

Applications of Extreme Value Theory in Nippon Professional Baseball

Subaru Ishizuka, Yosuke Jumonji, Kazuki Kawase
Faculty of Economics, Keio University

Abstract

In various sports, the records that have achieved the most or best events in history, such as Japanese records and world records, are often noted. These records are only relatively evaluated within each event, but are never compared across each event. In this study, we estimate the current Japanese record reproduction period by applying extreme value theory to the highest record of each season for batting average, hits, home runs, hit points, and stolen bases of NPB. Using the reproduced period as a unified index, it was possible to compare the "greatness" of each Japanese record of batting average, hits, home runs, hit points, and stolen bases. Furthermore, we estimated the reproduction level with a reproduction period of 100 years and the upper limit of recording.

日本プロ野球における極値理論の応用

石塚 昂・十文字 陽亮・川瀬 一貴

慶應義塾大学経済学部長倉大輔研究会

要旨

様々なスポーツにおいて、日本記録・世界記録といった、歴史上で最も多く、または最も良く、ある事象を実現した記録がよく注目される。それらの記録は、それぞれの事象の中での記録が相対的に評価されるだけに留まるが、それぞれの事象を超えて比較されることはない。本研究では、NPBの打率、安打数、本塁打数、打点数、盗塁数について、各シーズンの最高記録に対して極値理論を適用させることにより、現在の日本記録の再現期間を推定し、得られた再現期間を統一的な指標として、打率、安打数、本塁打数、打点数、盗塁数の各日本記録の「偉大さ」を比較することを可能にした。さらに、再現期間100年の再現レベル、記録の上限値を推定した。

キーワード

一般極値分布, 非定常, トレンド, 再現期間, 再現レベル, 日本プロ野球

1. はじめに

アメリカでは、スポーツにおける（特に、MLBにおいて）統計学的分析が発展、普及している。例えば、MLBの全スタジアムにはStatcastと呼ばれるプレイヤートラッキングシステムが設置されている。攻守問わず様々なプレーについてのデータがリアルタイムで記録され、取得されたデータは選手や戦術の評価に活用されている。その影響を受けて日本プロ野球においても統計学的分析が行われることが増加している。しかしながら、これまでMLBや日本プロ野球をはじめとする野球界で用いられてきた統計では、打率・本塁打数など選手個人のパフォーマンスを、あるシーズンの中で他の選手との比較で相対的に評価することがほとんどであった。そこで本研究では、シーズンの最高記録に注目して極値理論の応用を試みた。過去のシーズン最高記録のデータからその分布を仮定し、ある記録が達成される確率がどの程度なのか評価したり、さらに一定期間における最高記録を推定したりした。つまり、ある記録について、1シーズンを超えた長期的な視野から絶対評価を行う方法を実行したということである。

2. プロ野球における統計学の応用の歴史

(1) 世界における野球と統計学

鳥越(2012)によると、野球というスポーツに対して、統計的手法での研究がなされたのは、1959年のLindseyによるものが初めてだとされている[7]。それ以降、約60年で多くの統計的研究が世界中でなされており、Google Scholarにて、“baseball data”と検索すると314,000件の文献がヒットし、“baseball statistics”と検索すると約159,000件の文献がヒットすることがその膨大さを表しているだろう。

1970年代にはアメリカ野球において、セイバーメトリクスと呼ばれる、野球におけるデータを客観的に分析し、選手やチームを評価したり、戦略を決める概念が誕生した。2000年代初頭には、MLBのオークランド・アスレチックスが、本格的にセイバーメトリクスをチームの采配に活用することで良い結果を残し、世界で大きな反響を呼んだ。

そして現在、MLBではStatcastと呼ばれる、AIによって全自動・リアルタイムでデータを取得する技術が用いられ、それによって単なる一つ一つの打席での結果のみならず、1球1球の投げられたボールの回転数や、守備による打撃を追う走路など、非常に細かく選手自身も意識を強く向けてはいないデータすらも分析の対象となっている。

(2) 日本における野球と統計学

アメリカほどでは無いが、日本でも統計的手法での分析が進められている。Google Scholarで「野球 データ」と検索すると15,200件、「野球 分析」と検索すると15,500件の文献がヒットする。

株式会社グラッドキューブが運営するSPAIAでは、各試合のデータをまとめているだけでなく、勝敗予想をAIに行わせている。また、日本テレビとデータスタジアム社により、「AIキャッチャー」と呼ばれる、失点を防ぐための最適な球種とコースを、過去16年間のデータから予測する技術も開発され、野球中継の画面にそれを表示させることで、視聴者をより試合の行方に釘付けにしている。

このように、今では野球における統計学的手法は身近なものになっている。

3. 極値理論

(1) 極値理論の実用について

極値理論の目的は、「長期間または広領域等における最大値（または最小値）に関する推測である」[5]。

とある分布の母集団から観測される値のうち、一般的に焦点を当てられることの多いのは、平均値やその周辺の頻度の高い値であろう。しかし、極値理論が興味を持つのは、異常に大きかったり小さかったりする、いわゆる外れ値と呼ばれる値の出現である。この異常な値は大きなリスクを伴っていることがあり、その分布の予測をすることで、リスクを事前に最小限に抑えられる。

川越ら(2008)は、再現期間10年、30年、100年に対する降雨極値を用いることで、将来の土砂災害の発生確率を導き出し、日本列島を対象にした、土砂災害の発生確率分布を用いたハザードマップを作成した[1]。壇ら(1986)は、上下限値を有する極値分布に対して、日本の各地点で記録される観測値の年最大値を当てはめることで、日本における地震危険度マップを作成した[6]。

上記のように、極値理論は古くから水文学・工学の分野で、そして近年では環境や金融工学の分野などで、主に防災を目的として扱われている[5]。

Einmahlら(2008)は、男女それぞれ陸上競技13競技におけるトップアスリートの個人ベスト記録を集め、それらに極値理論を適用させることによって、それぞれの競技における限界値を推測した[9]。例を挙げると、当時の100m走の世界記録は、ジャマイカのアサファ・パウエル氏の記録した9.74秒であったが、当論文における推測値は9.29秒である。その推測から12年経った今でも、100m走世界記録は、2009年にジャマイカのウサイン・ボルト氏が記録した9.58秒であることから、到達することが容易ではないことが伺える。また、渋谷ら(2003)は、連続変数に対して議論されている極値理論を、厚生労働省によって発表されている年齢時代区分別のデータに対して適用するために新たな方法を提案し、日本人男性寿命分布における寿命限界を推測した[2]。中でも生存者数統計の解析において、寿命限界は134歳、死亡者数統計の解析では125-127歳と推測している。

上記のように、単に災害の被害最小化のためだけでなく、とある事象における限界値を推測するという目的でも利用される。

(2) 極値統計量

極値統計学において使われる極値統計量(極値データ)には大きく分けて、区分最大データ (block maxima)と水準超過データ (Peaks Over Threshold) がある。前者は「標本を同じ大きさのブロックに分けたときの各ブロックごとの最大値データ」[4]であり、本研究ではこちらの極値統計量を使用する。

(3) 極値分布

独立で同一な分布Fに従う確率変数列 X_1, X_2, \dots, X_n に対して、極値統計量 Z_n を

$$Z_n = \max \{X_1, X_2, \dots, X_n\}$$

とする。

このとき非退化な分布へ収束するように次のような基準化を考える。

定数列 $a_n > 0$, $b_n \in \mathcal{R}$ ($n = 1, 2, \dots$)と非退化な分布Gを持つ確率変数Zが存在して、 $n \rightarrow \infty$ のとき

$$\frac{Z_n - b_n}{a_n} \rightarrow^d Z : \text{分布収束}$$

すなわち

$$P\left(\frac{Z_n - b_n}{a_n} \leq x\right) \rightarrow P(Z \leq x) = G(x), \quad \forall x: G \text{の連続点}$$

また

$$\begin{aligned} P\left(\frac{Z_n - b_n}{a_n} \leq x\right) &= P(Z_n \leq a_n x + b_n) \\ &= F^n(Z_n \leq a_n x + b_n) \rightarrow^L G(x) \end{aligned}$$

このとき、 G を極値分布、 a_n, b_n を基準化係数と呼び、 F は G の吸引領域に属するという。極値分布は以下の3つの型になる事が知られている。詳しい証明はde Haan (1976)などを参照。

$$\text{Gumbel分布: } \Lambda(x) = \exp(-\exp(-x)), \quad x \in \mathbb{R}$$

$$\text{Fréchet分布: } \Phi_\alpha(x) = \exp(-x^{-\alpha}), \quad x \geq 0, \alpha > 0$$

$$\text{Weibull分布: } \Psi_\alpha(x) = \exp(-(-x)^\alpha), \quad x \leq 0, \alpha > 0$$

(4) 一般極値分布

3つの極値分布は1つのパラメータ($\xi \in \mathbb{R}$)を持つ分布で表される。これを一般極値(generalized extreme value, GEV)分布とよぶ。一般極値分布の分布関数は

$$G_\xi(x) = \exp\left\{-\left(1 + \xi x\right)^{-\frac{1}{\xi}}\right\}, \quad 1 + \xi x > 0$$

である。ここで $\xi=0$ の場合には

$$G_0(x) = \lim_{\xi \rightarrow 0} G_\xi(x) = \exp\{-\exp(-x)\}$$

とする。

一般極値分布はパラメータ ξ の値で3つの極値分布を表しており、 $\xi < 0$ でWeibull分布、 $\xi = 0$ でGumbel分布、 $\xi > 0$ でFréchet分布に対応する。

(5) 再現期間と再現レベル

統計量 X が、ある確率密度関数 $g(x)$ に従い、独立に発生すると仮定する。このとき X が x_p 以下となる確率 $G(x_p)$ は

$$G(x_p) = \int_{-\infty}^{x_p} g(x) dx$$

で与えられる。このとき、 $G(x_p)$ を非超過確率 p といい、 $1-p$ を超過確率という。

X が x_T 以上の値をとることが、 T 年に平均1度生起すると期待されるとき、 T を再現期間といい、 x_T を再現期間 T 年の再現レベル(再現期待値)という。再現期間 T と非超過確率 p の間には、以下の関係がある。

$$T = \frac{1}{1-p} = \frac{1}{1-G(x_T)}$$

4. 使用データ

本研究では、NPBの打率、本塁打数、打点数、安打数、盗塁数について1950年から2019年までの70年間のシーズン最高記録を集計した。当研究では、各ブロックを1年とした区分最大デー

データを極値データとして使用する。分析(1)-(4)ではそれらのデータから日本記録を除いたデータを使用し、分析(5)ではすべてのデータを使用する。

打率が丸められた値であることも考慮すれば、今回得られたデータは全て離散量である。離散確率分布は、一般に極値分布の吸引領域に属さないが、志村(2015)において、多くの離散分布で極値分布の吸引領域に属するような連続化が存在することが示されている。本研究では、そのような連続化がどのようなものであるか分かっていないこと、身長データや体重データなどは離散量だが便宜的に連続量として扱われる場合が多いこと、今回得られたデータの有効桁数がそれなりに多いこと、から連続化は行わずに今回得られたデータを連続量としてもとの分布 F が極値分布の吸引領域に属すると仮定して分析を行う。

5. 分析

(1) 一般極値(GEV)モデルの推定

打率、本塁打数、打点数、安打数、盗塁数の年最大値 z の累積分布関数が以下の(1)式で表されるGEV分布(μ, σ, ξ)に従うと仮定して、日本記録を除いたデータを用いて最尤推定を行った。最尤推定値($\hat{\mu}, \hat{\sigma}, \hat{\xi}$)をまとめたのが、表-1である。

$$G(z) = \exp\left\{-\left(1 + \xi \frac{z-\mu}{\sigma}\right)^{\frac{1}{\xi}}\right\} \quad (1)$$

形状パラメータ $\xi < 0$ であるため、打率、本塁打数、打点数、安打数、盗塁数の年最大値は裾の薄く上限値が存在するWeibull分布に従う。

表-1 GEV分布パラメータの最尤推定値

	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\xi}$
打率	0.345(0.0019)	0.014(0.0013)	-0.187(0.085)
本塁打数	40.6(1.02)	7.7(0.775)	-0.47(0.085)
打点数	110(1.61)	11.8(1.16)	-0.112(0.094)
安打数	169(1.86)	13.4(1.36)	-0.148(0.103)
盗塁数	48.5(1.5)	10.9(1.09)	-0.02(0.096)

(2) 当てはまりの診断

分析(1)において打率、本塁打数、打点数、安打数、盗塁数それぞれの極値データを用いて最尤推定したGEVモデルの診断図(確率プロット、分位プロット、再現レベルプロット、密度関数プロット)を図-1、図-2、図-3、図-4、図-5に示す。

推定分布関数を \hat{G} 、データを降順に並べ替えて $z_{(1)} \leq z_{(2)} \leq \dots \leq z_{(n)}$ とする。左上図の確率点プロットでは、以下の(2)で表される点を実線として表されている。右上図の分位点プロットでは、以下の(3)で表される点を実線として表されている。

$$\left\{ \left(\hat{G}(z_{(i)}), \frac{i}{n+1} \right) : i = 1, \dots, n \right\} \quad (2)$$

$$\left\{ \left(\hat{G}^{-1}\left(\frac{i}{n+1}\right), z_{(i)} \right) : i = 1, \dots, n \right\} \quad (3)$$

右下図の再現レベルプロットでは、推定分布関数から得られた再現期間に対する再現レベルが実線として表されている。確率プロット，分位プロット，再現レベルプロットにおける点プロットはそれぞれの実現値を表し，点線は推定値の95%信頼区間を表す。左下図の密度関数プロットでは，推定密度関数が実線として，実現値に基づいたノンパラメトリック推定が点線として表されている。

点プロットが実線から系統的にずれている場合は，モデルの当てはまりが悪いと判断することが出来る。打率，本塁打数，打点数，安打数，盗塁数全てのGEVモデルにおいて，実現値が信頼区間内におさまっており，診断図からは当てはまりが悪いとはいえない。

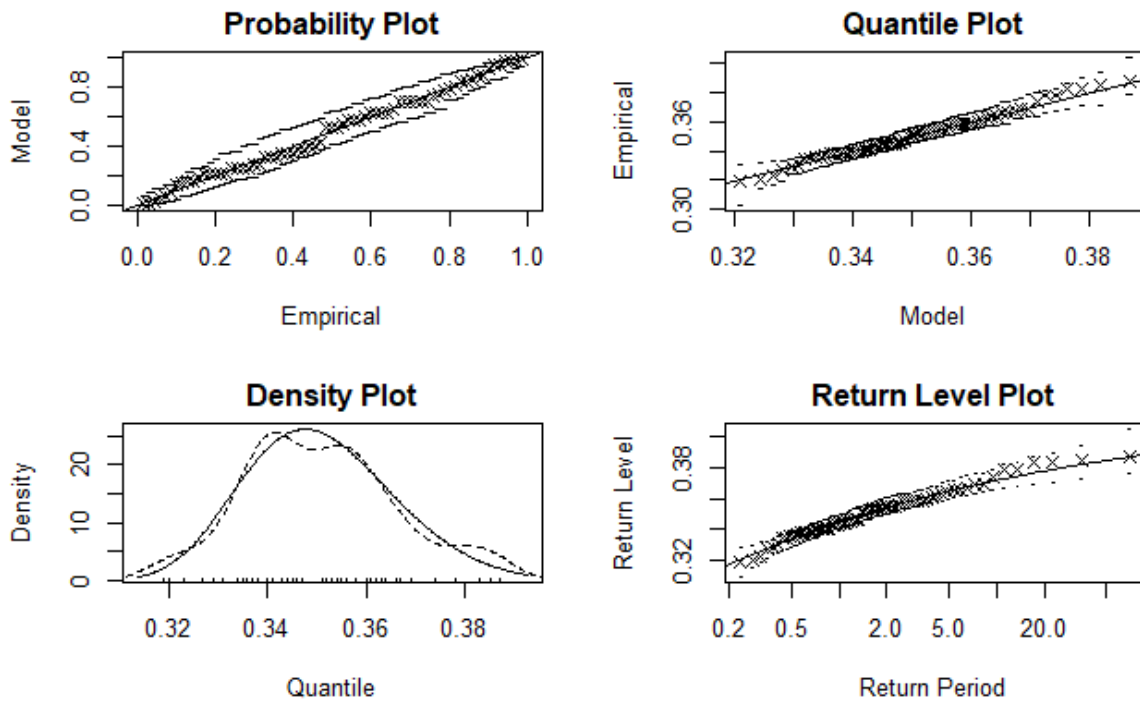


図-1 打率の診断図

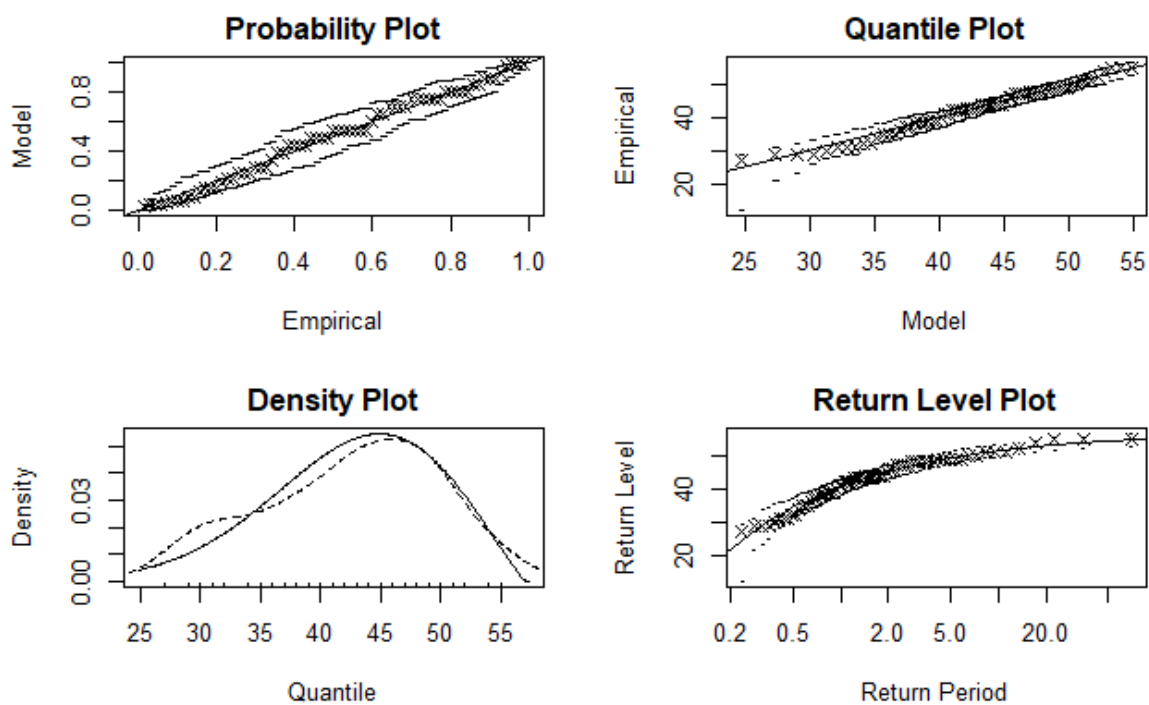


図-2 本塁打数の診断図

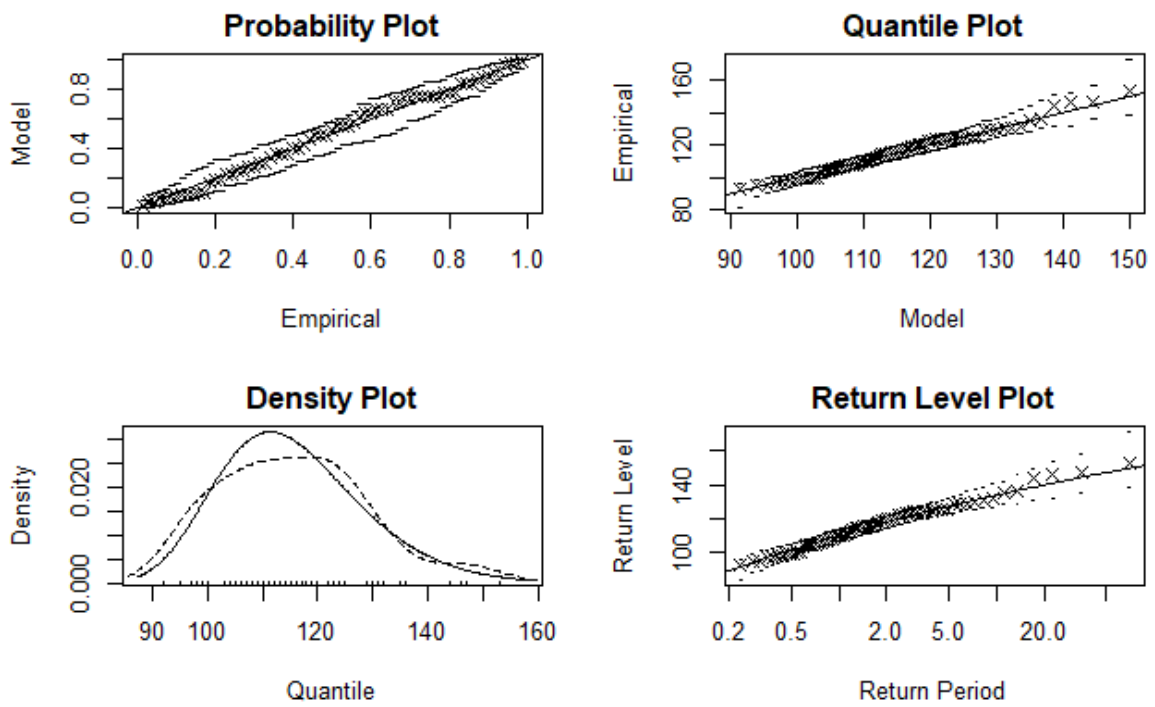


図-3 打点数の診断図

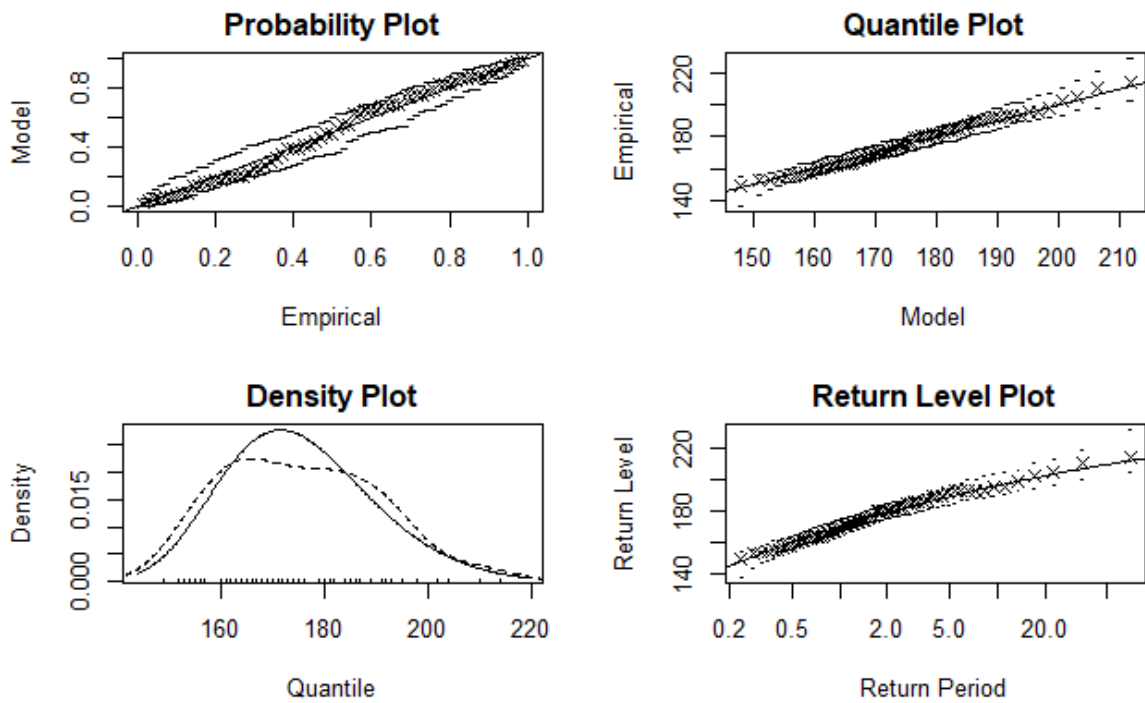


図-4 安打数の診断図

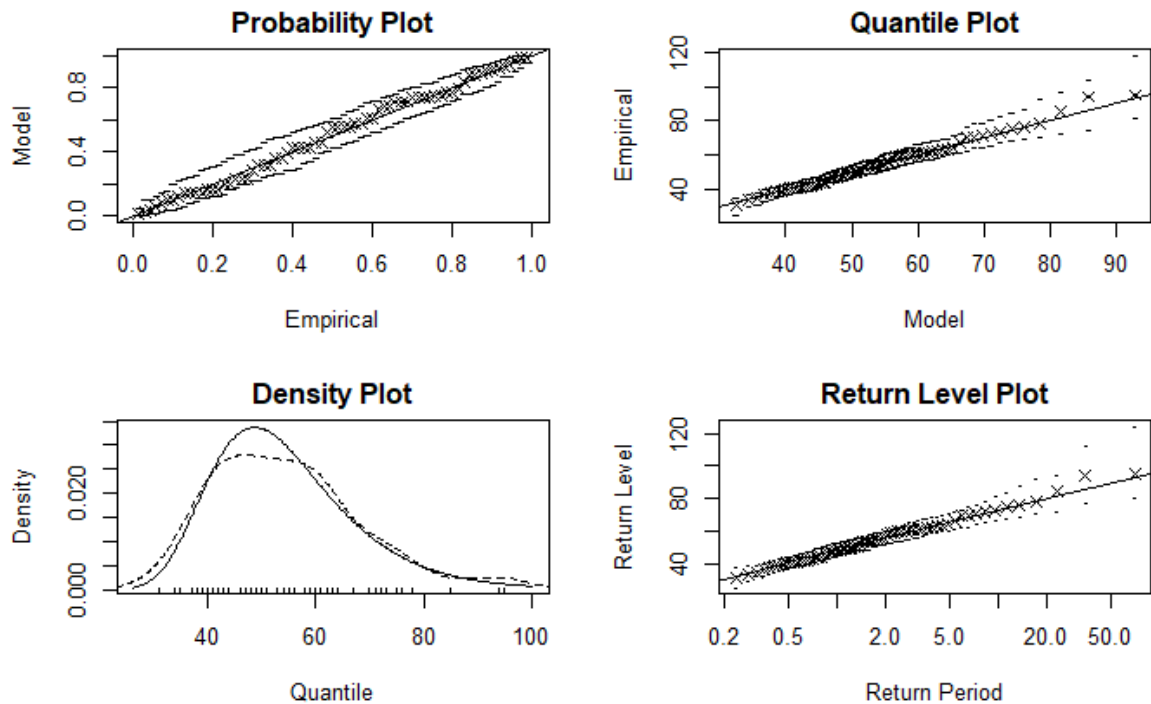


図-5 盗塁数の診断図

(3) 非定常GEVモデルの導入

分析(2)において診断図からGEVモデルへの当てはまりを診断したが、モデルへの当てはまりが悪い場合でも診断図にそれが現れない場合があり、診断図だけでモデルの当てはまりが良いと判断すべきではない。そこで、打率、本塁打数、打点数、安打数、盗塁数の年最大値を時系列でプロットしたのが、図-6である。この図から、年最大値が明らかにトレンドを持つことが分かり、定常を仮定したGEVモデルの適用は必ずしも適切ではない。

GEVモデルの位置パラメータ μ を $\mu(t)$ として年数 t に依存させた非定常GEVモデル(5)式を導入する。

$$G(z) = \exp \left\{ - \left(1 + \xi \frac{z - \mu(t)}{\sigma} \right)^{-\frac{1}{\xi}} \right\} \quad (5)$$

本研究では、位置パラメータ $\mu(t)$ を

$$\text{定常} : \mu(t) = \beta_0 \quad (6)$$

$$\text{1次式} : \mu(t) = \beta_0 + \beta_1 t \quad (7)$$

$$\text{2次式} : \mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2 \quad (8)$$

$$\text{3次式} : \mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 \quad (9)$$

として、(5)式に対して最尤推定を行う。ここで、年数 t に関して、そのままの年数を使用すると値が大きくなりすぎるため、以下の(10)式により[1950, 2019]を[0, 1]に線形変換する。

$$t = \frac{i-1950}{2019-1950}, \quad (i = 1950, 1951, \dots, 2019) \quad (10)$$

$\mu(t)$ を定常、1次式、2次式、3次式とした各モデルの下でのBICをまとめたのが、表-2である。打率、本塁打数、打点数、安打数、盗塁数においてBICを最小にするようなモデルは、それぞれ $\mu(t)$ が定数、2次、1次、3次、1次のモデルである。

打点数に関して、全体として当てはまりが良い（BICを最小にする）モデルは $\mu(t)$ が1次のモデルである。しかし、再現期間を求めたい1950年は、それ以前のトレンドを把握できないことから、1950年のパラメータ推定において、BICを最小にする非定常モデルが適当であるとはいえない。実際、図-6の時系列プロットを見ると1950年はそれ以降の数年と比較して大きく外れた値となっている。そのため、打点数では、BICが最小となる $\mu(t)$ が1次のモデルではなく、 $\mu(t)$ が定数のモデルを選択する。本塁打数、打点数、安打数、盗塁数ではBICが最小となるモデルを選択する。

選択したモデルのパラメータの最尤推定値と最大対数尤度をまとめたのが、表-3である。選択したモデルの位置パラメータの時系列プロットが、図-6の点線である。これから、それぞれのモデルは大まかなトレンドをとらえているといえる。

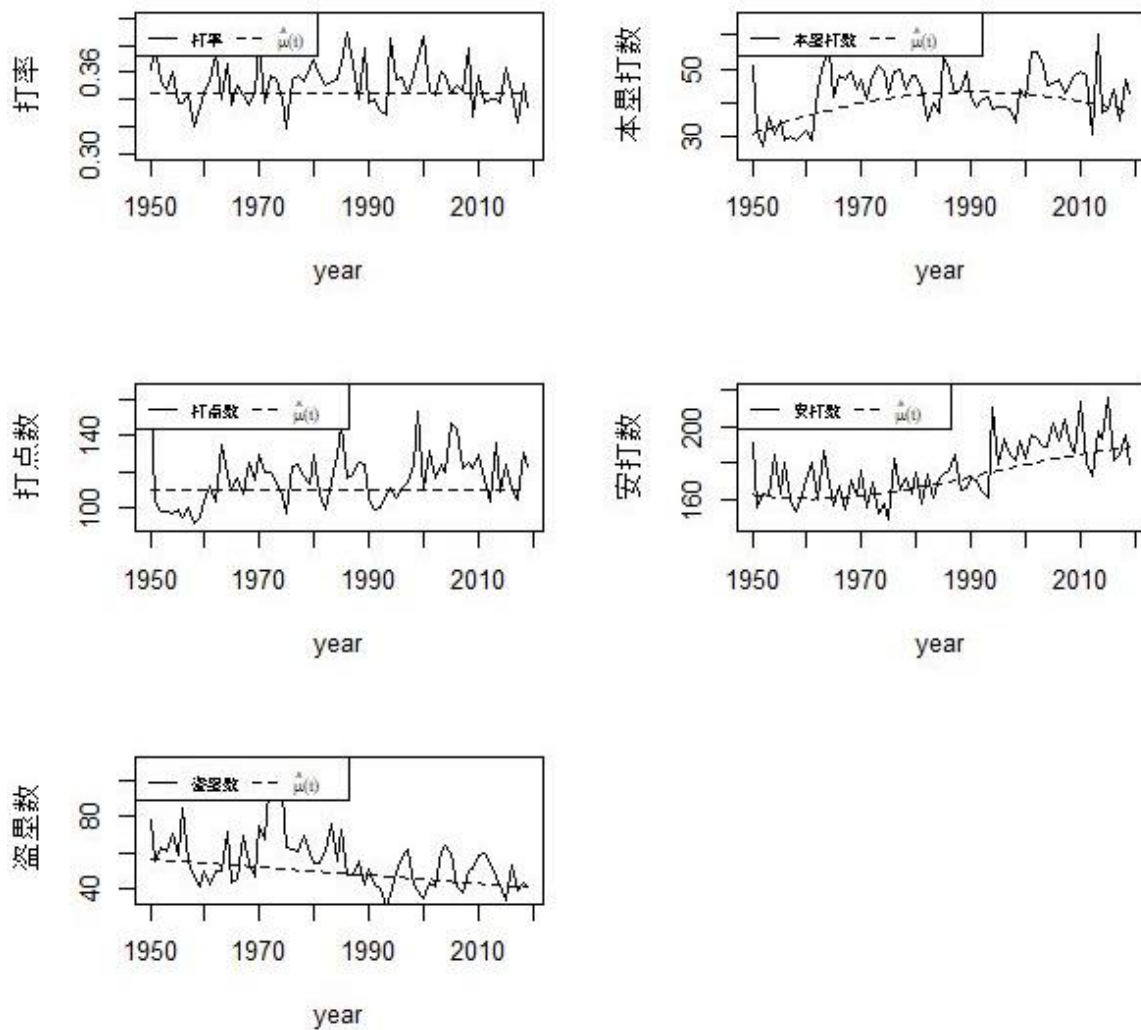


図-6 年最大値と位置パラメータの時系列プロット

表-2 位置パラメータが定数，1次，2次，3次のモデルにおけるBIC

	位置パラメータ $\mu(t)$ の次数			
	定数	1次	2次	3次
打率	-369.8	-365.7	-363.3	-363.1
本塁打数	475.5	477.7	472.5	473.9
打点数	562.7	550.1	551.4	555.7
安打数	578.2	548.6	548.2	546.1
盗塁数	559.0	551.5	555.7	559.8

表-3 選択したモデルの最尤推定値と最大対数尤度

	パラメータ数	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\sigma}$	$\hat{\xi}$	最大対数尤度
打率	3	0.345	-	-	-	0.0143	-0.187	191.3
本塁打数	5	30.0	41.5	-35.0	-	6.03	0.188	-225.6
打点数	3	110	-	-	-	11.8	0.113	-265.1
安打数	6	27.1	-47.6	156	-85.4	9.41	0.100	-260.3
盗塁数	4	56.6	-15.1	-	-	10.1	0.0320	-267.2

(4) 再現期間の推定と比較

分析(3)で選択したモデルを用いて各日本記録の再現期間を推定，比較する。

GEV分布の分布関数を $G(x)$ とする。このとき，ある値 z の再現期間 T は，以下の(11)式から求めることができる。

$$T = 1/(1 - G(z))$$

$$= 1/\left(1 - \exp\left\{-\left(1 + \xi \frac{z-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right\}\right) \quad (11)$$

打率，本塁打数，打点数，安打数，盗塁数の最高記録とその再現期間をまとめたのが，表-4である。再現期間の大きさから，打点，盗塁数，本塁打数，打率，安打数の順で「偉大な」記録であるといえる。再現期間に基づいた「偉大さ」とは，極値データと比較してどれだけ極値に位置しているか，すなわち，非凡な記録と比較しても傑出した記録であるかと解釈できる。打点数と盗塁数の再現期間が他の再現期間と比較してずば抜けて長い，1950年と1972年に打ち立てられた記録が2020年現在においても破られていないことやその年のもう一方のリーグの最高記録がそれぞれ106と45であったことから，直感と矛盾しない。

表-4 各最高記録の再現期間

	年数	最高記録	上側確率	再現期間(年)
打率	1986	0.389	0.009991	100
本塁打数	2001	60	0.000586	214
打点数	1950	161	0.004671	374
安打数	2015	216	0.022204	45
盗塁数	1972	106	0.002741	365

(5) 再現期間100年の再現レベルと上限値の推定

日本記録を含めた全ての極値データが定常GEV分布 (1) 式に従うと仮定して、最尤推定を行う。分析(1)において、診断図からは定常モデルへの当てはまりが良いといえること、そして非定常モデルを導入した場合どの年のパラメータを採用するかという問題が発生することから、本節では定常GEVモデルを採用した。

GEV分布(μ, σ, ξ)の $1 - 1/T$ 確率点 Z_T 、すなわち、 $G(Z_T) = 1 - 1/T$ となる確率点 Z_T は以下の(12)式で表される。

$$Z_T = G^{-1}(1 - 1/T)$$

$$= \mu + \sigma \left[\{-\log(1 - 1/T)\}^{-1} - 1 \right] / \xi, \quad \xi \neq 0 \quad (12)$$

式(12)における確率点 Z_T は再現期間(return period) T の再現レベル(return level)と呼ばれる。以下で推定する再現期間100年の再現レベルとは、再現期間100年のうちに平均1度生起すると期待される値と解釈できる。最尤推定したGEVモデルを用いて推定した再現期間100年の再現レベルと95%信頼区間をまとめたのが、表-5である。

形状パラメータの最尤推定値 $\hat{\xi}$ とプロファイル尤度95%信頼区間をまとめたのが、表-6である。打率、本塁打数、打点数、安打数は本塁打数を除き全て95%信頼区間が0をまたいでいるものの、形状パラメータの最尤推定値 $\hat{\xi} < 0$ であるため、打率、本塁打数、打点数、安打数の年最大値は上限値が存在するWeibull分布に従い、上限値を推定することが出来る。GEV分布の上限値 ω は以下の(13)式で表される。

$$\omega = \mu + \sigma / (-\xi) \quad (13)$$

打率、本塁打数、打点数、安打数、盗塁数の年最大値の上限値をまとめたのが、表-7である。盗塁数は、形状パラメータの最尤推定値 $\hat{\xi} > 0$ であるが、信頼区間が負の値を含むため、 $\hat{\xi} = -0.1$ として上限値を推定した。しかし、ここで得られた上限値はあくまでも数学上のものであり、参考程度に考えるべきである。実用を考える場合には、実際のデータに合うように ξ を調節する必要がある。

表-5 再現期間100年の再現レベルと95%信頼区間

	推定値	95%信頼区間
打率	0.392	[0.378, 0.405]
本塁打数	58	[56, 61]
打点数	159	[141, 167]
安打数	219	[201, 236]
盗塁数	104	[81, 127]

表-6 形状パラメータの最尤推定値と95%信頼区間

	$\hat{\xi}$	95%信頼区間
打率	-0.169	[-0.330, 0.013]
本塁打数	-0.338	[-0.453, -0.187]
打点数	-0.0566	[-0.211, 0.158]
安打数	-0.113	[-0.290, 0.123]
盗塁数	0.0332	[-0.128, 0.242]

表-7 分布の上限値

	上限値
打率	0.432
本塁打数	63
打点数	322
安打数	291
盗塁数	170

5. 最後に

本研究では、日本プロ野球の打者個人記録に対して極値理論を応用し、再現期間による各日本記録の「偉大さ」の比較、再現期間100年の再現レベルと記録の上限値の推定を行った。

異なる記録を統一的な指標で比較する方法としては、様々な方法（例えば平均値との距離など）が考えられるが、再現期間は、（1）全体の記録ではなく極値の記録と比較して導かれる点、（2）時系列トレンドを考慮することができる点、（3）何年に1度生起されると期待される記録であるかを示しており、指標自体が理解しやすく面白い点、以上の3点から興味深い指標であると考えている。

一方再現期間を指標として用いることの問題点として、（1）データ数が少ないと推定精度が低下すること、（2）極値モデルへの適合度によって再現期間が変化すること、以上の2点挙げられる。（1）に関して、今回使用したGEV分布であれば、各ブロックの最大値のみで推定が可能であるため、比較的データの入手は容易である。（2）に関して、モデルへの適合度を考慮するような指標にするのが望ましい。また極値モデルへの適合度によって再現期間がどの程度変化するかを検討する必要がある。これらは今後の課題としたい。

現在の日本プロ野球では、シーズンMVPが記者投票によって決定されている。本研究を応用して、各選手の個人成績の再現期間をMVP選定の基準の1つとして用いることで、より客観的かつ多面的にMVPの選定ができる。これは、他のスポーツでも同様である。

今回得られたデータが各年の最高記録のみであったため、一般極値分布を当てはめた。しかし、より多くのデータ（例えば各年の上位複数個の記録など）を得ることが出来れば、rGEVモデルやGPモデルなどの他の極値理論を用いて、より精度の高い推定ができる可能性がある。また、今回は非定常モデルにおいて位置パラメータ μ のみを年数 t に依存させたが、他パラメータの非定常のモデルも検討する必要がある。これは今後の課題としたい。

6. 謝辞

論文の作成にあたってご助言いただいた長倉大輔教授に深く感謝いたします。

各種推定やグラフの作成においてRのパッケージevd v2.3-3 (A. G. Stephenson 2002) とextRemes v2.0-12 (Eric Gilleland and Richard W. Katz 2016)を使用した。

参考文献

- [1] 川越清樹・風間聡・沢本正樹 (2008), 「数値地理情報と降雨極値データ を利用した土砂災害発生確率 モデルの構築」, 自然災害科学 Vol.27 (1), pp.69-83.
- [2] 渋谷政昭・華山宣胤 (2004), 「年齢時代区分データによる超高齢者寿命分布の推測」, 統計数理 第52巻 第1号, pp.117-134.
- [3] 志村隆彰 (2015), 「極値理論における離散化の影響」, 統計数理研究所オープンハウス.
- [4] 高橋倫也 (2009), 「極値統計学」, 情報論的学習理論テクニカルレポート.
- [5] 高橋倫也・志村隆彰 (2004), 「『特集 極値理論』について」, 統計数理 第52巻 第1号, pp.1-4.
- [6] 壇一男・神田順 (1986), 「上下限界を有する極値分布を用いた地震危険度解析」, 日本建築学会構造系論文報告集 第363号, pp.50-56.
- [7] 鳥越規夫 (2012), 「プロ野球の数理科学」, オペレーションズ・リサーチ Vol. 57 (1), pp.11-16.
- [8] de Haan, L. (1976). "Sample extremes: an elementary introduction", *Statistica Neerlandica Volume 30*, pp.161-172.
- [9] Einmahl, J., Magnus, J. R. (2006). "Records in athletics through extreme-value theory, " *Journal of the American Statistical Association* 103, pp.1382-1391.
- [10] my favorite giants, 「【NPB】歴代タイトル受賞者・首位打者」, https://www.my-favorite-giants.net/npb/title/batter_ba.html, (参照 2020-11-01)