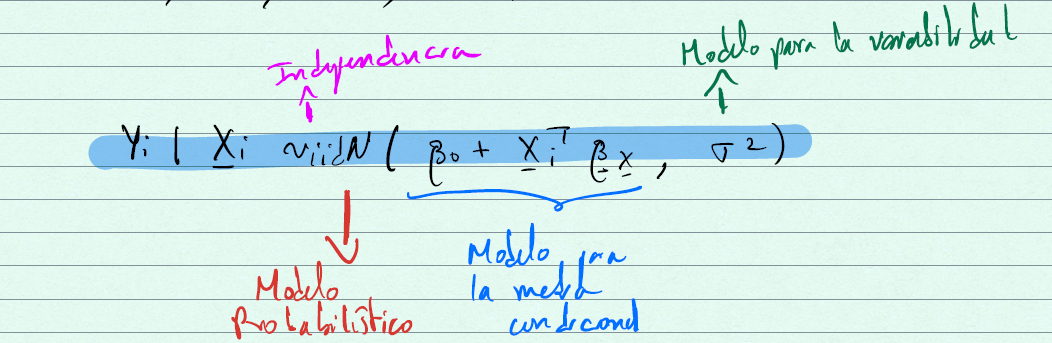
Modelo de Regresión Líneal general

L = { (X1, Y1), … , (Xn, Yn) }



Modelo de regresión lineal generalizado

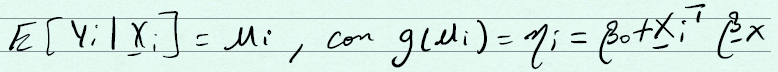
L = { (X1, Y1), … , (Xn, Yn) }

Modelo probabilístico: Distribuciones de la clase exponencial.

Ejemplos: Normal, binomial, Poisson, Gamma.

Modelo para la media :





monótona y diferenciable.

–



Ejemplo:

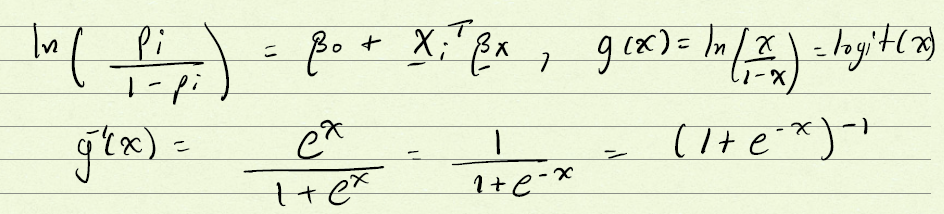






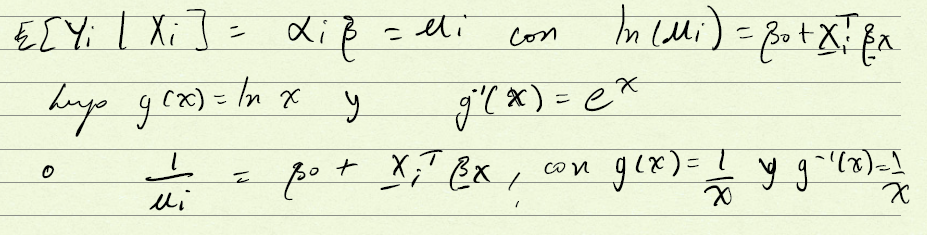
Ejemplo: :





Ejemplo:





¿Por qué la clase exponencial?

Estas distribuciones se dejan parametrizar de una manera que permita calcular la función de costo, es derivada y el Hessiano de una manera fácil.

Estimación: Ejemplo

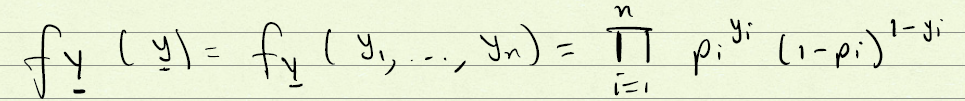
Sea { (X1, Y1), … , (Xn, Yn) } y con :, independientes y i=1, . . . , n

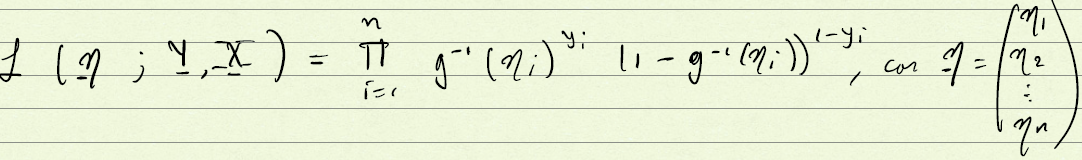


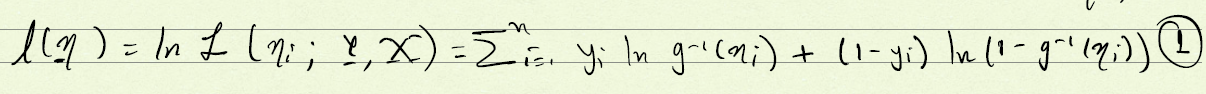
Como su función de masa está dada por ,

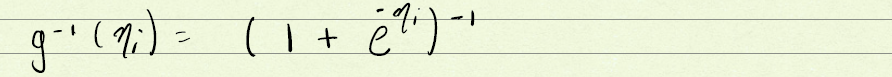


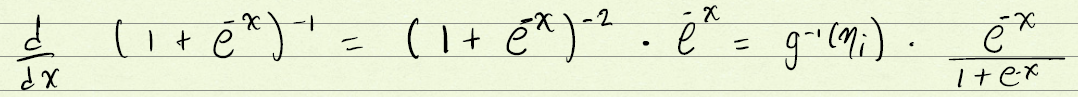
La función de masa de la muestra es el producto de las funciones de masa (por dependencia)

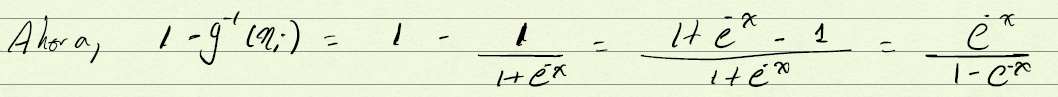




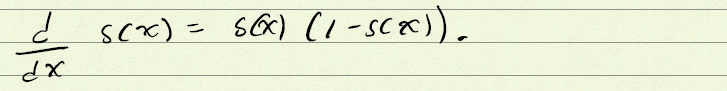


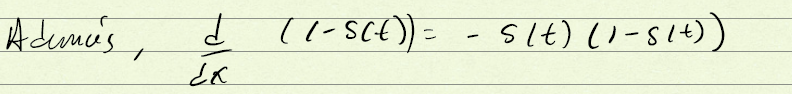












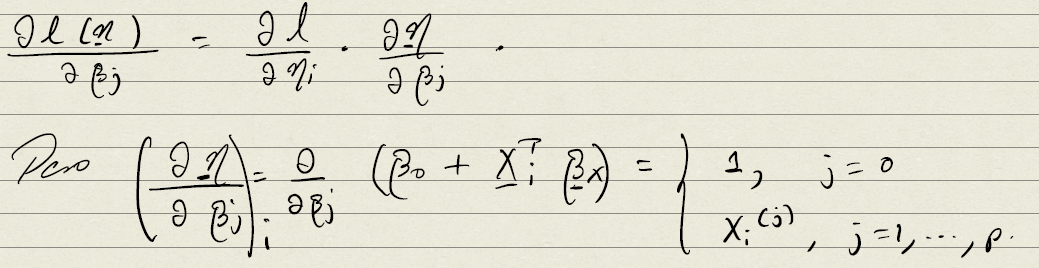
Asi, si S(X) = (1-e-X)-1, entonces .

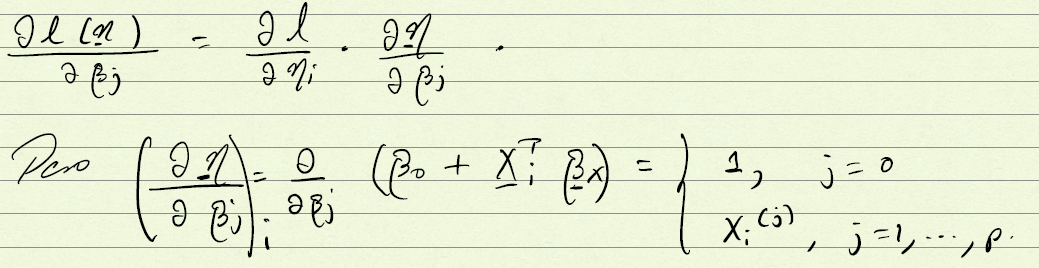


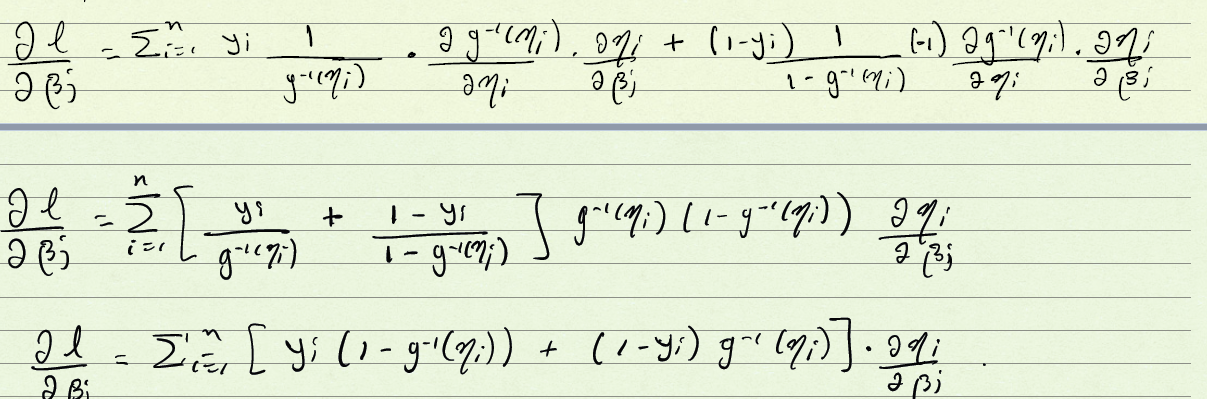
Además, .



Para maximizar con respecto a debemos hallar



Pero 

Asi, 

Asi, se forma el vector



Con , si , ,

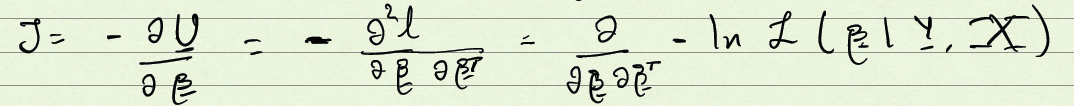




Entonces

A se le llama el vector Score. El estimador de máxima verosimilitud de β , β ̂, es la solución del sistema , que tiene p+1 ecuaciones y p+1 incognitas. Se puede mostrar que .

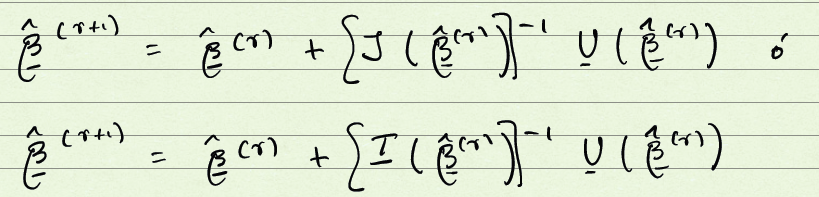
La matriz Hessiana es , pero se prefiere



Nota: Nos permitimos el abuso de notación , para indicar la misma cantidad, aunque realmente n es una función β.

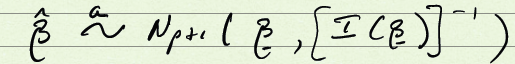
Se define la matriz de información como se puede mostrar que (la inversa de la información).

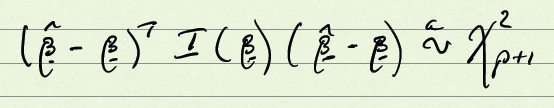
Para resolver el sistema se puede usar métodos iteractivos con J o con I

Esto requiere tener. I es una propiedad del modelo y J tiene una estrecha más compleja en general. Se puede mostrar que para el método de máxima verosimilitud.

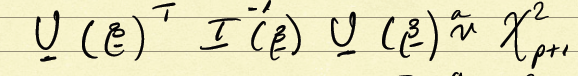
1. Si F es una función uno a uno



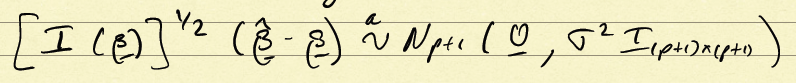
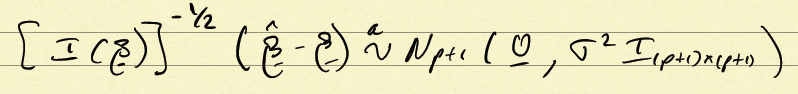
2. y oportunamente.



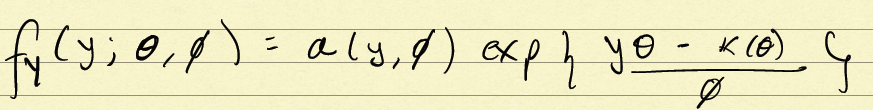
Se tiene tres estadísticas:

1. Wald: -
2. Score: -
3. Test de razón de verosimilitud: 

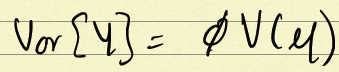
Para los test de Wald y Score, además se tiene:

1. Wald: 
2. Score: 

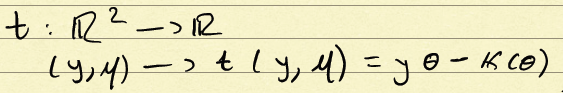
Un criterio importante de ajuste es el deviance. Esta cantidad surge al expresar el modelo de la familia exponencia en la forma

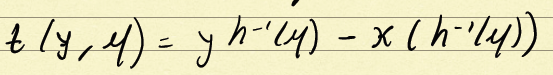


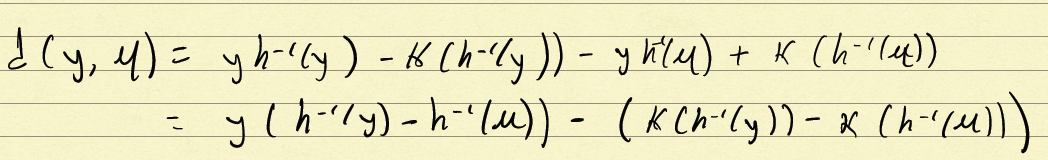
Donde es una función normalizada, se llama la función constante (Cumulant function), es el parámetro canónico y es la dispersión. Se puede mostrar que , con la monótona creciente.

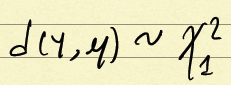
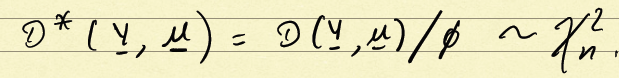
Además y .

A partir de esta parametrizada se define la función

, donde ϴ a una función de y que es el valor esperado de y.

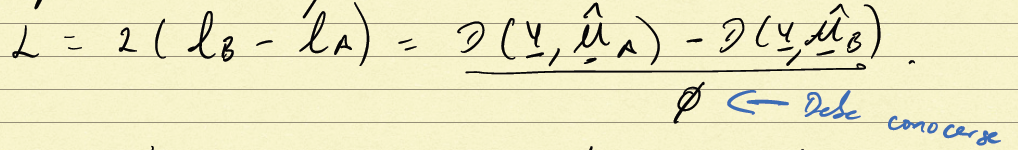
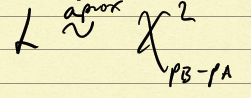
Asi,  El deviance se define como 



Se puede mostrar que . Ahora se define el deviance total como , donde los wi son ponderadores para cada observación. El deviance escalado es .

Si los modelos A y B tienen y parámetros, con < - y B contiene todas las variables que tienen el modelo A, se dice que A esta anidado en B. B es más complejo que A.

Para evaluar si la complejidad de B le representa un beneficio con respecto a A se utiliza

Si H0: A y B son iguales VS Ha: A y B son diferentes, entonces. Si se rechaza H0, se prefiere a B sobre A.

Cuando ф no es conocido (es decir, siempre) se lleva a cabo la prueba anterior usando un estadístico F.

Ejemplo: Registro de Autos.