquete usa diferentes restricciones. Los que sí son únicos son los contrastes como este: $\lambda_{11}^{XY} + \lambda_{22}^{XY} - \lambda_{12}^{XY} - \lambda_{21}^{XY} = log(\theta)$, que determina el cociente de momios en una tabla de 2x2.

2. Modelos Loglineales tablas de contingencia de 3 entradas

Suponemos que tenemos tres variables categóricas X,Y y Z.

El modelo saturado es este:

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}.$$

Al igual que en las tablas de dos entradas, para estimar los parámetros pueden usarse las condiciones simétricas o las basadas en las categorías de referencia, considerando el esquema de muestreo del que se trate, incluso puede haber más restricciones (fijar el total $\sum_i \sum_j \sum_k \lambda_{ijk}^{XYZ} = n_{\bullet\bullet\bullet}$, o las marginales $\sum_j \sum_k \lambda_{ijk}^{XYZ} = n_{i\bullet\bullet}$ o $\sum_k \lambda_{ijk}^{XYZ} = n_{ij\bullet}$)

Los modelos loglineales que se pueden interpretar en términos de sus independencias (condicionales) son los que nos resultan interesantes y estos son precisamente los **modelos jerárquicos**. Un modelo es llamado jerárquico cuando incluye todos los términos de menor orden, derivados de los de mayor orden. Por ejemplo si el modelo contiene a λ_{ij}^{XY} también contendrá a λ_i^X y a λ_j^Y.

La mayoría de los modelos jerárquicos loglineales para tablas de 3 entradas también son **modelos gráficos**. La idea es esta, se hace una gráfica, donde los puntos son las variables y los puntos se unen con una arista si el modelo contiene el término de la interacción. Si la gráfica resulta disconexa, los grupos de variables separados resultan independientes. Si dos variables resultan conectadas a través de una tercera, entonces estas son condicionalmente independientes, dada la tercera.

2.1. Tipos de independencia

- **Independencia mutua**

  Todas las variables son mutuamente independientes,

  $$\pi_{ijk} = \pi_{i\bullet\bullet}\pi_{\bullet j\bullet}\pi_{\bullet\bullet k}$$

  El modelo se denota como (X,Y,Z), solo tiene los efectos principales.

  $$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$$

- **Independencia Conjunta**

  Dos variables son conjuntamente independientes de la tercera,

  $$\pi_{ijk} = \pi_{ij\bullet}\pi_{\bullet\bullet k}$$

  El modelo se denota (XY,Z), tiene sólo la interacción.

  $$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$$

- **Independencia Marginal**

  Dos variables son independientes ignorando la tercera, es decir la tabla se **colapsa** sumando sobre la tercera y ahora se tendrá una tabla de dos entradas.

  $$\pi_{ij\bullet} = \pi_{i\bullet\bullet}\pi_{\bullet j\bullet}$$

  El modelo se denota se denota (X,Y)

  $$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y$$

- **Independencia Condicional**

  Dos variables son independientes dada una tercera,

$$\pi_{ij|k} = \pi_{i\bullet|k}\pi_{\bullet j|k}$$

  El modelo se denota (XZ,YZ), tiene dos interacciones.

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + +\lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

- **Asociación Homogénea**

$$\pi_{ijk} = \psi_{ij}\phi_{jk}\omega_{ik}$$

  El modelo se denota (XY,XZ,YZ), tiene tres interacciones

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$$

  Este es un caso especial que no corresponde a un modelo gráfico.

3. Relaciones entre Independencias

Independencia Mutua $\Rightarrow$ Independencia Conjunta. Todas las variables son independientes entre si.

Independencia Conjunta $\Rightarrow$ Independencia Marginal; una variable es independiente de las otras dos

OJO

Independencia Marginal $\neg\Rightarrow$ Independencia Conjunta

Independencia Marginal $\neg\Rightarrow$ Independencia Condicional

Independencia Condicional $\neg\Rightarrow$ Independencia Marginal

4. Relación de Modelos Loglineales y Logit

Con los modelos loglineales se modela cómo los conteos de las celdas dependen de los distintos niveles de las variables categóricas. Se modelan las asociaciones y las interacciones entre las variables categóricas. Son muy adecuados cuando no hay distinción entre variable respuesta y explicativa, o cuando hay más de dos variables respuestas. Cuando una de ellas es considerada como respuesta, existe una equivalencia entre modelos loglineales y los modelos logit que a continuación se enumeran

Se hará a través del ejemplo de Admisiones a Berkley (ver ejemplo de clase). Se pueden considerar todas relaciones entre:

A = Admisión, D = Departmento and S = Sexo.

Al considerar a **A como variable respuesta** y a D y a S como covariables, se tendrán los siguientes modelos logit:

1. logit model para A con solo el término constante;

2. logit model para A con el efecto principal de D;

3. logit model para A con el efecto principal de S;

4. logit model para A con los efectos principales de D y S; y

5. logit model para A con los efectos principales de D y S más la interaccion D x S .

Que corresponderán a los siguientes modelos loglineales:

1. Modelo de **independencia conjunta** de D y S con A (DS, A), que indica que ni D ni S tienen un efecto sobre A, es equivalente a un modelo logit con la constante;

2. Modelo de **independencia condicional** (DS, DA), que indica que el sexo no tiene efecto en A, después de que el efecto de departamento es incluido, es equivalente a un modelo logit con respuesta A y un efecto principal para D;

3. Otro modelo de **independencia condicional** es (DS, SA), que indica que el departamento no tiene efecto en A, después de que el efecto de sexo es incluido, que es equivalente a un modelo logit con respuesta A y un efecto principal para S;

4. El modelo **sin triple interacción** (DS, DA, SA) o de **asociación homogénea**, indica que el efecto del sexo sobre A es el mismo en cada nivel de D y es equivalente a un modelo logit para A con efectos principales de S y D;

5. El modelo con **triple interacción o saturado** (DSA), indica que el efecto del sexo sobre A varía en cada departamento y es equivalente a un modelo logit para A con efectos principales de S y D y la interacción S x D.

Equivalente significa que los dos modelos tienen la misma medida de bondad de ajuste relativa al modelo saturado y además que arrojan los mismos conteos esperados para cada celda.

Los modelos loglineales no son lo mismo que los modelos logit puesto que los loglineales describen la distribución conjunta de las tres variables, mientras que los logit sólo describen la distribución condicional de A dados S y D. Los modelos loglineales tienen más parámetros que los logit y en los logit los parámetros de la distribución conjunta de S y D nos son de interés.