

Probabilistische Bewertungsmodelle für Fußballergebnisse

Von Poisson-Intensitäten zu kalibrierten Ergebniswahrscheinlichkeiten

Klaus Hürbl

Unabhängiger Analyst für Quoten- und Marktmodellierung, Graz · Kontakt: kontakt@klaus-huerbl.de

ZUSAMMENFASSUNG

Die Bewertung von Fußballspielen verlangt keine Punktprognose, sondern eine Wahrscheinlichkeitsverteilung über alle möglichen Ergebnisse. Der Beitrag systematisiert die etablierte Modellfamilie — vom Poisson-Grundmodell nach Maher über die Niedrigergebnis-Korrektur von Dixon und Coles bis zum bivariaten Poisson-Ansatz — und zeigt an einem nachvollziehbaren Fallbeispiel, wie sich die Verschiebung einzelner Modellparameter (Wegfall eines Schlüsselspielers, neutraler Spielort) quantitativ auf die abgeleiteten 1X2-, Über/Unter- und Korrektergebnis-Wahrscheinlichkeiten überträgt. Der Schwerpunkt liegt nicht auf Prognosegüte im Einzelfall, sondern auf der korrekten Übersetzung von Intensitäten in marktfähige, kalibrierte Wahrscheinlichkeiten.

Schlagwörter: Poisson-Modell · Dixon-Coles · bivariate Zähldaten · Ergebniswahrscheinlichkeit · Modellkalibrierung · Sportwettenmärkte

1. Einleitung

Ein Fußballergebnis ist kein deterministischer Wert, sondern die Realisierung eines Zufallsprozesses mit geringer Ereignisdichte. Die zentrale Aufgabe eines

Bewertungsmodells besteht daher nicht darin, ein einzelnes Resultat zu „tippen“, sondern die gesamte Verteilung über mögliche Ergebnisse $P(X = x, Y = y)$ zu schätzen — wobei X und Y die erzielten Tore von Heim- und Auswärtsmannschaft bezeichnen. Aus dieser Verteilung lassen sich anschließend sämtliche relevanten Märkte konsistent ableiten: die Dreiwegwette (1X2), Torsummenmärkte (Über/Unter), Beide-Treffen sowie das Korrektergebnis.

Diese Sichtweise ist seit über vier Jahrzehnten methodischer Standard. Der vorliegende Beitrag fasst die tragende Modellfamilie kompakt zusammen, ordnet ihre Erweiterungen ein und demonstriert vor allem den oft unterschätzten Schritt von der geschätzten Intensität zur kalibrierten Ergebniswahrscheinlichkeit. Auf eine erschöpfende Darstellung dynamischer oder vollständig bayesianischer Varianten wird bewusst verzichtet; sie werden lediglich verortet.

2. Das Poisson-Grundmodell

Den Ausgangspunkt bildet die Annahme, dass die Toranzahlen beider Mannschaften unabhängig poissonverteilt sind (Maher, 1982):

$$X \sim \text{Pois}(\lambda_H), \quad Y \sim \text{Pois}(\lambda_A) \quad (1)$$

Die Erwartungswerte λ_H und λ_A werden log-linear aus mannschaftsspezifischen Stärkeparametern modelliert. Mit Angriffsstärke α , Defensivstärke δ , einem Grundniveau μ und einem additiven Heimvorteil γ gilt:

$$\log \lambda_H = \mu + \alpha_i + \delta_j + \gamma, \quad \log \lambda_A = \mu + \alpha_j + \delta_i \quad (2)$$

wobei i die Heim- und j die Auswärtsmannschaft indiziert. Die Parameter werden per Maximum-Likelihood aus historischen Spielen geschätzt; eine Identifikationsrestriktion (etwa $\sum_i \alpha_i = 0$) hält die Skala fest. Die gemeinsame Ergebniswahrscheinlichkeit ergibt sich unter der Unabhängigkeitsannahme als Produkt:

$$P(X = x, Y = y) = \frac{e^{-\lambda_H} \lambda_H^x}{x!} \cdot \frac{e^{-\lambda_A} \lambda_A^y}{y!} \quad (3)$$

Das Modell ist sparsam, interpretierbar und überraschend leistungsfähig. Seine bekannte Schwäche: Die Unabhängigkeitsannahme unterschätzt empirisch die Häufigkeit knapper Resultate, insbesondere von Unentschieden im niedrigen Bereich (0:0, 1:1).

3. Erweiterungen

3.1 Niedrigergebnis-Korrektur und Zeitgewichtung (Dixon-Coles)

Dixon und Coles (1997) führen einen Korrekturfaktor τ ein, der ausschließlich die vier Ergebniszellen mit höchstens einem Tor je Seite anpasst:

$$P^*(x, y) = \tau_{\lambda_H, \lambda_A}(x, y) \cdot P(X = x, Y = y) \quad (4)$$

mit einem Abhängigkeitsparameter ρ , der 0:0 und 1:1 leicht aufwertet und 1:0 bzw. 0:1 entsprechend dämpft. Zweitens gewichten die Autoren ältere Spiele exponentiell ab:

$$w(t) = \exp(-\xi \Delta t) \quad (5)$$

Damit trägt aktuelle Form stärker zur Schätzung bei als weit zurückliegende Begegnungen. Die Wahl von ξ ist ein klassischer Bias-Varianz-Kompromiss und sollte über Kreuzvalidierung bestimmt werden.

3.2 Bivariates Poisson-Modell

Eine strukturell sauberere Behandlung der Abhängigkeit liefert das bivariate Poisson-Modell (Karlis & Ntzoufras, 2003). Es zerlegt die Torzahlen in eine gemeinsame Komponente Z_0 :

$$X = Z_1 + Z_0, \quad Y = Z_2 + Z_0, \quad Z_k \sim \text{Pois}(\theta_k) \quad (6)$$

Die Kovarianz beträgt $\text{Cov}(X, Y) = \theta_0 \geq 0$. Der Ansatz modelliert positive Korrelation explizit (etwa gemeinsame Spielintensität), kann aber die in der Praxis ebenfalls auftretende negative Abhängigkeit nicht abbilden — hier bleibt die pragmatische Dixon-Coles-Korrektur häufig konkurrenzfähig.

3.3 Dynamische Stärken (Verortung)

Mannschaftsständen sind nicht konstant. State-Space-Formulierungen lassen $\alpha_{i,t}$ und $\delta_{i,t}$ einem latenten Random-Walk folgen (Rue & Salvesen, 2000); bayesianische Umsetzungen liefern zusätzlich Unsicherheitsbänder. Für den Zweck dieses Beitrags genügt die Feststellung, dass solche Modelle dieselbe Ausgabestruktur — eine Ergebnismatrix — erzeugen und sich der folgende Übersetzungsschritt unverändert anwenden lässt.

4. Von Intensitäten zu Ergebniswahrscheinlichkeiten

Unabhängig vom gewählten Modell ist die Ausgabe stets eine bei $x, y = 0, 1, 2, \dots$ ausgewertete Wahrscheinlichkeitsmatrix $P^*(x, y)$. In der Praxis genügt eine Abschneidung bei acht Toren je Seite; die Restmasse ist vernachlässigbar. Aus dieser Matrix werden alle Märkte durch Summation der zutreffenden Zellen gebildet:

$$P(\text{Heim}) = \sum_{x>y} P^*(x, y), \quad P(\text{Remis}) = \sum_{x=y} P^*(x, y), \quad P(\text{Über } 2,5) = \sum_{x+y \geq 3} P^*(x, y)$$

Damit ist der entscheidende Punkt benannt: Ein Modell prognostiziert nicht „Heimsieg“, sondern weist jeder Ergebniszelle eine Masse zu, deren Aggregation die Marktwahrscheinlichkeit ergibt. Erst die Umkehrung $o = 1/p$ führt zur fairen Quote — die im Markt beobachtete Quote enthält zusätzlich die Marge des Anbieters.

5. Fallbeispiel: Wie Parameter die Ergebnisse verschieben

Das folgende, bewusst einfach gehaltene Beispiel illustriert die Wirkungskette. Es dient der Veranschaulichung des Mechanismus, nicht einer realen Prognose. Aus den Stärkeparametern einer fiktiven Begegnung ergeben sich nach Gleichung (2) die Intensitäten $\lambda_H = 2,14$ und $\lambda_A = 1,65$. Tabelle 1 zeigt den oberen Ausschnitt der resultierenden Ergebnismatrix nach Gleichung (3); die fünf wahrscheinlichsten Resultate sind hervorgehoben.

Tabelle 1: Ergebnismatrix $P^*(x, y)$ in Prozent ($\lambda_H = 2,14$, $\lambda_A = 1,65$), Ausschnitt.

H \ A	0	1	2	3	4
0	2,3	3,7	3,1	1,7	0,7
1	4,8	8,0	6,6	3,6	1,5
2	5,2	8,5	7,0	3,9	1,6
3	3,7	6,1	5,0	2,7	1,1
4	2,0	3,2	2,7	1,5	0,6

Die Aggregation nach Gleichung (7) liefert die Baseline der Marktwahrscheinlichkeiten in Tabelle 2. Anschließend werden zwei Eingriffe modelliert: (a) der Wegfall eines Schlüsselspielers im Angriff der Heimmannschaft, abgebildet als Reduktion der Heim-

Angriffsstärke um rund 15 % auf der Intensitätsskala ($\lambda_H : 2,14 \rightarrow 1,82$); (b) die Austragung am neutralen Ort, also $\gamma = 0$.

Tabelle 2: Abgeleitete Wahrscheinlichkeiten je Szenario und Veränderung gegenüber Baseline (Prozentpunkte).

MARKT	BASELINE	(A) SCHLÜSSELSPIELER FEHLT	Δ	(B) NEUTRALER ORT	Δ
Heimsieg	49,2 %	42,3 %	-6,9	38,5 %	-10,7
Unentschieden	20,8 %	22,3 %	+1,6	23,0 %	+2,2
Auswärtssieg	30,0 %	35,4 %	+5,4	38,5 %	+8,5
Über 2,5 Tore	72,9 %	~67 %	-6	~66 %	-7

Der Mechanismus ist damit quantitativ greifbar. Eine 15-prozentige Reduktion der Heim-Angriffsintensität verschiebt nicht nur den Heimsieg (-6,9 Prozentpunkte), sondern verteilt diese Masse asymmetrisch: Der Auswärtssieg profitiert deutlich stärker (+5,4) als das Unentschieden (+1,6), weil eine niedrigere Heimintensität die Verteilung in Richtung knapper und auswärtsdominierter Ergebnisse zieht. In faire Quoten übersetzt, bedeutet das eine Bewegung des Heimsiegs von 2,03 auf 2,37 — eine Differenz, die jede Bewertung entscheidet. Der neutrale Spielort wirkt in dieselbe Richtung, aber stärker, und führt im Beispiel zur erwarteten Symmetrisierung von Heim- und Auswärtswahrscheinlichkeit (je 38,5 %).

Die praktische Lehre: Bewertungsrelevant ist nicht die Parameteränderung selbst, sondern wie sie sich nach Aggregation über die Ergebnismatrix auf die einzelnen Märkte verteilt — und diese Verteilung ist regelmäßig asymmetrisch.

6. Kalibrierung und Validierung

Ein Modell ist nicht an der Trefferquote einzelner Spiele zu messen, sondern an der Kalibrierung seiner Wahrscheinlichkeiten: Ereignisse, denen 30 % zugewiesen werden, sollten langfristig in etwa 30 % der Fälle eintreten. Geeignete Maße sind der Brier-Score und der logarithmische Verlust:

$$\text{Brier} = \frac{1}{N} \sum_{n=1}^N \sum_k (p_{nk} - y_{nk})^2 \quad (8)$$

ergänzt um eine Reliabilitätskurve (beobachtete vs. vorhergesagte Häufigkeit) und einen Vergleich gegen die markt-implizierten, margin-bereinigten Wahrscheinlichkeiten als anspruchsvolle Referenz. Ein Modell, das den Markt nicht schlägt, ist nicht wertlos — aber es ist kein Bewertungsvorteil.

7. Grenzen

Drei Einschränkungen sind redlich zu benennen. Erstens behandeln die dargestellten Grundmodelle Stärken über ein Spiel hinweg als konstant und ignorieren Innerspiel-Dynamik (Rote Karten, frühe Führung). Zweitens sind Punktschätzer der Parameter mit Unsicherheit behaftet; ohne deren Propagation werden die Ergebniswahrscheinlichkeiten zu selbstsicher ausgewiesen. Drittens ersetzt kein Modell die Sachprüfung der Eingangsdaten: Eine fehlerhaft kodierte Aufstellung verzerrt jede noch so saubere Likelihood. Die Modellklasse ist robust — sie ist nicht immun gegen schlechte Daten.

8. Fazit

Probabilistische Bewertungsmodelle für Fußballergebnisse bilden eine ausgereifte, sparsame und gut verstandene Familie. Der methodische Kern liegt weniger in der Wahl zwischen Poisson, Dixon-Coles oder bivariatem Ansatz als in der disziplinierten Übersetzung geschätzter Intensitäten in kalibrierte, aggregierte Marktwahrscheinlichkeiten — und im Bewusstsein, dass Parameteränderungen sich asymmetrisch über die Märkte verteilen. Wer diesen Schritt sauber beherrscht und konsequent gegen den Markt validiert, besitzt ein belastbares Bewertungswerkzeug; wer ihn überspringt, produziert lediglich präzise wirkende Scheingenauigkeit.

Literatur

1. **Dixon, M. J., & Coles, S. G.** (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. *Journal of the Royal Statistical Society: Series C*, 46(2), 265–280.
2. **Karlis, D., & Ntzoufras, I.** (2003). Analysis of Sports Data by Using Bivariate Poisson Models. *Journal of the Royal Statistical Society: Series D*, 52(3), 381–393.
3. **Maher, M. J.** (1982). Modelling Association Football Scores. *Statistica Neerlandica*, 36(3), 109–118.
4. **Rue, H., & Salvesen, Ø.** (2000). Prediction and Retrospective Analysis of Soccer Matches in a League. *Journal of the Royal Statistical Society: Series D*, 49(3), 399–418.

5. **Constantinou, A. C., & Fenton, N. E.** (2012). Solving the Problem of Inadequate Scoring Rules for Assessing Probabilistic Football Forecast Models. *Journal of Quantitative Analysis in Sports*, 8(1).

Manuskriptfassung des Autors. Korrespondenz: kontakt@klaus-huerbl.de · Graz, Österreich. Der Autor erklärt keine Interessenkonflikte; der Beitrag enthält keine Wetttempfehlung.