

Modelos de Pronóstico de Resultados

Rómulo A. Chumacero

Motivación

- Abundante información de resultados de partidos
- Nos interesa evaluar si existen maneras exitosas para predecir resultados o entender el juego
- Los determinantes de resultados pueden ayudar a elaborar mejores estrategias, definir contrataciones, etc.

Algunos ejemplos

- NBA: ¿Qué es más rentable, lanzar por 3 puntos o por dos puntos?
 - Probabilidad 3 puntos: 37%, Probabilidad 2 puntos: 50%
 - Porcentaje de tiro efectivo: $3 \times .37 = 1.11 > 2 \times .50 = 1$
- NFL: ¿Qué hacer en el cuarto down?
 - Patriots y Colts (2009)
 - 4 y 2 con 2 minutos
 - Elección: Patear o tratar por 6 puntos (60%)

Algunos ejemplos: Regla de $0 > 1$

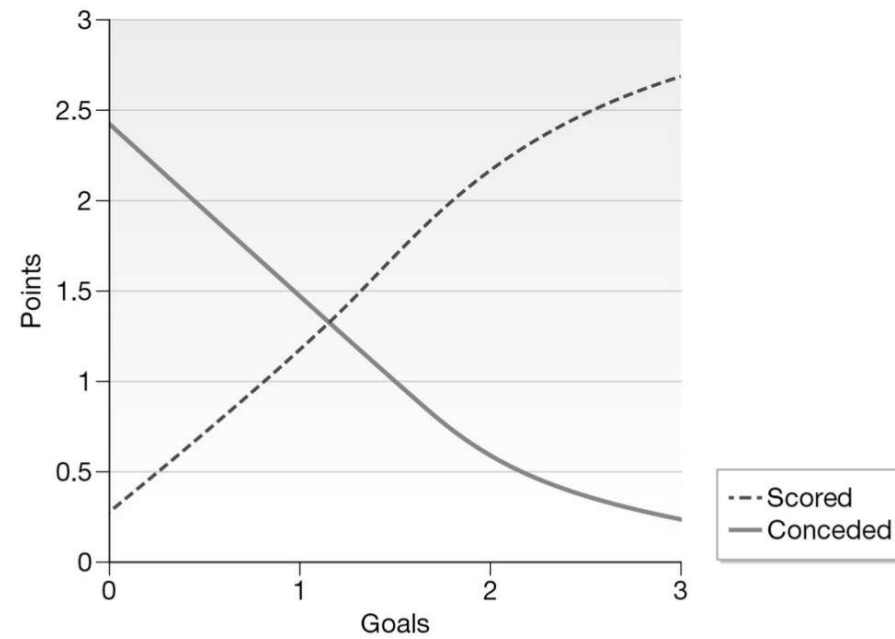


Figure 1:

Algunos ejemplos: El gol más valioso

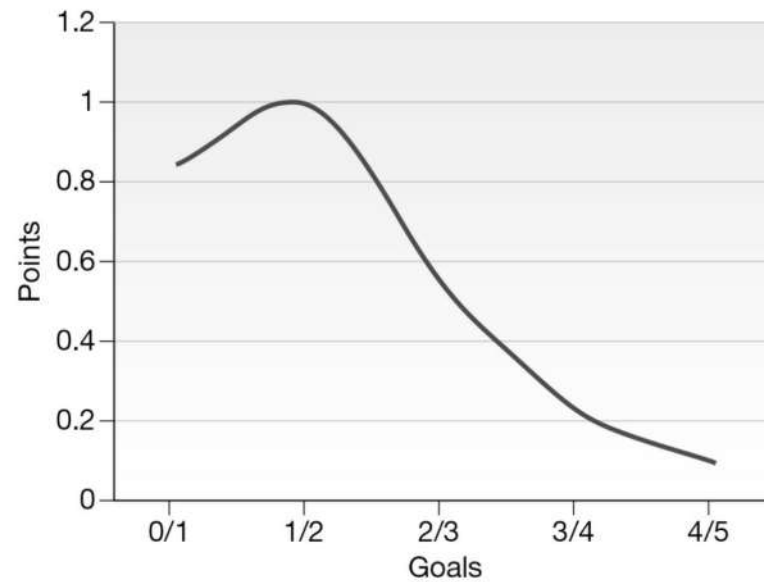


Figure 20 Marginal points produced by goals

Figure 2:

Las preguntas habituales

- ¿De qué depende el resultado específico de un partido (goles de local y de visita)?
- ¿De qué depende el resultado (ganar, empatar, perder)?
- Para responder estas preguntas utilizamos modelos econométricos que se acomodan a la naturaleza del problema

Resultados específicos

- Goles de local y goles de visita
- Consideradas como variables aleatorias
- Variables aleatorias discretas
- Desde 0 hasta ¿infinito?
- Elección de distribución apropiada

Resultado general

- Gana local, empate, gana visita
- Resultado considerado como variable aleatoria
- Discreta y ordenada
- Elección de distribución apropiada

Procedimiento

- Determinación del modelo de interés
- Proponer supuesto distributivo para variables de interés
- Obtener datos
- Estimar el modelo usando QMLE
- Testear restricciones
- Evaluar el modelo

Introducción a MLE

- $f(y; \theta_0)$ describe la probabilidad de ocurrencia de y dado θ_0 . En la práctica, tenemos muestra $\{y_1, \dots, y_T\}$ pero desconocemos θ_0
- La función de verosimilitud muestral describe la situación en que se trata y de $f(\cdot)$ como dado y a θ_0 como una variable
- En este cambio de roles, buscamos el vector de parámetros que maximiza la probabilidad de ocurrencia de la muestra

$$\begin{aligned} L(\theta; y) &\equiv f(y; \theta) \\ \ell(\theta; y) &= \ln L(\theta; y) \end{aligned}$$

- Propiedades: consistencia, normalidad asintótica, eficiencia asintótica, invariabilidad a reparametrizaciones
- No deseadas: dependencia en supuesto distributivo (QMLE) y propiedades de muestras pequeñas (sesgo)

La función de verosimilitud

$$\ell(\theta; Y) = \ln \prod_{t=1}^T f(y_t; \theta) \quad (1)$$

$$= \sum_{t=1}^T \ell(\theta; y_t)$$

$$\hat{\theta}_{MLE} = \arg \max_{\theta \in \Theta} \frac{1}{T} \sum_{t=1}^T \ell(\theta; y_t | x_t)$$

- Equivalente a maximizar

$$\prod_{t=1}^T L(\theta; y_t | x_t) \quad \text{or} \quad \sum_{t=1}^T \ell(\theta; y_t | x_t)$$

Inferencia

- Puede demostrarse que:

$$\begin{aligned}\hat{\theta} &\overset{a}{\sim} \mathcal{N}(\theta_0, H_0^{-1} O_0 H_0^{-1}) \\ O_0 &= \mathcal{E} \left[\frac{\partial \ell}{\partial \theta} \Big|_{\theta_0} \frac{\partial \ell}{\partial \theta'} \Big|_{\theta_0} \right] = \mathcal{E} [g_0 g_0'] \quad H_0 = \mathcal{E} \left[\frac{\partial^2 \ell}{\partial \theta \partial \theta'} \Big|_{\theta_0} \right] \\ \hat{O} &= \frac{1}{T} \sum_{t=1}^T \left[\frac{\partial \ell_t}{\partial \theta} \Big|_{\hat{\theta}} \frac{\partial \ell_t}{\partial \theta'} \Big|_{\hat{\theta}} \right] \quad \hat{H} = \left[\frac{\partial^2 \ell}{\partial \theta \partial \theta'} \Big|_{\hat{\theta}} \right]\end{aligned}$$

- Si el modelo está bien especificado:

$$\begin{aligned}\hat{\theta} &\overset{a}{\sim} \mathcal{N}(\theta_0, -H_0^{-1}) \\ \hat{\mathcal{V}}(\hat{\theta}) &= -\hat{H}^{-1} \quad \text{o} \quad \hat{\mathcal{V}}(\hat{\theta}) = (T\hat{O})^{-1}\end{aligned}$$

- La última expresión no requiere computar y evaluar segundas derivadas
- Existen tests para evaluar si $-\hat{H}$ es estadísticamente distinta a $T\hat{O}$ (Test de Matriz de Información)
- Más poderoso pero costoso:

$$2 \left[\ell(\hat{\theta}) - \ell(\bar{\theta}) \right] \xrightarrow{D} \chi_q^2$$

Newton-Raphson

- Considera una aproximación de Taylor de 2do orden:

$$\ell(\theta) \cong \ell(\theta_1) + g'_1(\theta - \theta_1) + \frac{1}{2}(\theta - \theta_1)' H_1(\theta - \theta_1), \quad g_1 = \left. \frac{\partial \ell}{\partial \theta} \right|_{\theta_1} \quad H_1 = \left. \frac{\partial^2 \ell}{\partial \theta \partial \theta'} \right|_{\theta_1}$$

- Conditional en θ_1 , la próxima iteración es: $\theta_2 = \theta_1 - H_1^{-1} g_1$
- Repetir hasta converger ($\theta_j \cong \theta_{j-1}$) y hacer $\theta_j = \hat{\theta}_{ML}$
- Componentes: *dirección* ($\delta_j = \theta_j - \theta_{j-1}$), *ángulo*: Inversa de Hessiano y *tamaño*: gradiente

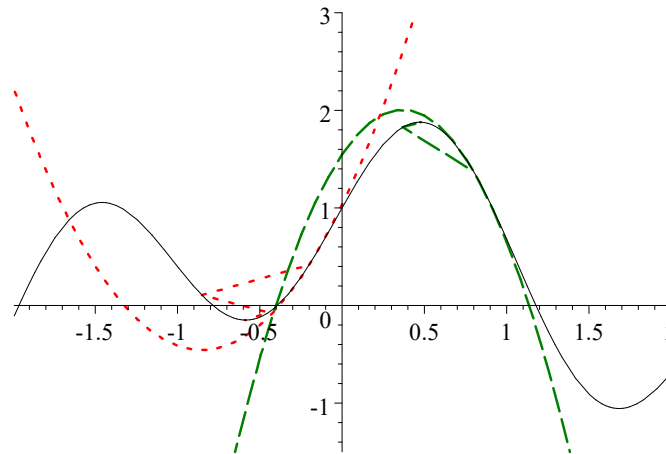


Figure 3: Newton-Raphson Method

Ejemplo 1 (elección binaria)

$$y_i = \begin{cases} 1 & \text{gana local} \\ 0 & \text{en otro caso} \end{cases}$$

- Determinantes:

$$y_i = \beta' x_i + u_i \tag{2}$$

- Alternativamente:

$$y_i^* = \beta' x_i + u_i$$

$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \tag{3}$$

$\beta' x_i$ se denota por una función índice

$$\begin{aligned} \Pr(y_i = 1 | x_i) &= \Pr(y_i^* = \beta' x_i + u_i > 0) \\ &= \Pr(u_i > -\beta' x_i) \\ &= 1 - \mathcal{F}(-\beta' x_i) \end{aligned}$$

- Dado $\mathcal{E}(u) = 0$, si $f(\cdot)$ es simétrica, $1 - \mathcal{F}(-\beta' x_i) = \mathcal{F}(\beta' x_i)$

Formas funcionales

$$\mathcal{F}(\beta'x_i) = \int_{-\infty}^{\beta'x_i} \phi(t) dt = \Phi(\beta'x_i) \quad \text{Probit-Normal} \quad (4)$$

$$\mathcal{F}(\beta'x_i) = \frac{e^{\beta'x_i}}{1 + e^{\beta'x_i}} = \Lambda(\beta'x_i) \quad \text{Logit-Logistic} \quad (5)$$

$$\mathcal{F}(\beta'x_i) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(\beta'x_i) \quad \text{Arctan-Cauchy} \quad (6)$$

- ¿Cuál elegir?
 - Logistic similar a normal (se asemeja a t de Student con 7 dof)
 - Para $\beta'x_i$ entre -1.2 and 1.2, entregan mismas probabilidades

Estimación

$$\begin{aligned}\ell(\beta; Y | X) &= \sum_{i=1}^N \ln \left[\mathcal{F}(\beta' x_i)^{y_i} (1 - \mathcal{F}(\beta' x_i))^{1-y_i} \right] \\ &= \sum_{i=1}^N \left[y_i \ln \mathcal{F}(\beta' x_i) + (1 - y_i) \ln (1 - \mathcal{F}(\beta' x_i)) \right] \\ &= \sum_{y_i=1} \ln \mathcal{F}(\beta' x_i) + \sum_{y_i=0} \ln (1 - \mathcal{F}(\beta' x_i)) \\ \frac{\partial \ell}{\partial \beta} &= \sum_{i=1}^N \left[y_i \frac{f_i}{\mathcal{F}_i} - (1 - y_i) \frac{f_i}{1 - \mathcal{F}_i} \right] x_i \\ \frac{\partial \ell}{\partial \beta} &= \sum_{i=1}^N \left[\frac{(y_i - \mathcal{F}_i) f_i}{\mathcal{F}_i (1 - \mathcal{F}_i)} \right] x_i\end{aligned}\tag{7}$$

Métodos numéricos

Ejemplo 2 (ordered logit/probit)

- Gana local, empate, gana visita
- Ordered logit (probit)
- Interpretación ordinal (no cardinal)

$$y_i^* = \beta' x_i + u_i$$

$$y_i = \begin{cases} 0 & \text{if } y_i^* < a_0 \\ 1 & \text{if } a_0 \leq y_i^* < a_1 \\ \vdots & \vdots \\ J & \text{if } a_J \leq y_i^* \end{cases}$$

Parámetros: β y $a = (a_1, \dots, a_J)$

$$\Pr(y_i = 0) = \Pr(y_i^* < a_0) = \Pr(\beta' x_i + u_i < a_0) = \Pr(u_i < a_0 - \beta' x_i) = \mathcal{F}(a_0 - \beta' x_i)$$

$$\Pr(y_i = j) = \Pr(a_{j-1} \leq y_i^* < a_j) = \Pr(a_{j-1} \leq \beta' x_i + u_i < a_j) = \mathcal{F}(a_j - \beta' x_i) - \mathcal{F}(a_{j-1} - \beta' x_i)$$

$$\Pr(y_i = J) = \Pr(y_i^* \geq a_J) = \Pr(\beta' x_i + u_i \geq a_J) = \Pr(u_i \geq a_J - \beta' x_i) = 1 - \mathcal{F}(a_J - \beta' x_i)$$

$$\ell(\beta, a) = \sum_{i=1}^N \sum_{j=0}^J d_{i,j} \ln [\Pr(y_i = j)]$$
$$a_0 < a_1 < \dots < a_J$$

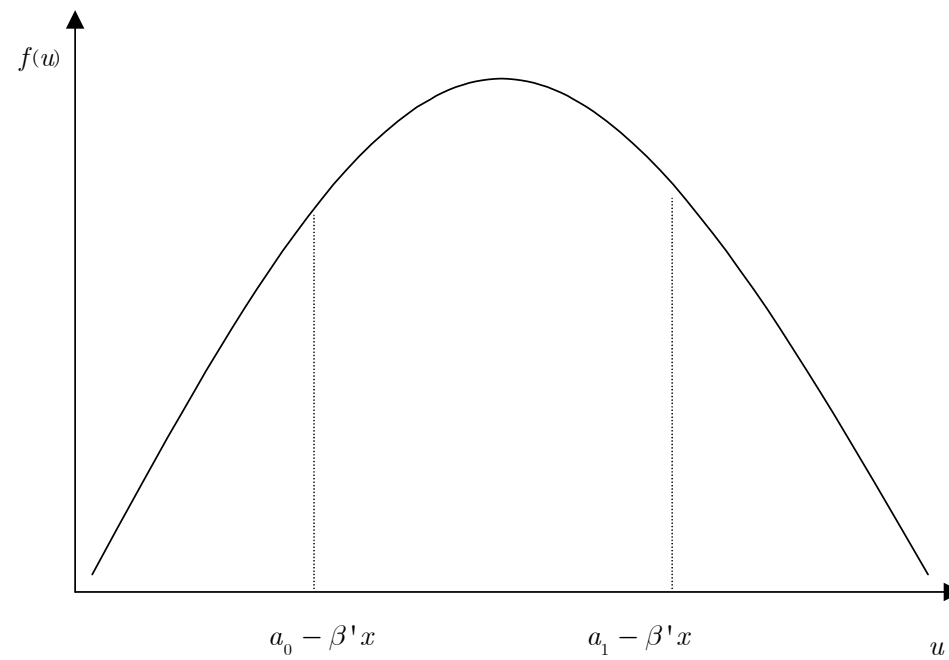


Figure 4: Probabilities with ordered observations

Ejemplo 3 (datos de conteo)

- Interpretación cardinal
- Goles anotados por cada equipo
- Discreta: Poisson regression

$$\Pr(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots$$

$$\ln \lambda_i = \beta' x_i, \quad \mathcal{E}[y_i | x_i] = \mathcal{V}[y_i | x_i] = \lambda_i = e^{\beta' x_i}$$
$$\frac{\partial \mathcal{E}[y_i | x_i]}{\partial x_i} = \lambda_i \beta$$

$$\ell(\beta) = \sum_{i=1}^N [-\lambda_i + y_i \ln \lambda_i - \ln y_i!]$$

$$\frac{\partial \ell}{\partial \beta} = \sum_{i=1}^N [y_i - \lambda_i] x_i$$

$$\frac{\partial^2 \ell}{\partial \beta \partial \beta'} = - \sum_{i=1}^N \lambda_i x_i x_i'$$

Globalmente cóncava. Crítica: misma media y varianza

Ejemplo

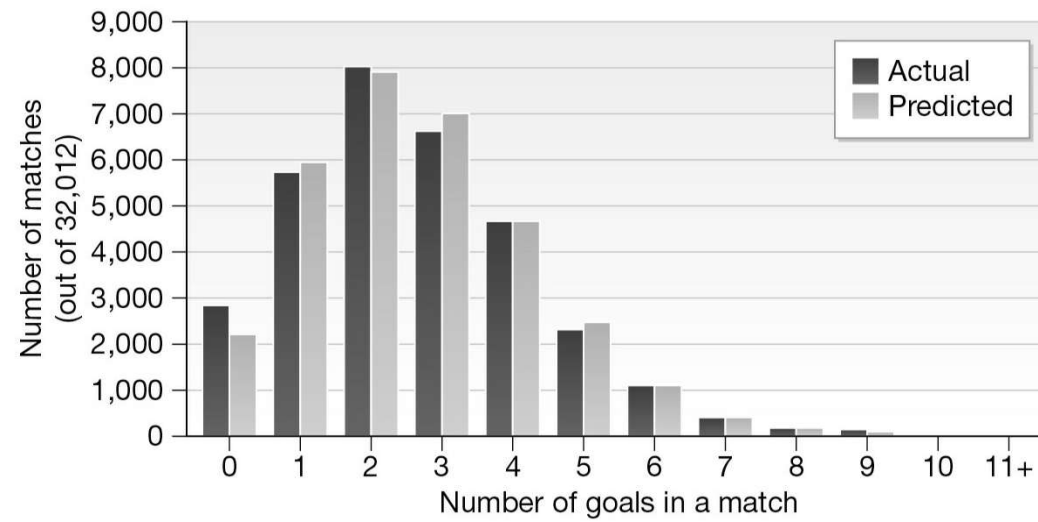


Figure 5:

Juzgando modelos

- Índices de coincidencia
 - Porcentaje de proyecciones correctas (probabilidad de ganar, empatar, perder)
- Pseudo – R^2
 - Promedio geométrico de probabilidades de resultados observados
- Comparación con “expertos” (bookmakers)