



HOKEJOVÉ GÓLY

Michal Friesl

*Katedra matematiky & NTIS
Fakulta aplikovaných věd
Západočeská univerzita v Plzni*

Úvod

Podíváme se na průběh počtu gólu během **hokejových** zápasů

- aneb co lze vyčíst ze statistik NHL
- pro sportovní znalce nic překvapivého, ale pro nás ostatní :-)

Kolegové, kteří rozumějí sportu

- Liam J. A. Lenten (La Trobe University)
- Jan Libich (La Trobe University & VŠB-TU Ostrava)
- Petr Stehlík (FAV ZČU)

Více úvah o sportovních a dalších okolnostech

- článek v *Journal of the Operational Research Society* 68 (2017), no. 9, 1006–1018

Datový soubor

Záznamy za období 20,5 sezón

- zápasy NHL z let 1995–2015
- přesněji 2. část sezóny 1994(-95) až celá 2014(-15)

Dostupné údaje o zápasech

- sezóna (rok)
- zda utkání základní části sezóny, nebo playoff
- časy gólů (min:sec)
- informace o prodloužení popř. nájezdech

Datový soubor

Záznamy za období 20,5 sezón

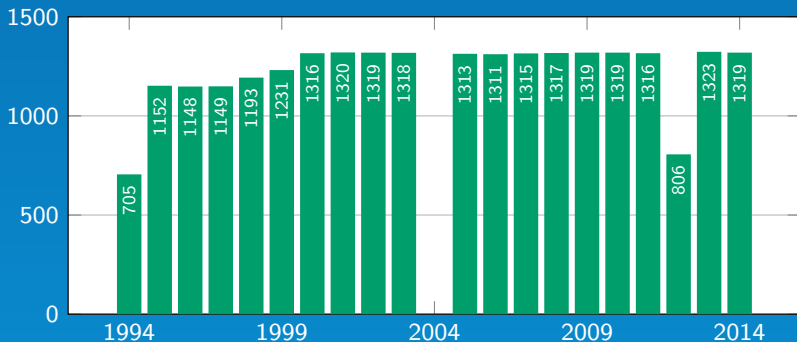
- zápasy NHL z let 1995–2015
- přesněji 2. část sezóny 1994(-95) až celá 2014(-15)

Souhrnně

- 24 509 utkání
- 132 159 gólů (132 670 včetně prodloužení)
- 1 469 880 minut základní hrací doby
- 88 192 800 sekund

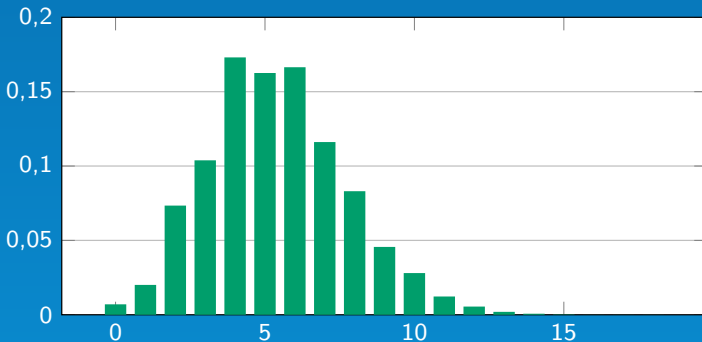
Základní přehled

Počty zápasů v sezónách



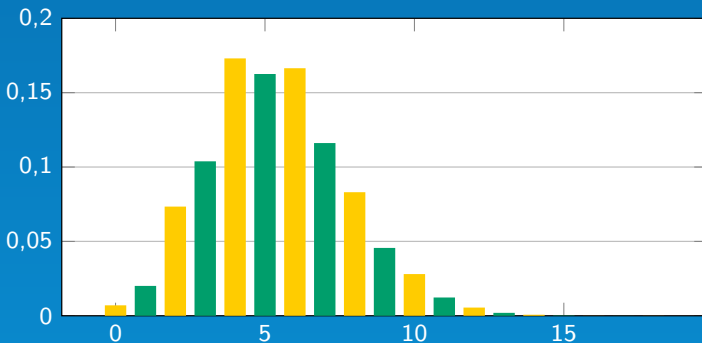
Základní přehled

Rozdělení počtu gólů po 60 minutách základní hrací doby



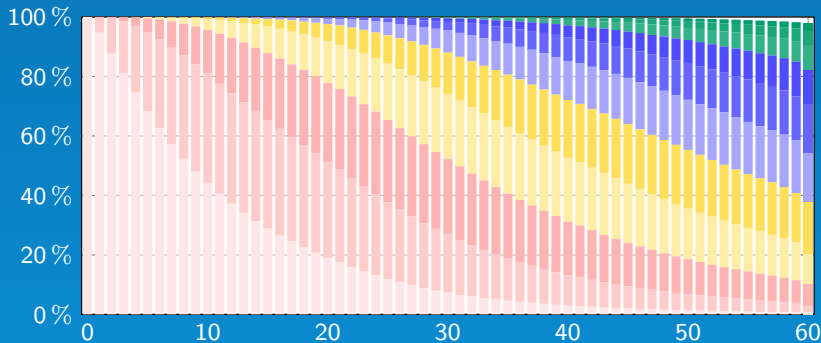
Základní přehled

Rozdělení počtu gólů po 60 minutách základní hrací doby



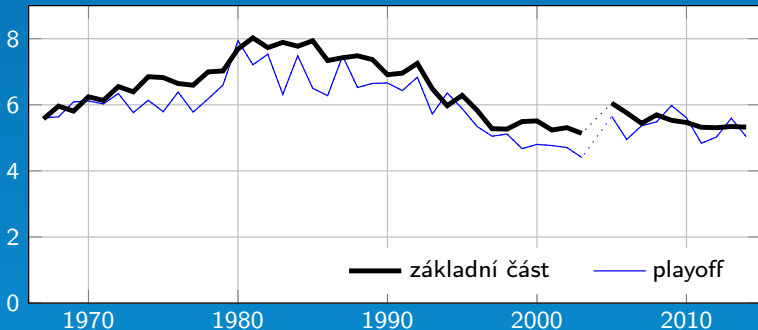
Základní přehled

Jak výsledné rozdělení postupně vzniká



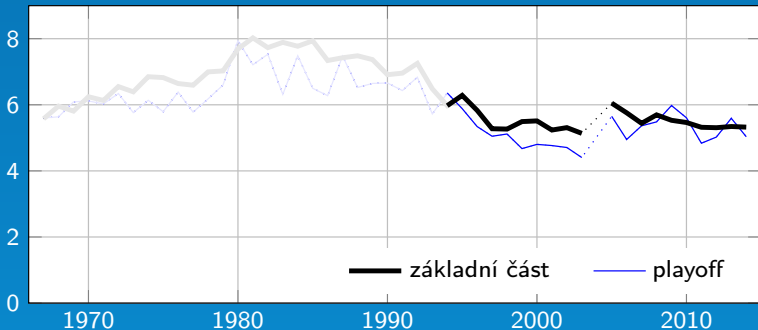
Základní přehled

Historický vývoj průměrného počtu gólů, včetně prodloužení



Základní přehled

Historický vývoj průměrného počtu gólů, včetně prodloužení



Přístup

Zajímá nás jen gólový průběh zápasu globálně

- zabývat se budeme **jen základní hrací dobou**
- jednoduché metody, nehledáme složité modely

Čím se zabývat nebudeme

- nebudeme rozlišovat po týmech, podle síly soupeřů apod.
- primárně nebudeme sledovat vývoj během let (i když...)
- nebudeme rozlišovat běžnou sezónou a playoff

Zvolené statistické modely

- použité naivně, často jen jako formální prostředek
- bez přesného splnění předpokladů atd.

Obsah

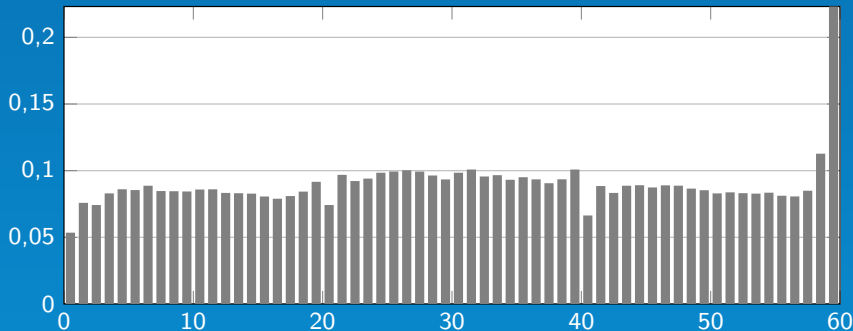
- Úvod ✓
- Minutový průběh
- Vteřinový průběh (jen zmíníme)
- Minutový průběh podrobněji
- Co to říká, k čemu to je dobré

Obsah

- Úvod ✓
- **Minutový průběh** ←
- Vteřinový průběh (jen zmíníme)
- Minutový průběh podrobněji
- Co to říká, k čemu to je dobré

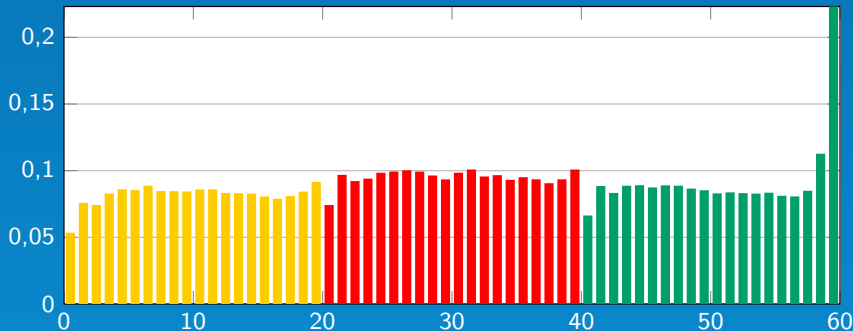
Minutový průběh

Průměrné počty gólů vstřelených v jednotlivých minutách utkání



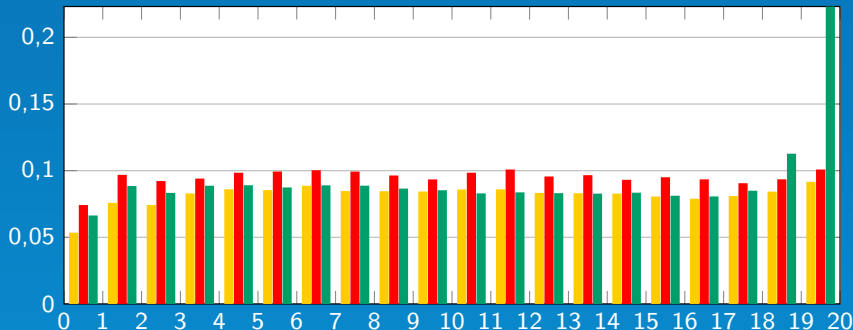
Minutový průběh

Průměrné počty gólů vstřelených v jednotlivých minutách utkání



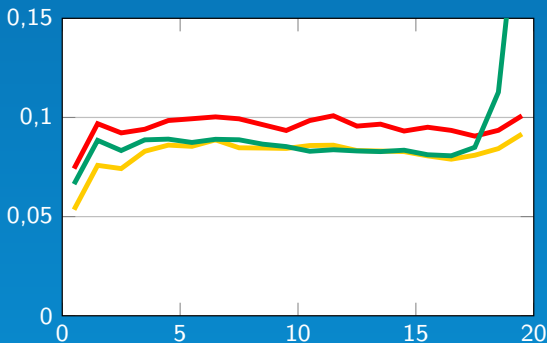
Minutový průběh

Průměrné počty gólů vstřelených v jednotlivých minutách třetin



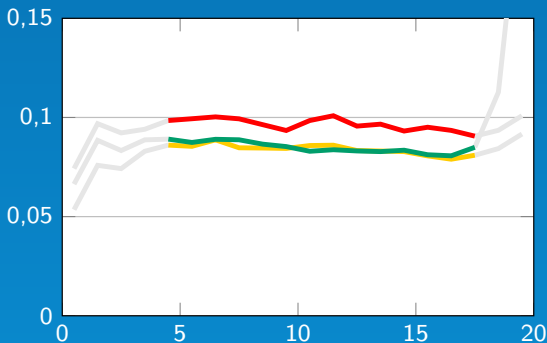
Minutový průběh

Průměrné počty gólů vstřelených v jednotlivých minutách třetin



Minutový průběh

Průměrné počty gólů vstřelených v jednotlivých minutách třetin



Model pro třetinu

Podle inspirace z obrázků

- odmyslíme si první čtyři a poslední dvě minuty třetin
- v každé třetině **lineární trend** se **stejnou směrnicí**, ale jiným absolutním členem

Regresní model pro počet gólů v t-té minutě i-té třetiny

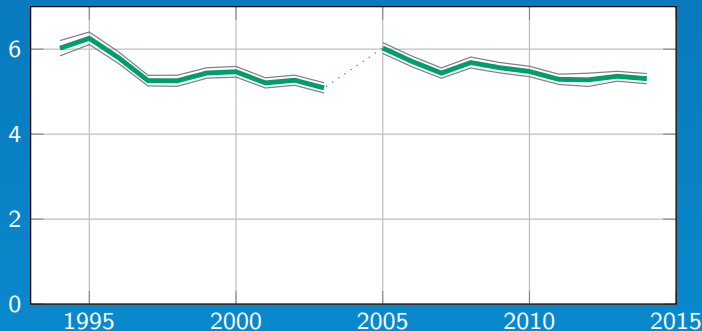
- $EY(i, t) = \mu_i + \beta(t - 10,5)$
- $EY(i, t) = \exp(\mu_i + \beta(t - 10,5))$

Odhady (z dat bez krajních minut) vycházejí:

- $\beta = -0,01085$, $20\mu_1 = 1,69$, $20\mu_2 = 1,94$, $20\mu_3 = 1,71$
- rozdíly větší než 10% gólů v třetině

Aplikace na každou sezónu zvlášť

Průměrné počty gólů se mezi sezónami liší, možná by se neměly použít stejné hodnoty parametrů pro všechny sezóny.



Aplikace na každou sezónu zvlášť

Průměrné počty gólů se mezi sezónami liší, možná by se neměly použít stejné hodnoty parametrů pro všechny sezóny.

Odhadneme-li v každé sezóně **samostatný** model

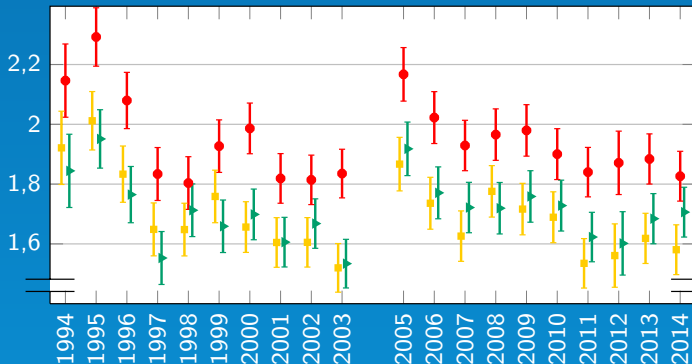
- $\mu_1 = \mu_2 = \mu_3$ se zamítá vždy ($p < 0,001$) kromě 1998
- rozdíl $\mu_1 - \mu_3$ je nevýznamný vždy ($p > 0,05$) kromě 2014

Směrnice

- $\beta = 0$ se zamítá u 9 z 20 sezón (a jeho odhad je tam záporný)
- případná interakce mezi trendem a třetinou je nevýznamná ($p > 0,05$) v 18 z 20 sezón (je jen $p_{2005} = 0,007$ a $p_{2013} = 0,030$)

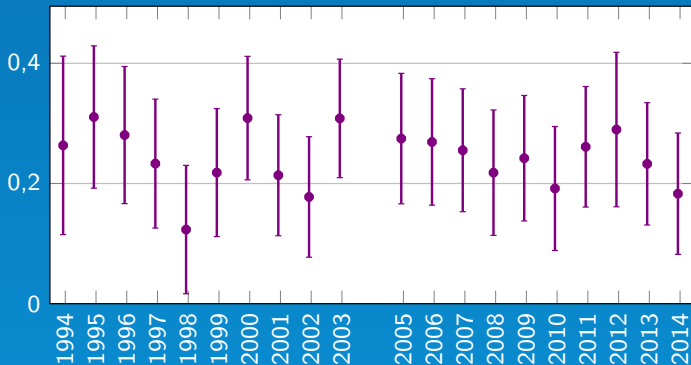
Aplikace na každou sezónu zvlášť

Odhady $20\mu_i$ (počtu gólů ve třetině) v jednotlivých sezónách — mezi sezónami jsou rozdíly.



Aplikace na každou sezónu zvlášť

Odhad rozdílu mezi druhou a krajními třetinami $20(\mu_2 - \frac{\mu_1 + \mu_3}{2})$ — je mezi sezónami celkem stabilní, číselně mezi 0,18 a 0,31 gólu.



Sezóny dohromady

Sezóny navrženy do jednoho modelu

- lineární trend v třetině
- faktory třetiny, sezóny
- případné interakce mezi těmito třemi proměnnými

Sezóny dohromady

Sezóny navrženy do jednoho modelu

- lineární trend v třetině
- faktory třetiny, sezóny
- případné interakce mezi těmito třemi

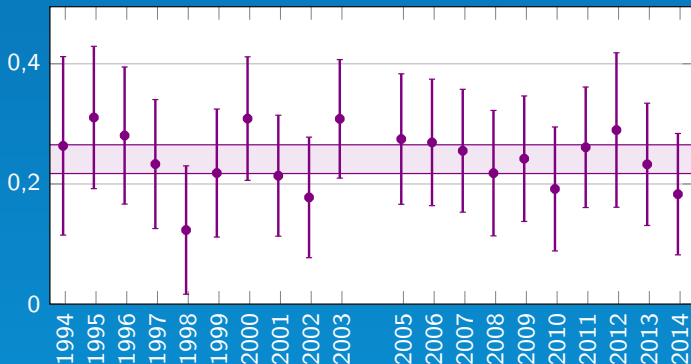
Lze vyřadit

- interakce mezi směrnici a třetinou ($p > 0,95$)
- interakce mezi sezónou a třetinou ($p > 0,5$)
- dejme tomu i interakci mezi směrnici a sezónou (sice $p = 0,027$, ale šla by, kdyby nebylo např. sezón 1994, 2001 a 2013)

Rozdíly třetin tedy **stejně ve všech sezónách**

Sezóny dohromady

V modelu s faktorem sezóny $EY(p, t) = \mu_{sez} + \mu_i + \beta t$ pak vychází 95 % odhad rozdílu druhé a ostatních třetin (0,22, 0,27) gólu.



Obsah

- Úvod ✓
- Minutový průběh ✓
- Vteřinový průběh (jen zmíníme)
- Minutový průběh podrobněji
- Co to říká, k čemu to je dobré

Obsah

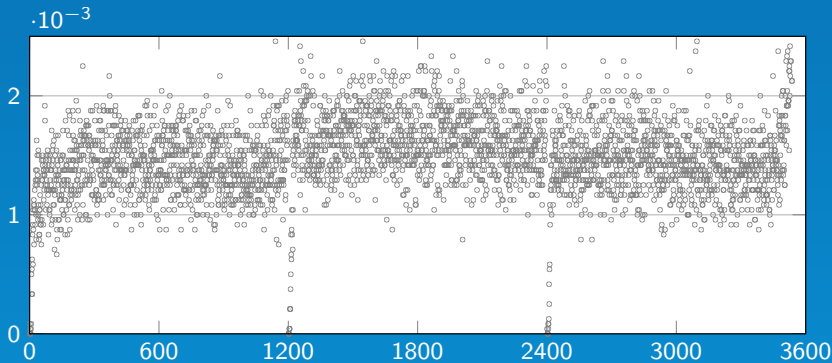
- Úvod ✓
- Minutový průběh ✓
- Vteřinový průběh (jen zmíníme) ←
- Minutový průběh podrobněji
- Co to říká, k čemu to je dobré

Vteřinový průběh

Pro každou z 3600 vteřin zápasu známe počet gólů (obvykle 0 nebo 1), můžeme tedy v každé odhadnout intenzitu skórování ...

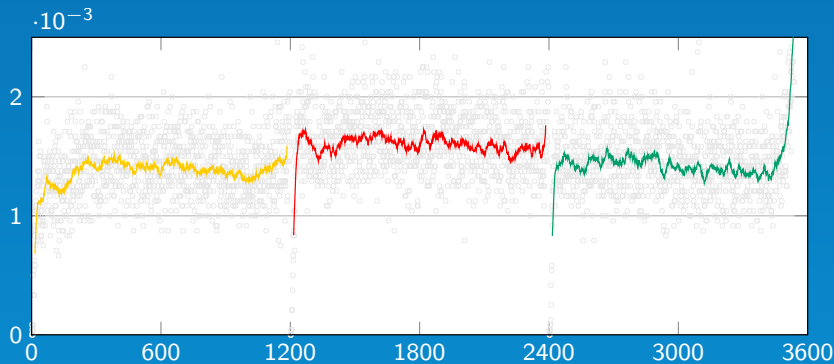
Vteřinový průběh

Průměrné počty gólů vstřelené za 1 s



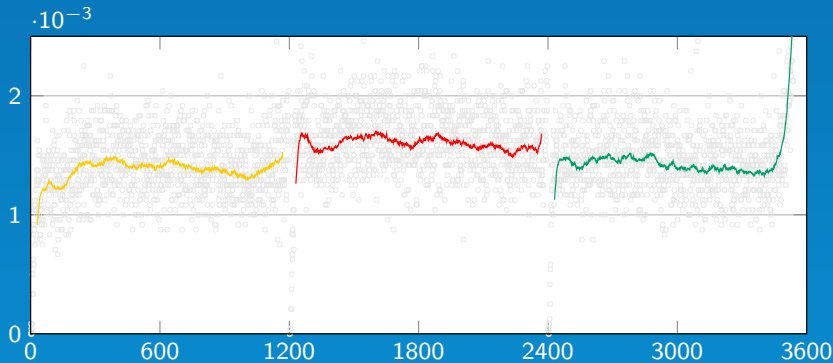
Vteřinový průběh

Průměrné počty gólů vstřelené za 1 s, klouzavé průměry délky 30 s



Vteřinový průběh

Průměrné počty gólů vstřelené za 1 s, klouzavé průměry délky 60 s



Skokově nebo spojitě

Modelování sekundové intenzity opět jednoduchou regresí

- uvnitř třetiny lineární (klesající?) průběh
- skoky mezi třetinami

A k tomu na krajích třetin domodelujeme pomocí funkce

- buď po částech konstantní, tedy doplníme **skoky**
- anebo po částech lineární spojitě, tj. bude se měnit **směrnice**

Volba okamžiků změn intenzity

Intuitivně

- stejné na začátku každé třetiny, např. v časech 10, 20, 30, 60, 120 s
- speciálně v poslední třetině navíc v časech 30, 60, 120 s před jejím koncem

Výběrovým algoritmem, např.

- na začátku uvažujeme model se změnami ve všech vteřinách resp. násobcích 10 nebo 30 s a příp. dodatečné změny ve 2. a dále ve 3. třetině
- postupné odebírání parametrů

Volba okamžiků změn intenzity

Intuitivně

- stejné na začátku každé třetiny, např. v časech 10, 20, 30, 60, 120 s
- speciálně v poslední třetině navíc v časech 30, 60, 120 s před jejím koncem

Výběrovým algoritmem, např.

- pro směrnice můžeme dospět k časům změn

| Změny | V každé třetině | Ve 3. navíc |
|----------|-------------------|--------------------|
| Po 10 s | 20, 250, -10 | 20, -170, -70, -60 |
| Po 1 min | 60, 120, 240, -60 | 60, -120 |

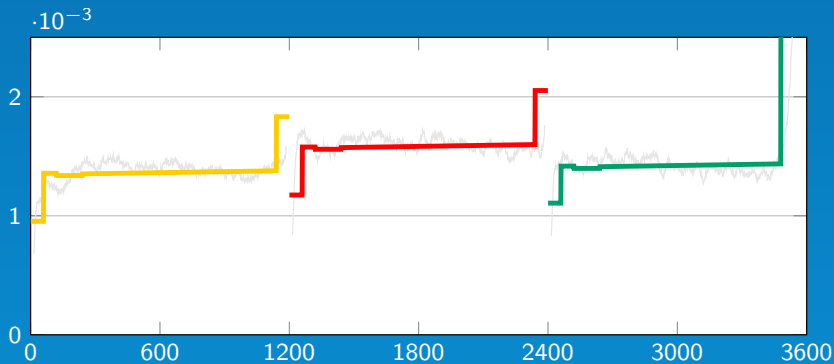
Porovnání modelů

Celkový součet čtverců $6,89 \cdot 10^{-4}$

| Model | Volba | Změn | RSS ($\times 10^{-4}$) |
|----------|------------|------|--------------------------|
| Skoky | intuitivní | 8 | 2,67 |
| Skoky | po 1 min | 6 | 3,86 |
| Skoky | po 10 s | 7 | 2,68 |
| Skoky | po 10 s | 6 | 2,67 |
| Směrnice | intuitivní | 8 | 2,56 |
| Směrnice | po 1 min | 6 | 2,64 |
| Směrnice | po 10 s | 7 | 2,38 |
| Směrnice | po 10 s | 6 | 2,40 |

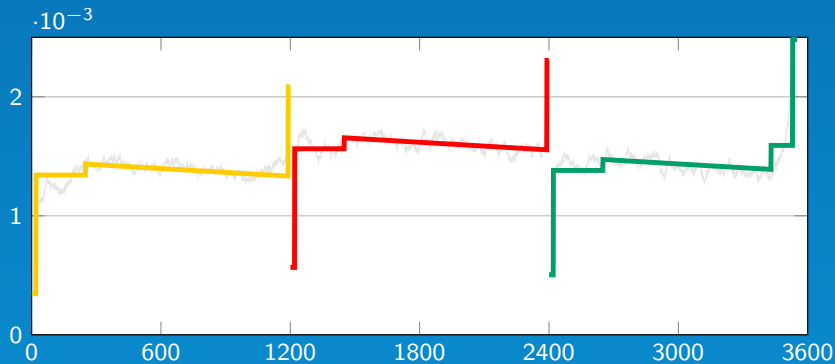
Porovnání modelů

Se skoky, změny po minutách



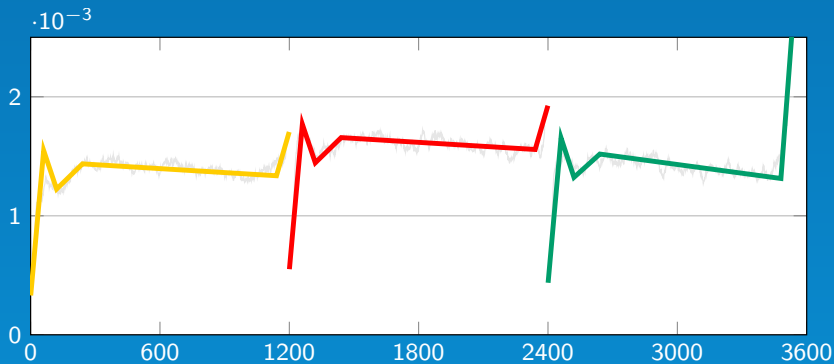
Porovnání modelů

Se skoky, změny po 10 s (7 změn)



Porovnání modelů

Po částech lineární, změny po minutách (6 změn)



Obsah

- Úvod ✓
- Minutový průběh ✓
- Vteřinový průběh (jen zmíníme) ✓
- Minutový průběh podrobněji
- Co to říká, k čemu to je dobré

Obsah

- Úvod ✓
- Minutový průběh ✓
- Vteřinový průběh (jen zmíníme) ✓
- **Minutový průběh podrobněji** ←
- Co to říká, k čemu to je dobré

Minutový průběh podrobněji

Možné další faktory

- běžná sezóna nebo playoff (v playoff méně gólů)
- nepoměr síly soupeřů
- stav utkání — průběžné skóre
- počet hráčů na ledě (vyloučení, nemáme údaje)

Podíváme se jen na souvislost s **aktuálním skóre**

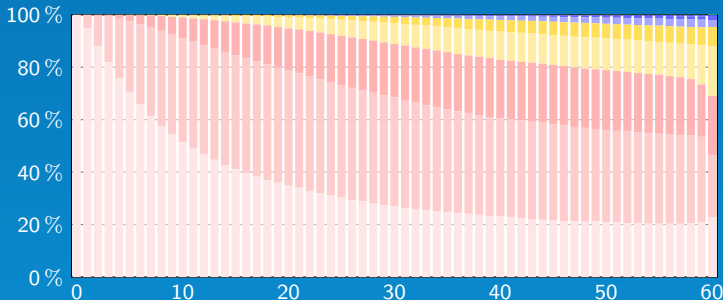
- velikost rozdílu
- rozdíl včetně znaménka

Jen minutově.

Vývoj skóre

Vývoj rozdělení velikosti rozdílu skóre během utkání

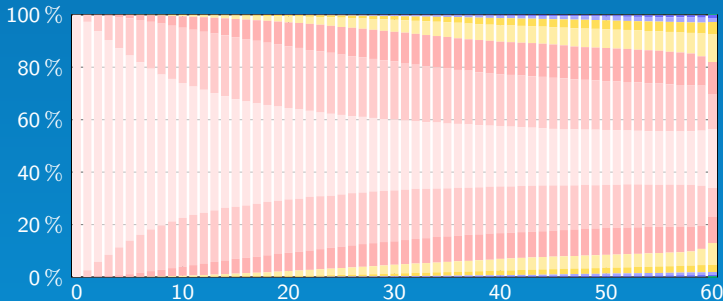
- remízou končí 22,7 % utkání (i díky prudší změně na konci)



Vývoj skóre

Vývoj rozdělení rozdílu **domáci** – **hosté** během utkání

- domáci vedou více než hosté, na konci 43,5 % proti 33,8 %



Intenzita podle rozdílu skóre

Ovlivňuje aktuální skóre intenzitu skórování?

Bude vidět zejména různý průběh intenzity na konci poslední třetiny

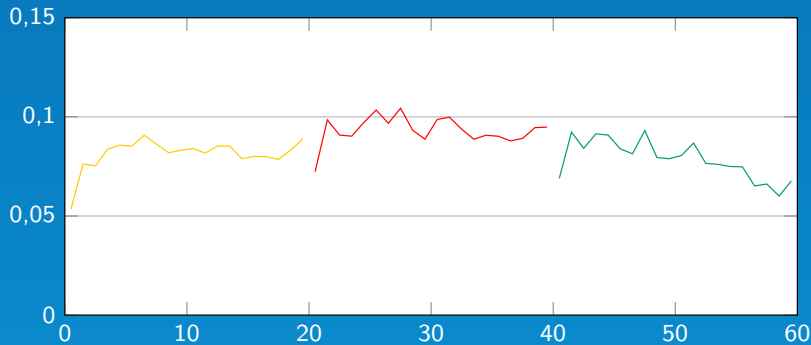
- velký **nárůst** při rozdílu skóre 1 nebo 2 (odvolání brankáře)
- při rozdílu vyšším než 2 už ne, a při rozdílu 0 **pokles**

V ostatních minutách

- **vyšší intenzita ve druhé** oproti ostatním třetinám **zůstává**, bez ohledu na rozdíl skóre
- s rostoucím rozdílem jen možná lehounce vyšší intenzita

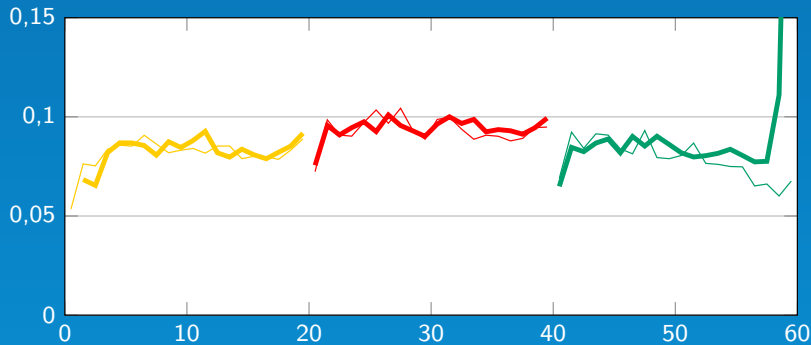
Intenzita podle rozdílu skóre

Při rozdílu 0:



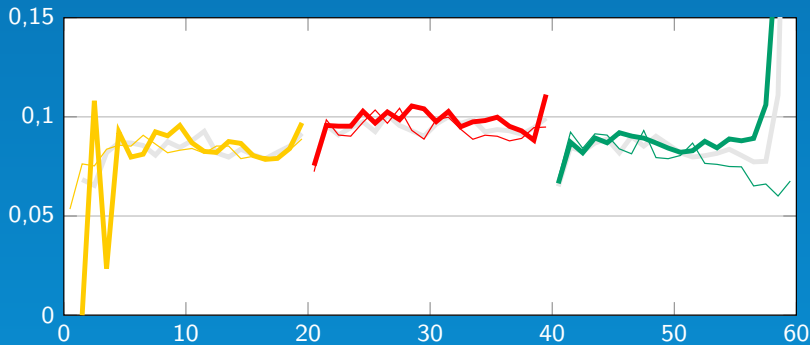
Intenzita podle rozdílu skóre

Při rozdílu 1 (vůči 0):



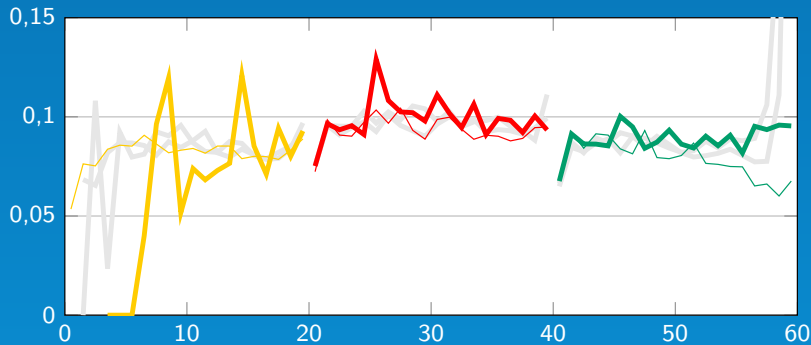
Intenzita podle rozdílu skóre

Při rozdílu 2 (vůči 0):



Intenzita podle rozdílu skóre

Při rozdílu **3** (vůči 0):



Intenzita podle rozdílu skóre

Doplníme-li do vteřinového modelu (počítaného bez okrajových minut) rozdíl skóre na začátku každé minuty jako další proměnnou

- všechny interakce se sezónou lze vynechat
- odhad vlivu druhé třetiny oproti první a třetí zůstává 0,23 gólu
- odhad vlivu rozdílu skóre je oproti nulovému rozdílu +0,025, +0,108, +...

Zjednodušit model jen na stav na začátku třetiny není užitečné

- v takovém případě se stav na začátku druhé třetiny jeví nevýznamný

Skórování prohrávajících vs. vyhrávajících

Budeme-li sledovat nejen rozdíl, ale i jeho znaménko

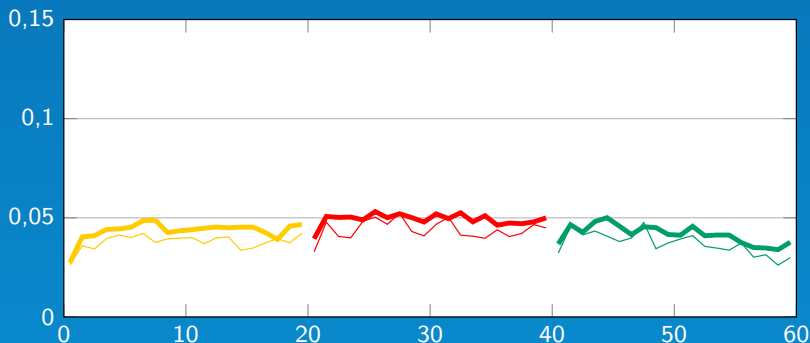
- kladné = vyhrávají domácí
- záporné = vyhrávají hosté

Kdo ze zvýšeného skórování při rozdílu skóre těží

- víme, že obecně více skórují domácí
- při nerovnosti vyhrávající tým skóruje s menší intenzitou oproti nerozhodnému stavu, prohrávající naopak s větší
- prohrávají-li hosté, zvýší se skórování hostů (a poklesne skórování u domácích) tak, že dochází k vyrovnání sil
- z nárůstu na konci poslední třetiny těží více vyhrávající

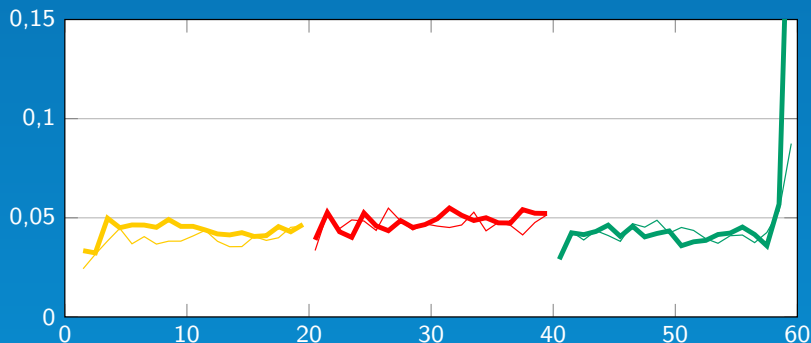
Skórování prohrávajících vs. vyhrávajících

Domáci (tlustě) a hosté při rozdílu 0



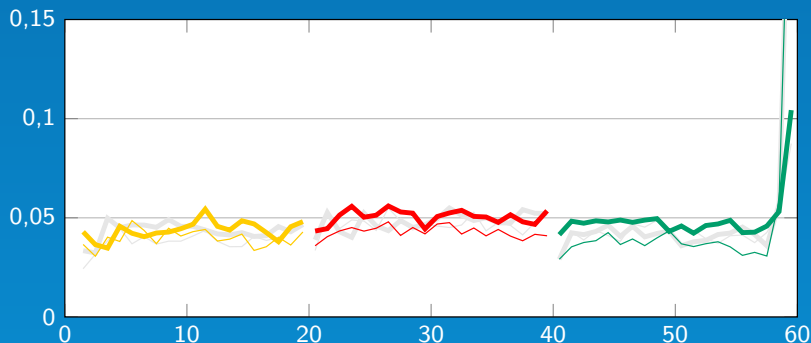
Skórování prohrávajících vs. vyhrávajících

Domáci (tlustě) a hosté při rozdílu +1



Skórování prohrávajících vs. vyhrávajících

Domáci (tlustě) a hosté při rozdílu -1



Obsah

- Úvod ✓
- Minutový průběh ✓
- Vteřinový průběh (jen zmíníme) ✓
- Minutový průběh podrobněji ✓
- Co to říká, k čemu to je dobré

Obsah

- Úvod ✓
- Minutový průběh ✓
- Vteřinový průběh (jen zmíníme) ✓
- Minutový průběh podrobněji ✓
- Co to říká, k čemu to je dobré ←

Vysvětlení rozdílu mezi třetinami

Ve druhé třetině padá více gólů

- intenzita je větší po celou třetinu (s výjimkou konce zápasu)
- nelze vysvětlit pouze dalšími faktory hry (skóre, vylučování, atd.)

Např. z dostupných statistik o počtech vyloučení v třetinách, délce přesilovek a skórování při vyloučeních

- ve druhé třetině je také vyšší počet vyloučení (o 0,5)
- při menším počtu hráčů na ledě se skórování výrazně zvyšuje, např. o 2,61 gólu/hod při počtu 5:4

Vyloučením by šlo připsat jen asi 0,036 gólu ze zjištěného nárůstu 0,25.

Vysvětlení rozdílu mezi třetinami

Pravděpodobné vysvětlení spočívá ve **střídání stran hřiště** mezi třetinami



Krátké střídání (short change)

- střídačka je u obranné třetiny (1. a 3. třetina)
- i pod tlakem mají obránci více příležitostí ke střídání během hry

Dlouhé střídání (long change)

- střídačka je daleko od obranné třetiny
- obránci mají méně příležitostí ke střídání během hry
- také jsou vyčerpaní a tak útočící má větší šanci dát gól

Možná změna

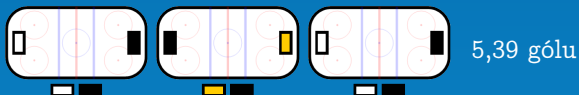
Stávající pravidla



5,39 gólu

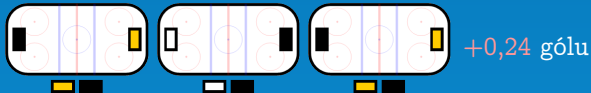
Možná změna

Stávající pravidla

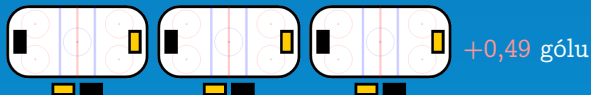


Zájemcům o vyšší počtu gólů v zápase se nabízí jednoduchá změna

- Začít dlouhým



- Vše s dlouhým



Možná změna

Provedení změny

- není organizačně náročné (začít hrát na jiné straně hřiště)
- podobná v roce 2014 u prodloužení s motivací snížit pravděpodobnost, že v prodloužení se nerozhodne

Vedlejší efekt

- kromě zvýšení počtu gólů by vedla k drobnému snížení počtu zápasů, které v řádné době skončí remízou

Problémy

- větší zátěž pro hráče
- fanoušci bývají za střídačkou svého klubu, po změně by jejich branka byla většinou dál od nich; daleko by to měl i odvolávaný brankář

Simulace změn

Markovský řetězec vývoje stavu utkání (rozdílu skóre)

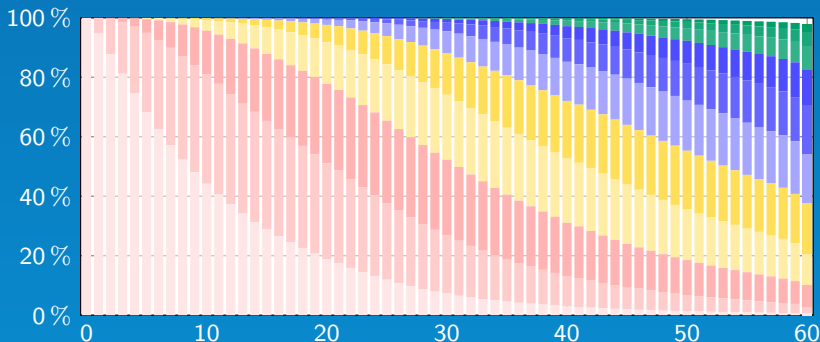
- nehomogenní se spojitým časem, intenzity se skoky po minutách
- pozorované intenzity upravené dle simulované varianty

Typický výsledek 100 000 simulací

| | gólů | domáci | hosté | remíza |
|-----------------|------|--------|--------|--------|
| Skutečnost | 5,39 | 43,5 % | 33,8 % | 22,7 % |
| Stávající | 5,40 | 43,7 % | 34,2 % | 22,1 % |
| Začít dlouhým | 5,63 | 44,2 % | 34,3 % | 21,5 % |
| Ve všech dlouhé | 5,88 | 44,1 % | 35,1 % | 20,8 % |

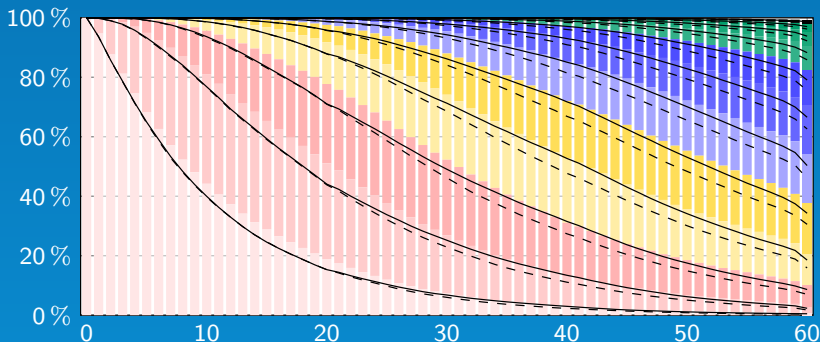
Simulace změn

Vývoj rozdělení počtu gólů během utkání



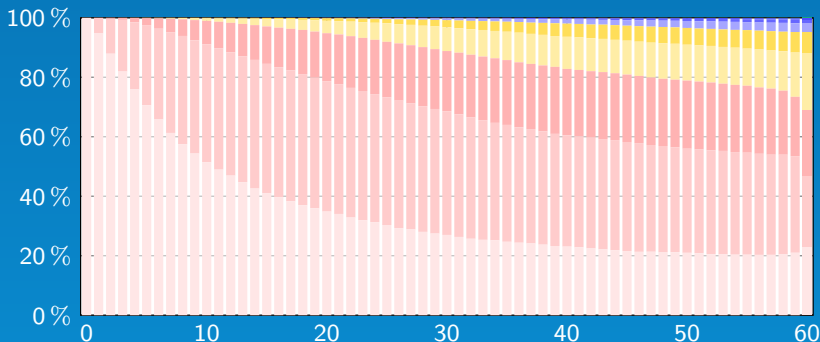
Simulace změn

Vývoj rozdělení počtu gólů během utkání



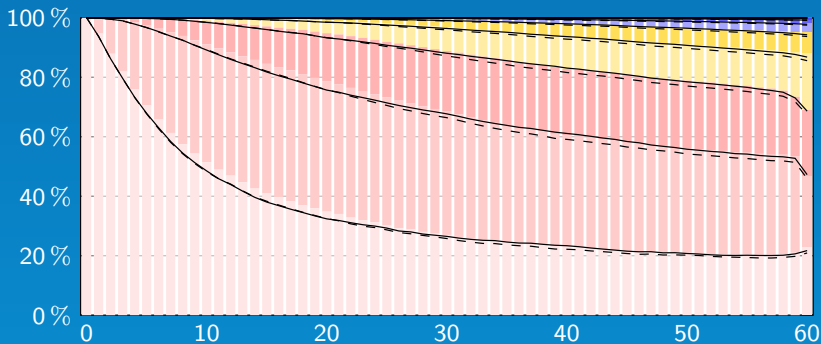
Simulace změn

Vývoj rozdělení velikosti rozdílu skóre během utkání



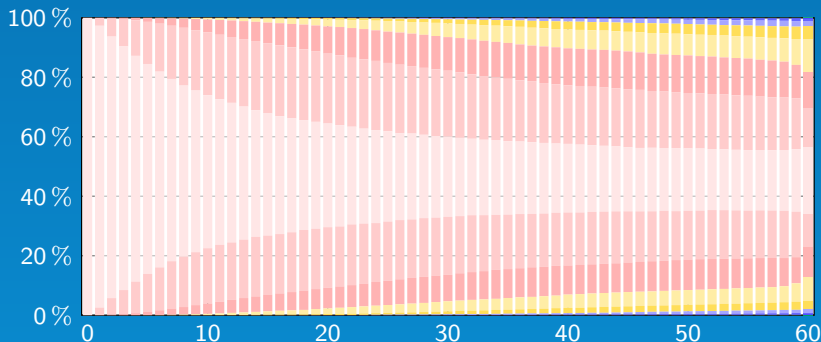
Simulace změn

Vývoj rozdělení velikosti rozdílu skóre během utkání



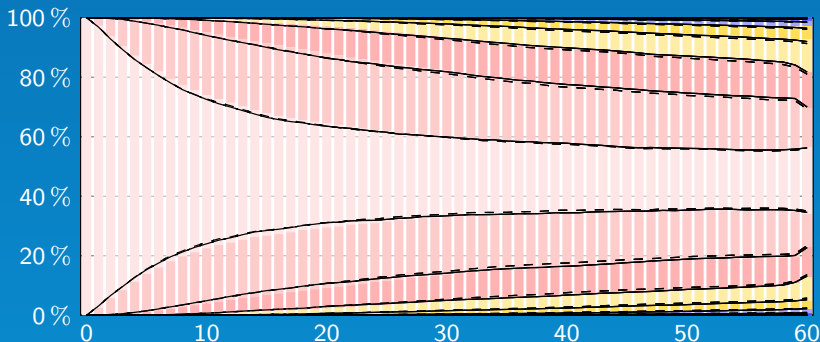
Simulace změn

Vývoj rozdělení rozdílu domácí – hosté během utkání



Simulace změn

Vývoj rozdělení rozdílu domácí – hosté během utkání





Friesl M., Lenten L. J. A.,
Libich J., Stehlík P.:

In search of goals: increasing ice hockey's
attractiveness by a sides swap

*Journal of the Operational Research So-
ciety 68 (2017), no. 9, 1006–1018*