Modelo Lineal Generalizado

Continuación TP 2

9. Consideremos la siguiente tabla

n_{11}	n_{12}	n_{1+}
n_{21}	n_{22}	n_{2+}
n_{+1}	n_{+2}	n

Cuadro 1: Marginales Fijas

a) Supongamos que todas las marginales n_{i+} y n_{+j} son fijas. Pruebe que el odd ratio $\theta = \dfrac{n_{11}n_{22}}{n_{12}n_{21}}$ es una función creciente de n_{11}.

b) Supongamos ahora que el muestreo es producto multinomial, es decir $n_{11} \sim Bi(n_{1+}, \pi_{1|1})$ y $n_{21} \sim Bi(n_{2+}, \pi_{1|2})$ independientes. Pruebe que $n_{11} = t|_{n_{11}+n_{21}=s}$ tiene distribución multinomial. De qué parámetros?

c) Se realizó un experimento psicológico con el fin de investigar el efecto de la ansiedad sobre el deseo de un individuo de permancer solo o acompañado (Schacter, 1959 ; Lehman, 1975). De un grupo de 30 individuos se formaron al azar dos grupos: uno de 13 individuos y otro de 17. A los todos los individuos se les dijo que recibirían un tratamiento con un electro shock, pero a un grupo se le dijo que el tratamiento sería doloroso, mientras que al otro se le dijo que sería indoloro. El primer grupo fue el de *alta ansiedad*, mientras que el segundo fue el de *baja ansiedad*. A todos se les dijo que tendrían 10 minutos de espera antes del tratamiento y a cada uno se le dio la posibilidad de elegir si deseaba esperar solo o acompañado. A través del test de exacto de Fisher testee si hay diferencia entre los dos grupos.

   Los resultados se muestran en la siguiente tabla

	Espera Acompañado	Espera Solo
Alta Ansiedad	12	5
Baja Ansiedada	4	9

Cuadro 2: Marginales Fijas

Modelo Lineal Generalizado

Trabajo Práctico 3

1. Sea X una variable con distribución Gamma de parámetros α y μ, $\Gamma(\alpha, \mu)$, donde $\mu = E(X)$.

Demuestre que la distribución de X pertenece a una familia exponencial. Identifique al parámetro canónico θ y a las funciones a, b y c.

Parece razonable usar el link canónico en esta familia? Tendría alguna desventaja?

2. Decimos que X tienen distribución gaussiana inversa de media μ y parámetro ϕ si su densidad es de la forma

$$f(x) = \left[\frac{\phi}{2\,\Pi\, y^3}\right]^{\frac{1}{2}} \exp\left[\frac{-\phi(y-\mu)^2}{2\,\mu^2\, y}\right] I_{[0,\infty)}(y)\,.$$

Esta familia se utiliza , por ejemplo, para el estudio del movimiento Browniano de partículas y en análisis de regresión con datos fuertemente asimétricos. Por ejemplo, suele ser utilizada para modelar tiempos de sobrevida, tiempos de espera, etc.

a) Grafique la función de densidad para distintos valores de los parámetros.

b) Demuestre que la distribución de X pertenece a una familia exponencial. Identifique al parámetro canónico θ y a las funciones a, b y c.

3. Decimos que X tienen distribución de Weibull de parámetros α y β si su densidad es de la forma

$$f(x) = \frac{\alpha\, y^{\alpha-1}}{\beta^\alpha} \exp\{-[\frac{y}{\beta}]^\alpha\} I_{[0,\infty)}(y)\,.$$

Verifique si esta familia de distribuciones es una familia exponencial para algún valor de α.

4. Sea $f_o(y)$ una función de densidad o de probabilidad arbitraria con función generadora de momentos dada por

$$M(\xi) = E\{\exp(\xi Y)\} = \exp\{b(\xi)\}\,,$$

que es finita para un rango de valores de ξ que incluye el 0. Consideremos la densidad *exponencial pesada* dada por

$$f_{Y,\theta}(y) \propto \exp(\theta y)\, f_o(y)\,.$$

Derive la constante normalizadora para esta densidad y muestre que $f_{Y,\theta}(y)$ pertenece a una familia exponencial con $a(\phi) = 1$.

5.

a) Sean $Y|\lambda \sim P(\lambda)$ y $\lambda \sim \Gamma(\alpha, \beta)$, donde $\Gamma(\alpha, \beta)$ es la distribución Gamma con esperanza $\alpha\beta$ y varianza $\alpha\beta^2$, cuya densidad está dada por

$$f(\lambda) = \frac{1}{\Gamma(\alpha)\, \beta^{\alpha}}\, \lambda^{\alpha-1} e^{-y/\beta} I_{[0,\infty)}(y)\,.$$

Pruebe que Y tienen distribución binomial negativa con función de probabilidad puntual

$$P(Y = y) = \frac{\Gamma(\alpha + y)}{\Gamma(\alpha)\; y} \left(\frac{\beta}{1+\beta}\right)^{y} \left(\frac{1}{1+\beta}\right)^{\alpha}\,.$$

b) La distribución binomial negativa dada en el item anterior suele parametrizarse en términos de $\mu = \alpha\beta$ y $\kappa = 1/\alpha$ como

$$P(Y = y) = \frac{\Gamma(\kappa^{-1} + y)}{\Gamma(\kappa^{-1})\; y!} \left(\frac{\kappa\beta}{1+\kappa\beta}\right)^{y} \left(\frac{1}{1+\kappa\beta}\right)^{1/\kappa}\,.$$

Verifique si para cada κ fijo esta distribución pertenece a una familia exponencial. En caso afirmativo, cuánto valdría el parámetro canónico θ?

6. Consideremos el caso en que $E(Y_i) = \mu_i = \mathbf{x}_i'\boldsymbol{\beta}$, siendo $Y_i \sim N(\mu_i, \sigma^2)$, es decir tomando como función link la identidad.

a) Deduzca el método de Fisher–scoring para este caso. Especifique la variable de trabajo **z** y la matriz **W** utilizadas en el método **IRWLS**. Qué diferencias tiene el estimador resultante con el de mínimos cuadrados?

b) Calcule la deviance.

c) Calcule la matriz de información y la matriz de covarianza de $\widehat{\boldsymbol{\beta}}$.

d) Compruebe que la distribución de $\widehat{\boldsymbol{\beta}}$ y de la deviance pueden determinarse exactamente en este caso.

7. Supongamos que $Y_i \sim N(\mu_i, \sigma^2/w_i)$, donde w_i son pesos conocidos, $1 \leq i \leq n$, y que $log(\mu_i) = \mathbf{x}_i'\boldsymbol{\beta} = \eta_i$.

a) Deduzca la variable de trabajo **z** y la matriz **W** utilizadas en el método **IRWLS** para este caso. Es necesario conocer a σ^2 para estimar a $\boldsymbol{\beta}$?

b) Cómo estimaría a σ^2 ? Es su propuesta insesgada?

8. Supongamos un modelo de regresión logística en el que $Y_i \sim Bi(n_i, \Pi_i)$, donde $log\left(\frac{\Pi_i}{1-\Pi_i}\right) = \mathbf{x}_i'\boldsymbol{\beta} = \eta_i$, $1 \leq i \leq n$.

Una función no lineal que interesa estimar en este caso es $\Pi_i = \Pi_i(\boldsymbol{\beta}) = \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}}}{1+e^{\mathbf{x}_i'\boldsymbol{\beta}}}$.

Proponga un estimador para Π_i. Cuál sería la distribución asintótica de dicho estimador? Cómo estimaría la varianza asintótica de la probabilidad estimada?

Deduzca un intervalo de nivel aproximado $1-\alpha$ para Π_i.