

〈資料解説〉時系列データ中の最大値について

Kitagawa, Tetsuya / 北川, 徹哉

(出版者 / Publisher)

法政大学人間環境学会

(雑誌名 / Journal or Publication Title)

The Hosei Journal of Sustainability Studies / 人間環境論集

(巻 / Volume)

23

(号 / Number)

1

(開始ページ / Start Page)

75

(終了ページ / End Page)

86

(発行年 / Year)

2022-10-31

(URL)

<https://doi.org/10.15002/00026181>

時系列データ中の最大値について

北川 徹哉

1. はじめに

たとえば病気やケガで治療を受けると医療費が発生し、わが国の場合、通常は健康保険や共済組合などの保険制度が医療費を部分的にカバーすることで本人の支払額は軽減される。保険料を継続的に納めることの対価として、本人と家族の健康にかかわるリスクの一部を保険制度に引き受けてもらっているのである。また、例えばある企業の債務不履行リスクが上昇すると、その債権を所有する投資家は損失を被る可能性が高まる。そこで投資家は、あらかじめクレジット・デフォルト・スワップ（デリバティブの一種）を購入するなどにより、損失の補償を受け取れるように先手を打っておく。これらの例のように、何らかの負の要因による損害や損失を軽減あるいは相殺したり、そこから回復させることがリスク対策の主な役割である。

上記のような健康・人命や資本・金融だけでなく、公共システム、製造・運輸、販売・サービス、情報・通信、資源・鉱物、農水産物、各種の安全保障ほか、社会を構成するほとんどのファクターやプレイヤーがそれぞれ固有の負の要因とリスクをかかえている。これに対して多くの場合に共通の特性の一つは、負の要因の大きさは時間に関して変動し、それに応じて損害・損失の規模もダイナミックに変化することである。これをリスク対策に反映するための常套的なプロセスは、まず負の要因によって被りうる損害・損失の期待値を推算することであ

る。そして、その期待値に見合ったリスク対策の種類、規模、条件やコストなどを検討・決定し、運用となる。ところが、予期せずして負の要因があまりに強く発現し、期待値を大きく上回る損害・損失が生じてしまい、運用中のリスク対策では焼け石に水となることがある。これは警戒すべき局面であろう。であるならば、負の要因が最も大きく現れる場合をにらみ、それによる損害・損失を相殺できるほど万全なリスク対策を備えておけば憂いはない、と私たちは考える。また、例えばインフラやライフラインでは、いったん深刻なトラブルが生じると膨大な損害・損失に見舞われ、都市や地域は麻痺あるいは孤立するとともに2次的損害・損失が連鎖する。その甚大さゆえ、起こりうる損害・損失を期待値ベースで評価するのではなく、負の要因が最大の強さで襲来しても損害・損失を生じさせないような予防的・構造的な備えが必要となるケースもある。いずれにせよ、負の要因が最大となるときの方策を練るのであれば、まず必要なのは負の要因の最大値を知るための具体的な手立てである。

最大値を推定するための基本情報の一つはどの程度の長さの期間をターゲットとするかである。結論を先に言ってしまうと、対象とする期間が長いほど、その間に記録されたデータ中の最大値は大きくなる。この理由は簡単であり、しばしば引用される説明の一つに以下のようなものがある。一つのサイコロを所定の回数だけ振って出た目の値を順次記録し、その記録の中の最大値に着目する。まず、サイコロを振る回数を1回のみと設定すると、1回だけ振って出た目の値がそのまま最大値となる。すると、1～6の目が出る確率はそれぞれ等しいことから、1～6のどの目も同等の確率で最大値になりえる。次に、サイコロを振る回数を2回と設定する。1が最大値となるとき目の出方は一つのパターンのみである（1回目に1が出て2回目も1が出る）。しかし、2が最大値となるときは三つのパターンがある（1回目に1が出て2回目に2が出る、1回目に2が出て2回目に1が出る、1回目に2が出て2回目も2が出る）。同様に、3、4、5、6が最大値となるほどにパターン数は多くなる。パターン数が多いほど全パターン数に対して占める割合が多いので、相対的に起こりやすいことを意味する。したがって、大きい目ほど最大値になる確率は高く、小さい目ほど最大値になる確率は低い。上述したようにサイコロを振る回数が1回のみであれば、どの目も同

じ確率で最大値になりえたが、サイコロを振る回数を2回にただけで、大きい目ほど最大値になる確率が高くなった。この傾向はサイコロを振る回数をさらに多く設定するほど顕著になり、すなわち、サイコロを振る回数が多いほど最大値は大きくなるのである。ここで、回数は機会の時間進行と解釈することもできるから、期間と等価である。これを考慮すると、「回数が多いほど最大値は大きくなる」は「期間が長いほど最大値は大きくなる」と言い換えられる。

さらに、小さい目ほど高い確率で出て、大きい目ほど低い確率で出るように細工されたサイコロを思い浮かべてみる。説明は省略するが、前述の通常のサイコロだけでなく、このような偏ったサイコロであっても振る回数を増やしてゆくと最大値は大きくなるのである。ゆえに、発生確率に偏りがある事象においても、やはり期間が長いほど最大値は大きくなる。ここで、発生確率に偏りがある事象とはさぞかし特殊なケースであろうと思うかもしれないが、そうではない。むしろ偏っていることが普通である。例えば株価の変動（ボラティリティ）を見ると、小幅な変動はほぼ途切れることなく続いている一方、大きな変動は間欠的である。また、各種の工業製品ではちょっとした故障や動作不良はそれなりの頻度で発生しても、修理不可能となるほど大きく損壊することは稀である。あるいは、地盤のごく小規模な震動（常時微動）は今この瞬間も発生しているが、大災害を招くほどの巨大地震の発生頻度は非常に低い、など例をあげるときりがない。これらのように、小さなものほど高い確率で発生し、大きなものほど低い確率で発生することは社会、経済、自然などにおいてはありふれた光景である。したがって、私たちを取り巻く環境のほとんどの事象においては発生確率に偏りがあり、また、対象とする期間が長いほど最大値が大きくなることは必然なのである。

以上のことから、いくら長期間にわたる安全安心を確保したいからといって、むやみに長い期間を設定すると負の要因の最大値は異常に大きく見積もられ、これに真正面から対策を講じようとするれば、とても導入できないほど高コストなりリスク対策案ができあがってしまう。極端なケースとして、無限の長さの期間を想定すると、負の要因の最大値は無限大となり、これはどれほど莫大な資金をリスク対策に投入しても太刀打ちできない破滅的状况にいつかは直面することを示唆

する。企業経営に例えて言い換えると、いくら危機管理に予算を割いても、同じ経営方針のまま未来永劫に存続できる企業はありえないと解釈でき、感覚的にも妥当に思える。しかしながら、宇宙物理のように想像を絶する時空間スケールを取り扱う分野であればともかく、私たちが常日ごろ悪戦苦闘している社会という狭小で移ろいやすい世界での諸問題については、無限の期間を考えることはナンセンスである。実状に照らして適度な、あるいは物理的裏付けに基づいた有限の期間を設定するのであり、気象であれば例えば風速の平均化時間（10 分間）か、長期予報期間（1 箇月、3 箇月）か、ビジネスなら会計年度（1 年間）ごとか中期経営計画期間（3～5 年間）か、風力発電事業であれば風車の耐用年数（20 年間前後）か、ビルや橋梁などの管理運用なら供用期間（50～60 年間程度）か、などである。つまり、対象ごとに視野に入れるべき時間のスパンや周期、区切りなどがあり、そのタイムホライズンに発生するであろう負の要因の最大値を考えるのである。

ここまで見てきたように、最大値の大きさは確率的であり、かつ期間の長さに応じて変わる。これが意味するところは、私たちがデータ中に目撃する最大値は、所定の期間の最大値を構成する母集団から、たまたま生起したサンプル（標本、実現値）の一つにすぎない、ということである。サンプルである最大値の一つ一つについて事例的な検討を繰り返すことに意義はあっても、それだけでは「木を見て森を見ず」であり、リスク対策の設計に役立つ情報はなかなか出てこない。このためには、最大値を生起させる本体、すなわち「森」である母集団がどのような姿であるかを浮かび上がらせることが必要かつ合理的である。この種の課題を考究する分野は極値統計とよばれる。筆者はけっして極値統計に明るいわけではないが、ウェスタンオンタリオ大学の教授であった Alan Garnett Davenport 先生が示した最大値の評価方法はよく知られており、これを本稿にて紹介したい。筆者が若かった頃、ある国際会議で Davenport 先生にお会いしたときには聡明な初老の紳士という印象を持ったが、筆者にとってあまりに遥か雲の上の存在であった。Davenport 先生は 2009 年に逝去されたが、多くの先駆的な研究の成果を世に送り出し、それらは学術でも実務でも広く応用されている。量的にも質的にもこれほどの業績をたった一人の人間が、しかも若い頃から成せ

るものなのか、いったい頭の中はどうなっているのか、と驚きを通り越して畏敬の念を覚えたものである。次章以降では、Davenport 先生の業績の中のほんの一部である最大値評価理論について、原著論文¹⁾に筆者なりの補足や展開を加えながら、あるいは時には簡略化しながら概説し、最後に筆者による簡単な試行例を示す。

2. 最大値評価の理論

まず、時間を t 、時系列データを $x(t)$ と表し、 $x(t)$ の記録時間長（第 1 章で説明した期間の長さに相当する）を T とする。簡単のため、 $x(t)$ の平均値はゼロとする。そして、 $x(t)$ 中の極大値の発生がポアソン過程であると仮定すると、極大値が a を超える時間間隔が T 以下である確率 $F(a)$ は

$$\begin{aligned} F(a) &= \int_0^T N_a^+ \exp(-N_a^+ t) dt \\ &= 1 - \exp(-N_a^+ T) \end{aligned} \quad (1)$$

である。ここで、 N_a^+ は $x(t)$ の極大値が単位時間内に a のレベルを正方向に横切る回数の期待値であり、

$$N_a^+ = \int_0^\infty \dot{x} p(x, \dot{x}) d\dot{x} \Big|_{x=a} \quad (2)$$

と表される（例えば文献 2)）。 $\dot{x}$ は x の速度 dx/dt であり、 $p(x, \dot{x})$ は x と $\dot{x}$ の結合確率密度関数である。 x と $\dot{x}$ が互いに独立で、正規過程であると仮定すると、 $p(x, \dot{x})$ は

$$p(x, \dot{x}) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \frac{1}{\sqrt{2\pi}\sigma_{\dot{x}}} \exp\left(-\frac{\dot{x}^2}{2\sigma_{\dot{x}}^2}\right) \quad (3)$$

と表され、 σ_x と $\sigma_{\dot{x}}$ はそれぞれ x と $\dot{x}$ の標準偏差である。式 (3) を式 (2) に代入すると、

$$\begin{aligned}
N_a^+ &= \int_0^\infty \dot{x} p(x, \dot{x}) d\dot{x} \Big|_{x=a} \\
&= \frac{1}{2\pi\sigma_x\sigma_{\dot{x}}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \int_0^\infty \dot{x} \exp\left(-\frac{\dot{x}^2}{2\sigma_{\dot{x}}^2}\right) d\dot{x} \Big|_{x=a} \\
&= \frac{1}{2\pi\sigma_x\sigma_{\dot{x}}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \left[\frac{1}{2(-1/2\sigma_{\dot{x}}^2)} \exp\left(-\frac{\dot{x}^2}{2\sigma_{\dot{x}}^2}\right) \right]_0^\infty \Big|_{x=a} \\
&= \frac{1}{2\pi\sigma_x\sigma_{\dot{x}}} \exp\left(-\frac{x^2}{2\sigma_x^2}\right) \sigma_{\dot{x}}^2 \Big|_{x=a} \\
&= \frac{1}{2\pi} \frac{\sigma_{\dot{x}}}{\sigma_x} \exp\left(-\frac{a^2}{2\sigma_x^2}\right)
\end{aligned} \tag{4}$$

と得られる。 $a=0$ とすると、

$$N_0^+ = \frac{1}{2\pi} \frac{\sigma_{\dot{x}}}{\sigma_x} \tag{5}$$

であり、これはゼロクロッシング数を意味する。式(5)を用いると式(4)は以下のように書ける。

$$N_a^+ = N_0^+ \exp\left(-\frac{a^2}{2\sigma_x^2}\right) \tag{6}$$

次に a の確率密度関数を $f(a)$ とし、 $f(a)$ を

$$\xi = N_a^+ T = N_0^+ T \exp\left(-\frac{a^2}{2\sigma_x^2}\right) \tag{7}$$

のように変数変換したものを $\Psi(\xi)$ とすると、両者には

$$f(a) da = \Psi(\xi) d\xi \tag{8}$$

の関係がある。これと式(1)の $F(a)$ を用いると、

$$f(a) da = \frac{dF(a)}{da} da = \Psi(\xi) d\xi \quad (9)$$

であるから、

$$dF(a) = \Psi(\xi) d\xi \quad (10)$$

となる。これに式(1)、(7)を適用すると

$$\begin{aligned} \Psi(\xi) &= \frac{dF(a)}{d\xi} \\ &= \frac{d}{d\xi} (1 - \exp(-\xi)) \\ &= \exp(-\xi) \end{aligned} \quad (11)$$

のように $\Psi(\xi)$ が求まる。改めて式(11)を式(9)に代入すると

$$f(a) da = \exp(-\xi) d\xi \quad (12)$$

となる。さらに、 $f(a)$ の1次モーメントをとれば a の期待値が得られ、これはすなわち $x(t)$ の最大値 x_{max} の期待値 η_{max} を指す。ゆえに、式(12)より、

$$\begin{aligned} \eta_{max} &= \int_0^{\infty} a f(a) da \\ &= \int_0^{\infty} a \exp(-\xi) d\xi \end{aligned} \quad (13)$$

と書くことができる。式(7)より $a^2 = 2\sigma_x^2 \ln(N_0^+ T/\xi)$ であるから、これを式(13)に用いると

$$\begin{aligned}
\eta_{max} &= \int_0^{\infty} a \exp(-\xi) d\xi \\
&= \sigma_x \int_0^{\infty} \sqrt{2 \ln \left(\frac{N_0^+ T}{\xi} \right)} \exp(-\xi) d\xi
\end{aligned} \tag{14}$$

である。ここで、式(14)第2式右辺の平方根の部分に $\ln(\xi)=0$ 近傍でテーラー展開すると

$$\begin{aligned}
\sqrt{2 \ln \left(\frac{N_0^+ T}{\xi} \right)} &= \sqrt{2 \ln(N_0^+ T) - 2 \ln(\xi)} \\
&= \sqrt{2 \ln(N_0^+ T) \left(1 - \frac{\ln(\xi)}{\ln(N_0^+ T)} \right)} \\
&= \sqrt{2 \ln(N_0^+ T)} - \frac{\ln(\xi)}{\sqrt{2 \ln(N_0^+ T)}} - \frac{1}{2} \frac{\ln^2(\xi)}{\sqrt{(2 \ln(N_0^+ T))^3}} + \dots
\end{aligned} \tag{15}$$

となる。式(15)第3式において右辺第3項以下を微小項として省略し、改めて式(14)第2式に代入すると、

$$\begin{aligned}
\eta_{max} &= \sigma_x \int_0^{\infty} \sqrt{2 \ln \left(\frac{N_0^+ T}{\xi} \right)} \exp(-\xi) d\xi \\
&\cong \sigma_x \int_0^{\infty} \left(\sqrt{2 \ln(N_0^+ T)} - \frac{\ln(\xi)}{\sqrt{2 \ln(N_0^+ T)}} \right) \exp(-\xi) d\xi \\
&= \sigma_x \left(\sqrt{2 \ln(N_0^+ T)} \int_0^{\infty} \exp(-\xi) d\xi - \frac{1}{\sqrt{2 \ln(N_0^+ T)}} \int_0^{\infty} \ln(\xi) \exp(-\xi) d\xi \right) \\
&= \sigma_x \left(\sqrt{2 \ln(N_0^+ T)} + \frac{\gamma}{\sqrt{2 \ln(N_0^+ T)}} \right)
\end{aligned} \tag{16}$$

のように最大値の期待値が得られる。ここで、 γ はオイラー定数

$$\begin{aligned}
\gamma &= - \int_0^{\infty} \ln(\xi) \exp(-\xi) d\xi \\
&= 0.57721 \dots
\end{aligned} \tag{17}$$

である。式(16)の両辺を σ_x で除した η_{max}/σ_x はピークファクターとよばれ、各

種の観測データ中の最大値の議論によく用いられている（例えば文献3）。

一方、最大値 x_{max} の分散 σ_{max}^2 は $f(a)$ の2次モーメントであり、式(15)、(16)を用いて

$$\begin{aligned}
 \sigma_{max}^2 &= \int_0^\infty (a - \eta_{max})^2 f(a) da \\
 &\equiv \sigma_x^2 \int_0^\infty \left(\sqrt{2\ln(N_0^+ T)} - \frac{\ln(\xi)}{\sqrt{2\ln(N_0^+ T)}} - \sqrt{2\ln(N_0^+ T)} \right. \\
 &\quad \left. - \frac{\gamma}{\sqrt{2\ln(N_0^+ T)}} \right)^2 \exp(-\xi) d\xi \\
 &= \sigma_x^2 \frac{1}{2\ln(N_0^+ T)} \int_0^\infty (\ln(\xi) + \gamma)^2 \exp(-\xi) d\xi \\
 &= \sigma_x^2 \frac{1}{2\ln(N_0^+ T)} \int_0^\infty (\ln^2(\xi) + 2\ln(\xi)\gamma + \gamma^2) \exp(-\xi) d\xi
 \end{aligned} \tag{18}$$

である。式(17)ならびに

$$\int_0^\infty \ln^2(\xi) \exp(-\xi) d\xi = \frac{\pi^2}{6} + \gamma^2 \tag{19}$$

の関係を用いると式(18)は

$$\begin{aligned}
 \sigma_{max}^2 &= \sigma_x^2 \frac{1}{2\ln(N_0^+ T)} \left(\frac{\pi^2}{6} + \gamma^2 + 2\gamma(-\gamma) + \gamma^2 \right) \\
 &= \sigma_x^2 \frac{\pi^2}{6} \frac{1}{2\ln(N_0^+ T)}
 \end{aligned} \tag{20}$$

と求まり、最大値 x_{max} の標準偏差 σ_{max} は次式となる。

$$\sigma_{max} = \sigma_x \frac{\pi}{\sqrt{6}} \frac{1}{\sqrt{2\ln(N_0^+ T)}} \tag{21}$$

図1、2に、 $N_0^+ T$ を横軸として、式(16)、(21)の両辺を σ_x で除した η_{max}/σ_x 、 σ_{max}/σ_x をそれぞれ示す。 $N_0^+ T$ が大きくなるほど η_{max}/σ_x は増大する一方、

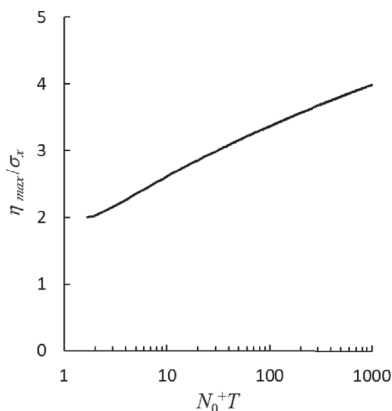


図 1 η_{max}/σ_x の概要

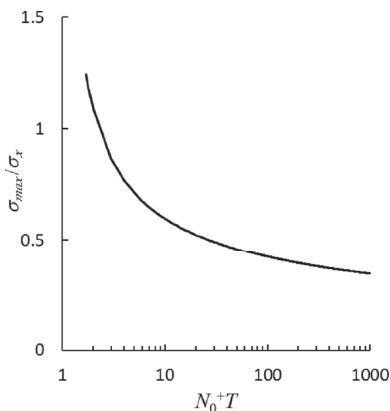


図 2 σ_{max}/σ_x の概要

σ_{max}/σ_x は小さくなる傾向がある。なお、図 1 には描いていないが、 N_0^+T がより一層小さい値になると η_{max}/σ_x は反転して急激に増大する。ただし実際は、統計的に有意となるほどの多くの波数が時系列に含まれる必要があり、必然的に時系列に含まれる各波長に比して T を十分長くすることが求められるため、 N_0^+T はそれほど小さな値にはならない。

3. 最大値評価式の適用例

前章に示した理論展開においては、対象とする時系列データが正規過程であることを前提としていた。それゆえ、導出された最大値の期待値 η_{max} (式(16)) と標準偏差 σ_{max} (式(21)) は時系列 $x(t)$ が正規過程の場合のみ適用可能である。本稿においては、気象庁が観測・公開⁴⁾している日最大風速の時系列データを Box-Cox 変換⁵⁾によって正規過程に近づけ、平均値がゼロで標準偏差が1となるように標準化し、さらにドリフト成分を除去したものを $x(t)$ とする。日最大風速のデータが元であるから時間 t の単位は日であり、また、年間のデータを対象とすることから記録時間長 T は 365 日である。

例えば、図 3 は 1994 年に東京管区気象台において観測された日最大風速から作成した $x(t)$ である。この $x(t)$ を対象に式(16)、(21)を用いて、 η_{max} 、 σ_{max} を

求めた結果を図4に示す。比較のために図3の $x(t)$ もあわせて示しているが、わかりやすくするために $x(t)$ は正の値のみを描いた。また、太い実線が η_{max} であり、その上方の破線が $\eta_{max}+\sigma_{max}$ 、下方が $\eta_{max}-\sigma_{max}$ を表す。 $t=273$ 日において最大値 x_{max} がみられ、その値は $\eta_{max} \pm \sigma_{max}$ の範囲に収まっており、なおかつ η_{max} に近い。したがって、Davenport 先生が導出した式(16)、(21)はこの時系列の最大値を良い精度で推定できている。同様に、同年に淡路島の洲本特別地

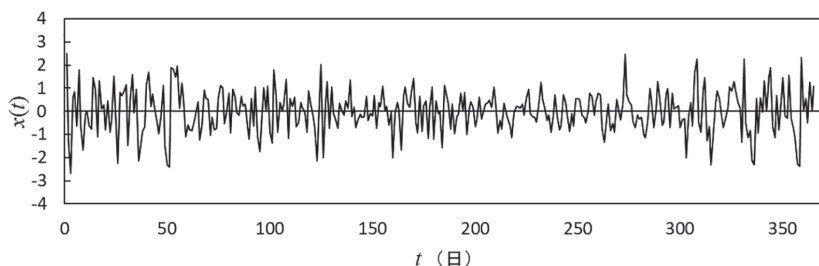


図3 1994年に東京管区気象台において観測された日最大風速から作成した $x(t)$

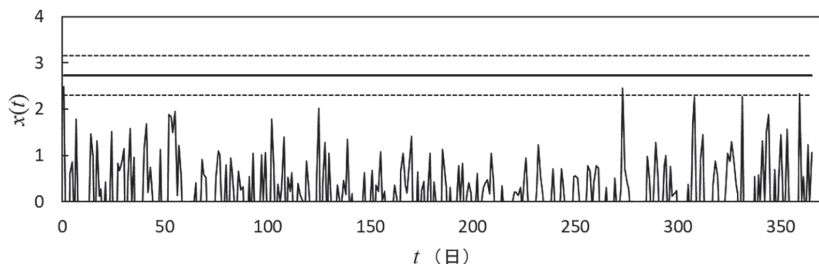


図4 図3の $x(t)$ (正の値のみ) と η_{max} (太い実線) および $\eta_{max} \pm \sigma_{max}$ (破線) の比較

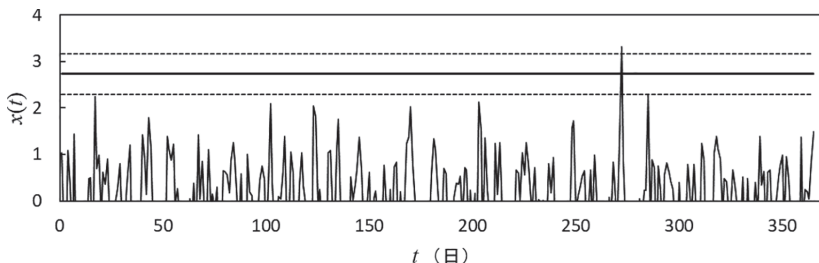


図5 1994年に洲本特別地域気象観測所において観測された日最大風速から作成した $x(t)$ (正の値のみ) と η_{max} (太い実線) および $\eta_{max} \pm \sigma_{max}$ (破線) の比較

域気象観測所において観測された日最大風速の時系列を対象に $x(t)$ を作成し、 η_{max} と σ_{max} を求めた。これらについて、図 4 と同様に、 $x(t)$ と比較したものが図 5 である。 x_{max} は $t=272$ 日に発生しており、図 4 と比べると、より大きな値となっている。図 5 の x_{max} はおよそ $\eta_{max} + \sigma_{max}$ に等しく、やはり式 (16)、(21) で最大値を推定できている。なお、1994 年の $t=272$ 日前後は台風 26 号が襲来したときであり、東京管区気象台での日最大風速は 11m/s、洲本特別地域気象観測所では 15.3m/s を観測した。図 4 の $t=273$ 日および図 5 の $t=272$ 日の各 x_{max} はそれらによるものである。

4. おわりに

確率統計や応用数学を用いているが、導出された最大値の期待値（式 (16)）とその標準偏差（式 (21)）は非常にシンプルな形にまとまっている。シンプルであることは、研究者であれ実務者であれ、ユーザーにとって大きなメリットであるだけでなく、各種の指針や設計にも取り入れやすい。シンプルだからこそ便利であり、広く受容され、利用されるのである。たとえ研究のプロセスは複雑になっても、回り道をしても、たどり着く先ではシンプルでありたいと筆者も常々思っているものの反省の連続である。なお、引用した Davenport 先生の論文には、本稿 2 章で概説した部分の後に続きがあり、空港における最大瞬間風速の特性（ガストファクター）への応用例が掲載されていることを付記しておく。

参考文献

- 1) Davenport, A. G. : Note on the distribution of the largest value of a random function with application to gust loading, *Proceedings of the Institution of Civil Engineers*, Vol. 28, Issue 2, pp. 187-196, 1964.
- 2) 中村大輔、南原英生：振幅情報に基づく任意不規則信号の簡易的ピーク値分布評価法、情報処理学会論文誌、Vol. 49, No. 1, pp. 375-380, 2008.
- 3) 岡野創、日下彰宏、桜庭潮理：塑性率と累積塑性率から推定される線形累積損傷度の期待値、日本建築学会構造系論文集、第 83 巻、第 747 号、pp. 669-676, 2018.
- 4) 気象庁： <http://www.jma.go.jp/jma/index.html>, 2021.
- 5) Box, G. E. P. and Cox, D. R. : An analysis of transformations, *Journal of Royal Statistical Society, Series B*, Vol. 26, pp. 211-252, 1964.