

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
- ❑ Los determinantes de resultados pueden ayudar a elaborar mejores estrategias, definir contrataciones, etc.

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

Resultado específico

- ❑ Goles de local y goles de visita
- ❑ Consideradas como 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

Preliminares

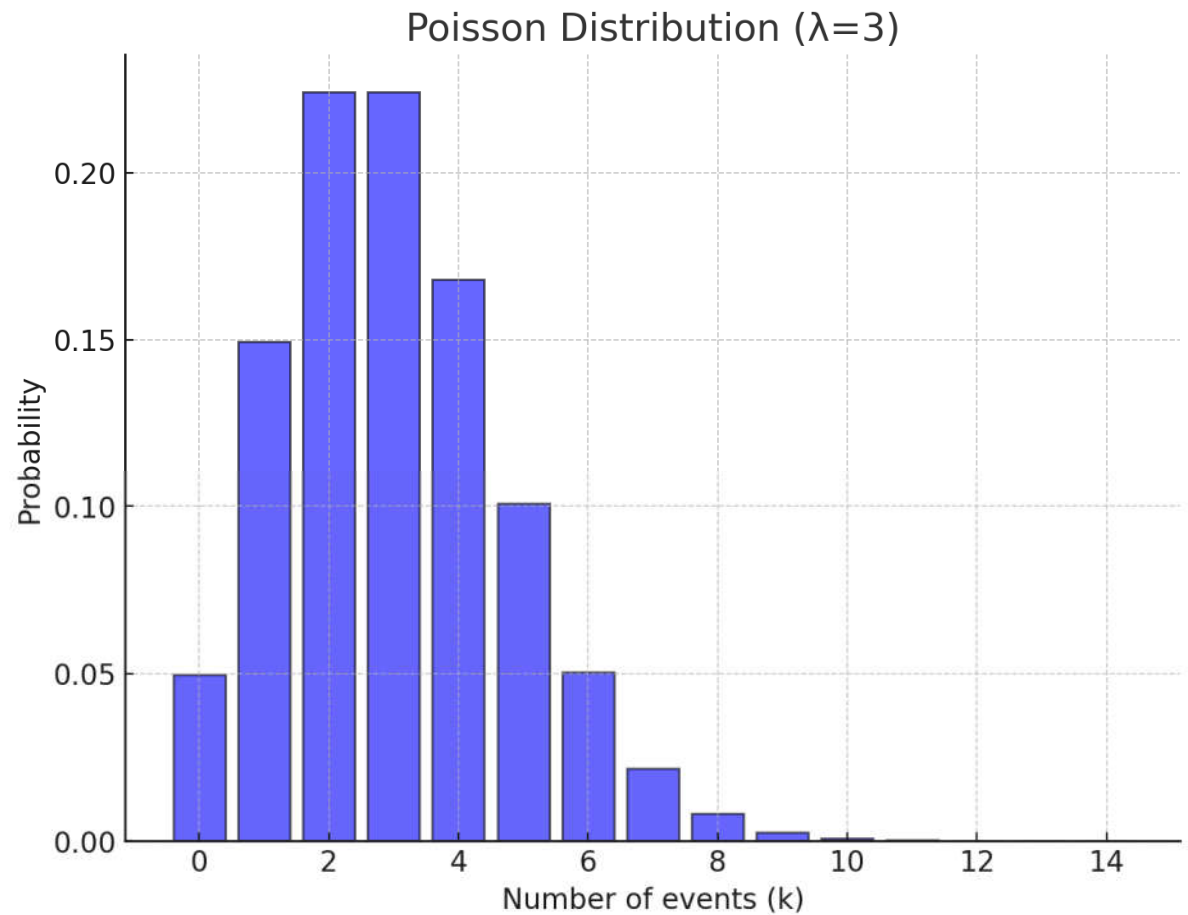
- Sea $f(y;a)$ la probabilidad de ocurrencia de y dado a
- En la práctica, tenemos muestra de y , pero desconocemos a
- La función de verosimilitud muestral describe la situación en que se trata y de $f(\cdot)$ como dado y a a como variable
- Buscamos el vector de parámetros que maximiza la probabilidad de ocurrencia de la muestra
- Propiedades deseadas: consistencia, normalidad asintótica, eficiencia asintótica, invariabilidad a reparametrizaciones
- Propiedades no deseadas: dependencia en supuesto distributivo (QMLE) y propiedades de muestras pequeñas (sesgo)

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

Resultado específico

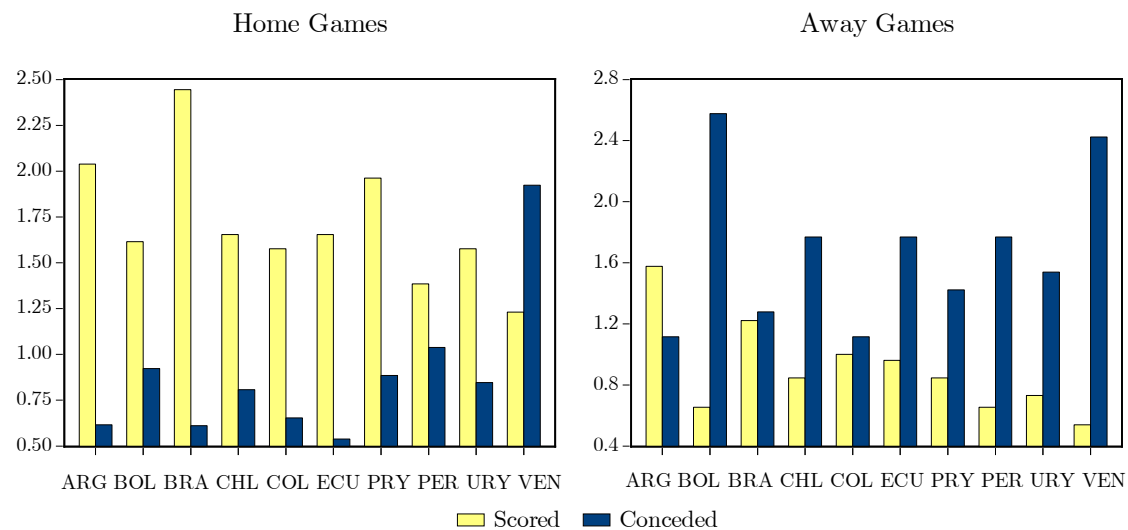
- Goles anotados
- Variable discreta
- Poisson (datos de conteo)



Bivariate Poisson model

$$\Pr(g_{ij,t}^h = f, g_{ij,t}^a = r) = e^{-(\lambda_{ij,t} + \theta_{ij,t} + \gamma_{ij,t})} \frac{\lambda_{ij,t}}{f!} \frac{\theta_{ij,t}}{r!} \sum_{k=0}^{\min(f,r)} \binom{f}{k} \binom{r}{k} k! \left(\frac{\gamma_{ij,t}}{\lambda_{ij,t} \theta_{ij,t}} \right)^k.$$

$$\ln(\lambda_{ij,t}) = \beta' x_{ij,t}, \quad \ln(\theta_{ij,t}) = \delta' x_{ij,t}, \quad \gamma_{ij,t} = 0.$$



Doble Poisson

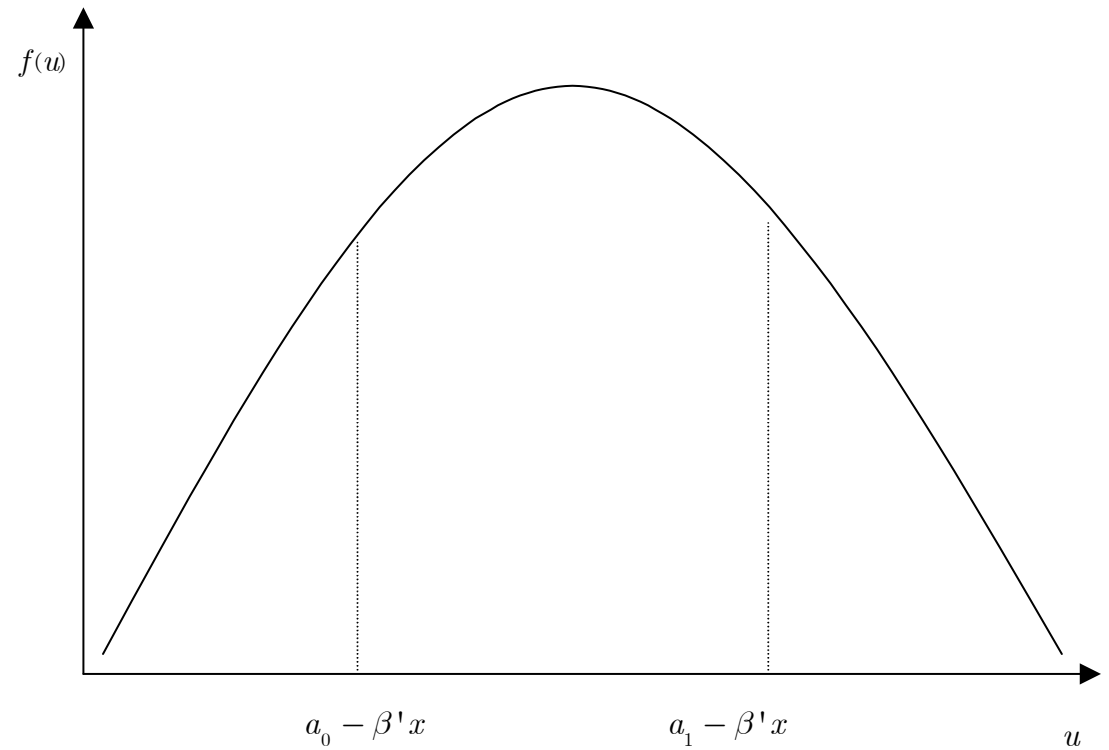
- ❑ Si se asume que la covarianza entre goles del equipo local y el de visita es 0
 - ❑ Dos procesos Poisson independientes
- ❑ Se proponen los determinantes del resultado
 - ❑ Características de equipos
 - ❑ Historial de cada equipo
 - ❑ Características del partido
 - ❑ Situación de los equipos en el campeonato
 - ❑ Localización del partido
 - ❑ Aspectos geográficos
 - ❑ Capacidad del estadio
 - ❑ Otros factores
 - ❑ Árbitros, etc

Extensiones

- ❑ Una vez estimados los modelos, pueden predecirse probabilidades de resultados específicos
- ❑ Predecirse resultado final (gana local, empate, pierde local)
- ❑ Simularse resultados
- ❑ Hacer predicciones fuera de muestra
- ❑ Un problema que se presenta en ocasiones:
 - ❑ Tendencia a subestimación de empates con pocos goles (0-0, 1-1, 2-2)
 - ❑ Zero-inflated (diagonally inflated) model, ajusta probabilidades para disminuir este problema

Resultado general

- ❑ Gana local, empate, gana visita
- ❑ Variable discreta
- ❑ Ordered logit (probit)



Ordered Probit model

$$V_{ij,t} = \begin{cases} 2 & \text{if } \mu_2 < v_{ij,t}^* + \varepsilon_{ij,t} \\ 1 & \text{if } \mu_1 < v_{ij,t}^* + \varepsilon_{ij,t} \leq \mu_2. \\ 0 & \text{if } v_{ij,t}^* + \varepsilon_{ij,t} \leq \mu_1 \end{cases}$$

$$v_{ij,t}^* = \rho' x_{ij,t},$$

Ordered Probit

- ❑ Tiene la ventaja de no requerir evaluar la covarianza entre goles de local y visita
- ❑ Se centra en resultado (sin goles)
- ❑ Goles pueden ser relevantes para desempate
- ❑ Determinantes similares a los ya discutidos

Ordered Probit

- ❑ Tiene la ventaja de no requerir evaluar la covarianza entre goles de local y visita
- ❑ Se centra en resultado (sin goles)
- ❑ Goles pueden ser relevantes para desempate
- ❑ Determinantes similares a los ya discutidos

Juzgando los 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)

Modelos de Pronóstico de Resultados

RÓMULO A. CHUMACERO

