

Modelování a odhadování výsledků ledního hokeje

Pavλίna Hellusová¹

1 Úvod

Hlavním cílem této práce je představit upravený model využívající dvojité Poissonovo rozdělení a porovnat jeho predikční schopnosti s původním modelem, z článku autorů Marka, Šedivé a Āoupala (2014) s využitím dat z české Extraligy.

2 Použité modely

2.1 Dvojitý Poissonův model

Ačkoliv v článku autorů Marka, Šedivé a Āoupala (2014) byl jako nejlepší vyhodnocen dvojitý Poissonův model s úpravou prvků na diagonále, pro práci byl vybrán model bez úpravy, kvůli jeho relativní jednoduchosti a téměř stejným predikčním schopnostem.

Předpokladem je, že se počty vstřelených gólů domácích a hostů řídí Poissonovým rozdělením a jsou navzájem nezávislé, pak $X_{ij} \sim \text{Po}(\lambda_H = \mu\alpha_i\beta_j\gamma)$ a $Y_{ij} \sim \text{Po}(\lambda_A = \mu\alpha_j\beta_i)$, kde α_i je parametr popisující útok, β_i je parametr popisující obranu, μ je parametr měřítka a γ zachycuje takzvaný efekt domácího týmu. Parametry α_i a β_i splňují $\sum_i \alpha_i = N$ a $\sum_i \beta_i = N$, kde N je počet týmů.

2.2 Upravený dvojitý Poissonův model

Na základě informací z článku autorů Marka a Vávry (2017) byl navržen model, který globální parametr γ upravuje na individuální a zároveň je vhodný pro hokejová data. Poskytuje možnost odlišit hru domácích, ale zároveň zachovává jednoznačnou identifikaci týmů pomocí parametrů α_i a β_i , které jsou při hře na domácí půdě upravovány.

$X_{ij} \sim \text{Po}(\lambda_H = \mu\alpha_i\gamma_i\gamma\beta_j)$ a $Y_{ij} \sim \text{Po}(\lambda_A = \mu\alpha_j\beta_i\delta_i\delta)$, kde γ_i slouží k individualizaci globálního vlivu domácího prostředí a tím pádem dochází k úpravě síly útoku domácího týmu (parametru α_i). Parametr δ_i upravuje slabost obrany domácího týmu (β_i). Globální parametry γ a δ ve spojení s podmínkou $\sum \alpha_i = \sum \gamma_i = \sum \beta_i = \sum \delta_i = N$ (N je počet týmů), zajišťují relativnost parametrů. Podmínky na součet zároveň zaručují jednoznačnou identifikaci parametrů.

3 Srovnání modelů

Modely byly porovnávány dvěma způsoby. Prvním z nich bylo vypočítání několika kritérií, díky nimž bylo možné určit kvalitu modelů a druhým bylo ověření účinnosti modelů na fiktivním sázení proti sázkové kanceláři.

¹ studentka bakalářského studijního programu Matematika, obor Matematika a finanční studia, e-mail: cervenk4@students.zcu.cz

3.1 Srovnání dle kritérií

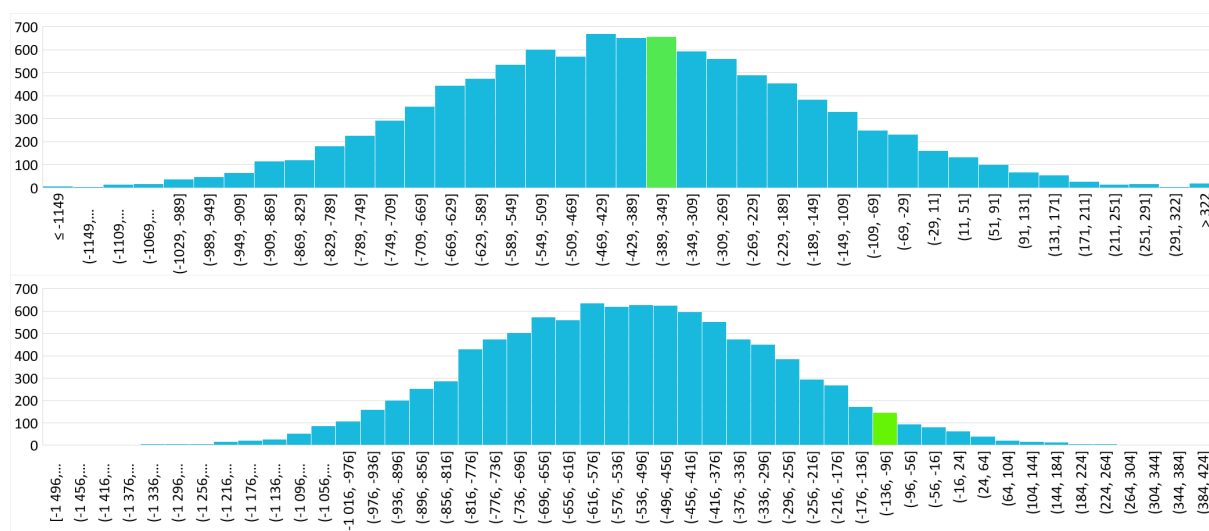
Odhad probíhal od 16. hracího dne (kolo) s využitím parametrů získaných na základě zápasů do předcházejícího kola. Tento postup byl opakován až k poslednímu kolu. V každém z kol byla následně vypočítána pravděpodobnost výhry domácích, výhry hostů a remízy. Hodnotící kritéria jsou *LogLoss* (menší hodnota - lepší), kalibrace *Cal* (blíže 1 - lepší) a přesnost *Ac* (nad 1/3 lepší než náhoda, 1 maximum).

Kritérium	Původní model	Upravený model
<i>LogLoss</i>	1,090	1,121
<i>Cal</i>	1,214	1,212
<i>Ac</i>	0,481	0,484

Tabulka 1: Výsledné hodnoty jednotlivých kritérií pro ověřované modely

3.2 Srovnání dle sázení

Předpokládejme „neomezený“ bank a výši sázky $h = 10$ Kč. Nejprve je definováno kritérium pro vložení sázky $\varphi = p_m^R \cdot o_m^R$, kde $p_m^R, R \in \{H, D, A\}$ je pravděpodobnost výhry domácích, remízy a výhry hostů a o_m^R je kurz sázkové kanceláře. Pokud je hodnota tohoto kritéria vyšší, než stanovená mez L , pak dojde k vložení sázky. Vždy musí platit $L \geq 1$, aby byla sázka výhodná. V případě, že $L = 1$ (tedy minimální), je u původního modelu vsazeno na 85,43% zápasů, zatímco u upraveného modelu na 86,09%. Výsledky byly porovnány s naivní metodou náhodných sázek (v obrázku 1 modře).



Obrázek 1: Histogramy četností výher/proher v porovnání s původním a upraveným modelem

Literatura

Marek, P., Šedivá, B., a ěoupal, T., (2014). Modelling and prediction of ice hockey match results. *Journal of Quantitative Analysis in Sports*, 357–365, ISSN: 1559-0410.

Marek, P. a Vávra, F. (2017) *Home team advantage in english premier league.*. Padova University Press, pp. 244–254.