

Kostbare Zeit: Analyse und Modellierung von Bearbeitungszeiten

Bernard Rummel
SAP User Experience
SAP SE
bernard.rummel@sap.com

ABSTRACT

Bearbeitungszeit ist die bei weitem beliebteste Metrik zur Erfassung von Effizienz als Komponente der Gebrauchstauglichkeit [1,2,3]. Statistische Besonderheiten von Zeitverteilungen sind jedoch unter UX-Praktikern noch wenig bekannt. In diesem Beitrag wird eine Anleitung gegeben, wie man durch geeignete Datenvisualisierung zufällige von nicht-zufälligen Einflüssen auf Bearbeitungszeiten unterscheiden und geeignete Modelle für quantitative Aussagen auswählen kann.

CCS CONCEPTS

Human-centered computing → HCI design and evaluation methods

KEYWORDS

Gebrauchstauglichkeit, Effizienz, Bearbeitungszeit, Modellierung

1 Einführung

Zeitdaten folgen in der Regel einem typischen Muster, der Exponentialverteilung. Diese Verteilung tritt gewöhnlich bei Zufallsprozessen auf, die mit einem beobachtbaren Ereignis enden (time-to-event). Bekannte Beispiele sind radioaktiver Kernzerfall, Wartezeiten in Restaurants, oder die Zeit, bis man beim Mensch-Ärgere-Dich-Nicht eine 6 würfelt. Da die Exponentialverteilung typisch ist für Zufallsprozesse, ist sie ein Standardmodell zur Analyse von Zeitdaten, in der Bedeutung etwa vergleichbar zur Normalverteilung in anderen Bereichen. Abweichungen von diesem Modell hingegen deuten auf Einflüsse hin, die nicht zufällig sind – und daher besonders interessant.

Ein bekanntes Beispiel für eine Exponentialverteilung ist die aus dem Physikunterricht bekannte Zerfallskurve einer radioaktiven Substanz (Abb. 1a). Wann ein einzelnes Atom zerfällt, ist völlig zufällig und unvorhersagbar. Wenn man dagegen sehr viele Atome betrachtet, ergibt sich die bekannte Kurve. Der Anteil einer

radioaktiven Substanz, der innerhalb einer bestimmten Zeit zerfällt, ist konstant; ebenso ist die Zeit konstant, zu der ein bestimmter Anteil zerfällt – man spricht z.B. von der Halbwertszeit. Aus Modellierungssicht ist an der Exponentialverteilung interessant, dass sie sich mit einem einzigen Parameter vollständig beschreiben lässt. Ihr Mittelwert ist gleich der Standardabweichung; beide sind gleich der sogenannten „charakteristischen Zeit“, zu der nur noch $1/e$ der Substanz verbleibt.

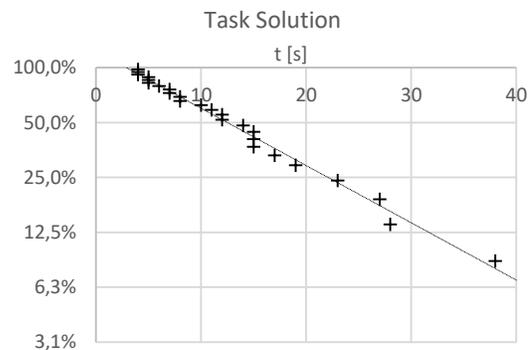
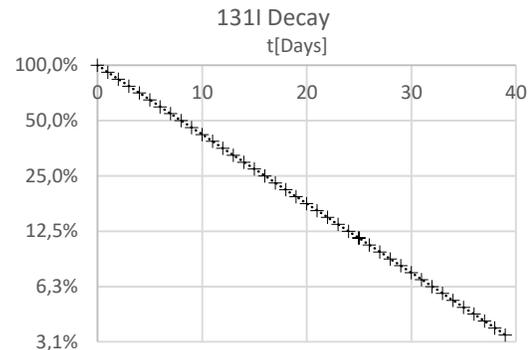


Abb. 1 a,b: Radioaktiver Zerfall des Isotops ^{131}I vs. Lösungsrate einer Usability-Testaufgabe. Abb. aus [4]

Trägt man die verbleibende Masse der Substanz wie in Abb. 1a logarithmisch ab, bilden die Messpunkte über die Zeit eine Gerade, da die Zerfallsrate konstant ist. Der Kehrwert ihrer Steigung ist gleich der charakteristischen Zeit.

Visualisiert man Bearbeitungszeiten zu einer Aufgabe in einem Usability Test auf die gleiche Weise, ist die Ähnlichkeit oft verblüffend. Wie beim radioaktiven Zerfall folgt der Anteil der Testteilnehmenden, der zu einer bestimmten Zeit noch an der Aufgabe arbeitet, in Abb. 1b der typischen Gerade.

2 Verschiebungskonstante vs. Halbwertszeit: „Click Time“ und „Think Time“

So frappierend die Ähnlichkeit zwischen Abb. 1a und b ist, so gibt es doch einen wichtigen Unterschied. Die Lösungsgerade in Abb. 1b beginnt nicht im Ursprung, sondern bei einer Zeit von etwa 3 Sekunden. Während im Kern ein Prozess mit einer konstanten „Zerfallsrate“ oder Halbwertszeit vorzuliegen scheint, sieht es so aus, als ob allen Zeiten eine Konstante von 3s hinzuaddiert worden wäre.

Wahrscheinlich ist genau das der Fall. Jedes interaktive System hat eine gewisse Reaktionszeit. Auch Benutzende brauchen eine endliche Zeit, um den Lösungspfad der Aufgabe rein mechanisch durchzuklicken. Benutzende können die Aufgabe nicht schneller abarbeiten als in der Summe dieser Zeiten, und weil alle den Lösungspfad durchlaufen müssen, ist es sinnvoll, einen konstanten Verschiebungs-Parameter in das Verteilungsmodell mit aufzunehmen.

Es ist plausibel anzunehmen, dass die Verschiebungskonstante die technische Performance und die reine Click-Effizienz der getesteten Benutzungsschnittstelle abbildet, die vom „Zufallsprozess“ der eigentlichen Aufgabenlösung unabhängig ist. Dieser Befund ist hochinteressant, erlaubt er doch die Trennung von Systemperformance und mechanischer Effizienz („click time“) von solchen Zeitbeiträgen, die auf das Konto von Zufallskomponenten wie Denk-, Such- und Fehlerbehebungsprozessen gehen („think time“). Während erstere technisch interessant sind, beschreiben letztere die kognitive Effizienz der Benutzungsschnittstelle. Erstere ist von der technischen Implementierung abhängig, letztere nicht und kann damit plausibel z.B. aus Tests mit Prototypen abgeleitet werden. Besonders interessant ist hier, dass die „Think Time“ ein starker Prädiktor für die Zufriedenstellung der Teilnehmenden ist [5], was nicht überrascht, weil sie ja zu einem guten Teil mit Suchprozessen sowie dem Begehen und Korrigieren von Benutzungsfehlern verbracht wird.

In der Praxis werden bisher aus Usability-Studien noch selten Verteilungsmodelle und -parameter berichtet, sondern in der Regel Minimal- und Maximalwerte, Mittelwerte und Standardabweichungen. Die Minimalzeit ist in der Regel ein guter Anhaltswert für die Verschiebungskonstante (Click Time). Die Think Time lässt sich schätzen: in einer reinen, nicht verschobenen Exponentialverteilung wären Mittelwert und Standardabweichung identisch, und zwar gleich der Charakteristischen Zeit. Zieht man also die Minimalzeit von der mittleren Bearbeitungszeit ab, und das Resultat entspricht in etwa der Standardabweichung, hat man einen guten Näherungswert für die Charakteristische Zeit. Diese entspricht der „Think Time“, die 63% der Teilnehmer für die Lösung der Aufgabe benötigten, zusätzlich zu und unabhängig von der „Click Time“.

3 Praktische Theorie, aber stimmt sie?

In der Praxis wird man Zeitdaten zunächst prüfen müssen, ob sie dem beschriebenen Muster der verschobenen Exponentialverteilung folgen.

Eine genauere Untersuchung der Zeitverteilung ist mit einem sog. Probability Plot [6] wie in Abb. 1b möglich. Hierzu werden die beobachteten Zeiten (x-Achse) gegen ihre jeweiligen Perzentilwerte auf einer logarithmischen Skala (y-Achse) geplottet (zur Bestimmung der Perzentilwerte später mehr). In etwa 80% der Fälle zeigt sich ein etwa geradliniger Verlauf der Datenpunkte, aus dem sich Click Time und Think Time leicht bestimmen lassen. Die Click Time ist der x-Achsenabschnitt, die Think Time der Kehrwert der Steigung der Regressionsgeraden.

Passt dieses Modell (zum Modelltest siehe [7]), beschreiben die beiden Parameter die Verteilung vollständig, d.h. es lässt sich für jede beliebige Zeit die entsprechende Erfolgsquote berechnen, sowie für jede Erfolgsquote die dafür erforderliche Zeit.

Ist der Verlauf der Datenpunkte im Plot nicht geradlinig, kann das mehrere Ursachen haben, von denen wir einige im Folgenden systematisch durchgehen werden. Erfahrungsgemäß lohnt sich eine Inspektion der Abweichungen unbedingt, da sie auf Probleme in der Datenqualität und andere, systematische Einflüsse auf die Bearbeitungszeit hindeuten, und letztlich erst die Auswahl eines geeigneten quantitativen Modells ermöglichen.

3.1 Schummler

In unmoderierten Online-Studien ist ein gängiges Problem, dass einige Teilnehmer Aufgaben nicht wirklich bearbeiten, sondern sich möglichst schnell durch die Aufgaben klicken, um die ausgelobte Vergütung zu erhalten. Viele Tools erlauben die Festlegung von Minimal-Bearbeitungszeiten, doch wie soll man diese ermitteln?

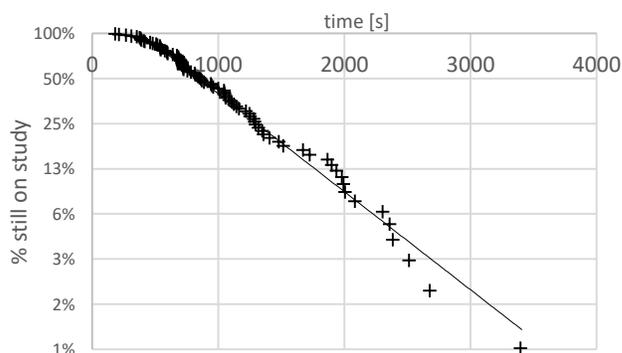


Abbildung 2: Bearbeitungszeit einer unmoderierten Usability-Studie. Beachte Abweichung der Datenpunkte von der Regressionsgeraden oben links. Abb. aus [4]

Ein Probability Plot ist eine elegante Methode, sowohl die Minimalzeit zu bestimmen – die „Click Time“ – als auch auffällige Teilnehmer zu identifizieren. Der typische lineare Verlauf deutet darauf hin, dass es ein kohärentes Modell gibt, das den Lösungsprozess im Zeitverlauf beschreibt. Abweichungen von diesem Modell

deuten darauf hin, dass die betreffenden Personen an einem anderen Prozess als der Aufgabebearbeitung teilgenommen haben - z.B. Schummeln.

Abbildung 2 zeigt Bearbeitungszeiten einer unmoderierten Usability-Studie (die Methode kann auf einzelne Aufgaben wie auch die gesamte Studiendauer angewendet werden). Im oberen Bereich des Plots, nahe der Zeitachse, sind 5 Datenpunkte zu sehen, die deutlich von der Regressionslinie der „Mainstream“-Datenpunkte abweichen, und zwar hin zu kürzeren Bearbeitungszeiten. Alle fünf haben sich bei näherer Inspektion als Schummler herausgestellt, die die Aufgaben nicht wie erwartet bearbeitet hatten. Kann man solche Personen noch zur Laufzeit der Studie identifizieren, werden sie oft von der Rekrutierungsagentur kostenfrei durch nachrekrutierte Teilnehmer ersetzt.

3.2 “Knick” im Plot

In unmoderierten Studien und allgemein unbeobachteten Arbeitsprozessen ist es nicht ungewöhnlich, dass Teilnehmende nach einiger Zeit unterbrochen werden, eine Pause einlegen oder anderweitig ihre Arbeitsweise ändern. Dies äußert sich im Probability Plot oft dahingehend, dass eine ansonsten geradlinige Verlaufskurve abknickt und einen anderen Verlauf nimmt. Abb. 2 zeigt einen solchen Knick nach etwa 1500s (25 min), wie er oft in Online-Studien vorkommt, wenn die Studie länger dauert als von Teilnehmenden erwartet, und sie sich in ihrer Zeitplanung neu orientieren müssen.

3.3 Muster

Wenn Abweichungen vom erwarteten Verlauf ihrerseits in Mustern auftreten, ist dies besonders interessant, da hier offenbar zwar eine Regelmäßigkeit vorliegt, doch eine andere als die erwartete.

In einer Studie zu Ladeanimationen sollten Teilnehmende bestimmte Suchziele auf Bildschirmbildern anklicken, die mit verschiedenen Animationen geladen wurden. Die Hypothese war, dass sog. Skeleton Screens - Animationen, die die Struktur des ladenden Bildschirmbildes andeuten - die Suchgeschwindigkeit verbessern.

Die Ergebnisse der Studie waren dahingehend auffällig, dass auffällig viele Datenpunkte im Probability Plot nicht dem linearen Muster folgten, sondern in auffälligen Gruppen auftraten. Ihre Anzahl reichte für eine genauere Ursachenbestimmung jedoch nicht aus.

In einer Folgestudie, unter Verwendung eines Eyetrackers, konnten wir das Datenmuster replizieren und seine Ursache feststellen. Abb.3 zeigt Daten aus zwei Versuchsbedingungen in einem kombinierten Probability Plot. Datenpunkte erscheinen in drei gut erkennbaren Gruppen, wobei einzelne Gruppen scheinbare Ausreißer-Datenpunkte aus beiden Versuchsbedingungen enthalten. Die Gruppen liegen etwa 1500ms auseinander, mit jeweils etwa paralleler Orientierung der Datenpunkte. Würde man für jede Versuchsbedingung eine Regressionsgerade durch die entsprechenden Punkte legen, wäre das mit Sicherheit kein angemessenes Modell der beobachteten Daten.

Eine Inspektion der Eyetracking-Daten der “Ausreißer“-Datenpunkte brachte die Erklärung. Teilnehmende begannen die Suchaufgabe, indem sie das Bildschirmbild nach dem Zielreiz absuchten. Fanden sie diesen nicht sofort, begannen sie von vorn. Die parallele Orientierung der Gruppen von Datenpunkten deutet darauf hin, dass der Suchprozess an sich ähnlich war – die Gruppen zeigen lediglich den ersten, zweiten und dritten Durchgang der Suche, mit jeweils ähnlicher Verschiebung um eine konstante Zeit. Die „bessere“ Versuchsbedingung war also nicht unbedingt effizienter im Sinne höherer Suchgeschwindigkeit. Sie war aber effektiver im Hinblick auf das wesentlich häufigere Erkennen des Zielreizes schon im ersten Durchgang.

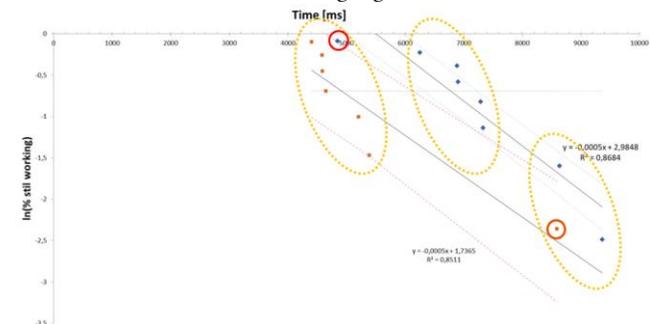


Abb.3: Probability Plots von zwei Versuchsbedingungen in einer visuellen Suchaufgabe mit verschiedenen Ladeanimationen. Beachte die Gruppierung der Datenpunkte, wobei Gruppen Datenpunkte beider Versuchsbedingungen enthalten. Regressionsgeraden mit 90% Vertrauensbändern. Abb. aus [4].

Aus praktischer Sicht ist hier interessant, dass die Folgestudie erst aufgrund der Inspektion der Probability Plots der Erststudie aufgesetzt wurde, während die Plots der Folgestudie die Analyse der Eyetracking-Daten maßgeblich unterstützten, indem sie die kritischen Datenpunkte auswiesen. In jedem Fall zeigten die Plots, dass ein Vergleich der Versuchsbedingungen mit klassischen quantitativen Verfahren irreführend gewesen wäre.

3.4 Kurvilineare Plots: Ausreißer oder stetiger Einfluss?

Nicht selten brauchen einzelne Testteilnehmende erheblich länger zur Lösung einer Aufgabe als andere. Für eine Exponentialverteilung ist das nicht überraschend. Angenommen, eine Aufgabe mit ausschließlich zufälligen Einflüssen auf die Bearbeitungszeit hat eine „Halbwertszeit“ t , zu der 50% der Teilnehmenden die Aufgabe lösen. Zur doppelten Zeit $2t$ arbeiten noch 25%, nach $3t$ 12,5%. Es ist zu erwarten, dass 6,3% der Teilnehmenden, also in einer Stichprobe von 20 einer, viermal so lange braucht wie die mittlere Bearbeitungszeit! Die übliche Definition eines „Ausreißers“ als „mehr als 3 Standardabweichungen vom Mittelwert entfernt“ wäre hier völlig irreführend; sie würde pauschal jeden: 20. Teilnehmende:n zum Ausreißer erklären.

Es kommt durchaus vor, dass Testteilnehmende noch länger für die Lösung einer Aufgabe brauchen, als nach der Exponentialver-

teilung zu erwarten wäre. Sie können durch viele Gebrauchstauglichkeitsprobleme frustriert, gelangweilt oder schlicht ermüdet sein. Einflüsse auf die Bearbeitungszeit können punktuell eintreten, oder aber sich systematisch über die Zeit aufaddieren. Sehr langsam arbeitende Testteilnehmende können daher durchaus Ausreißer sein, oder aber Betroffene eines systematischen, „bremsenden“ Einflusses.

Abb. 4a zeigt einen Probability Plot, in dem drei Teilnehmende deutlich länger zur Lösung der Aufgabe benötigen, als selbst nach einer Exponentialverteilung zu erwarten wäre. Da die übrigen Datenpunkte linear angeordnet sind, gibt es zwei mögliche Erklärungen. Eine ist, dass es sich tatsächlich um Ausreißer handelt, d.h. nicht valide Datenpunkte. Entfernt man sie, erhält man eine plausible und gut parametrisierbare Exponentialverteilung.

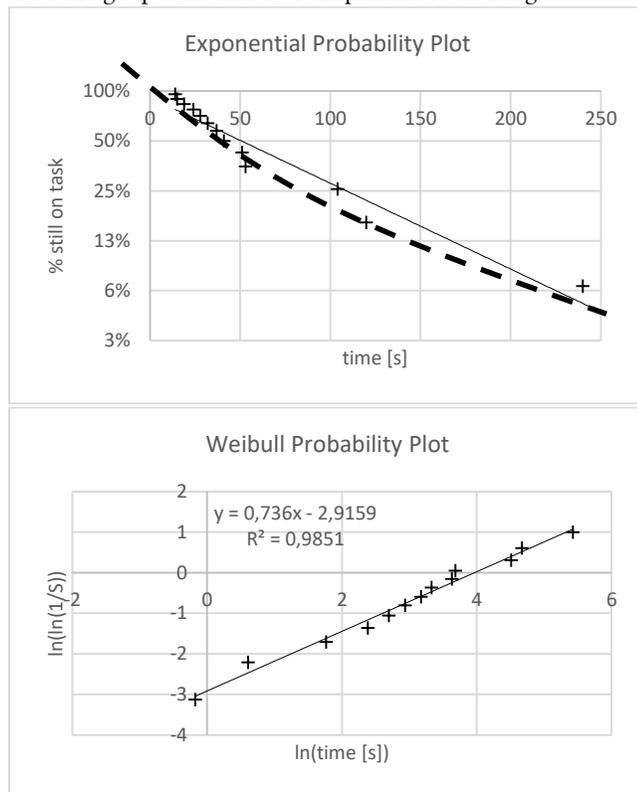


Abb. 4 a,b: Exponential- (a) und Weibull- (b) Probability Plots der gleichen Bearbeitungszeiten zu einer Aufgabe aus einer Gebrauchstauglichkeitsstudie. Abb. aus [4].

Eine alternative Hypothese wird in Abb. 4b untersucht. Dieser Probability Plot ist für eine Weibull-Verteilung konfiguriert, eine einfache Erweiterung des Exponentialverteilungsmodells [8]. Die Weibull-Verteilung enthält einen „Form“-Parameter, der einen systematischen, stetigen Einfluss auf die Bearbeitungszeit modelliert. Ist der Form-Parameter >1 , verkürzt sich die Bearbeitungszeit gegenüber dem Exponentialmodell, ist er <1 , verlängert sie sich (z.B. durch Ermüdung oder Motivationsverlust der Teilnehmenden). Der Probability Plot für die Weibullverteilung zeigt die

Zeit auf einer logarithmischen, die Perzentile auf einer doppelt logarithmischen Skala. Wenn die Datenpunkte linear angeordnet sind – was bei Abb. 4b der Fall ist – indiziert dies das Vorliegen einer Weibull-Verteilung, was wiederum nahelegt, dass es einen systematischen Einfluss auf das Bearbeitungsgeschehen gibt.

Für die Modellierung der Bearbeitungszeiten bedeutet dies, dass eine Entscheidung getroffen werden muss. Entweder sind die drei auffälligen Datenpunkte Ausreißer und zu verwerfen, oder die gesamte Verteilung folgt dem Weibull- und nicht dem Exponentialverteilungsmodell. Für die Analyse bedeutet dies, dass die Rohdaten erneut inspiziert werden müssen, aber unter verschiedenen und jeweils klaren Perspektiven. Finden sich bei den drei auffälligen Datenpunkten Indizien, die sie als nicht valide Datenpunkte identifizieren, stützt dies die Ausreißer-Hypothese. Finden sich andererseits im Bearbeitungsgeschehen Hinweise auf einen stetigen, leistungsmindernden Einfluss, stützt dies die Hypothese einer Weibull-Verteilung. Für die Modellierung müssen dann die entsprechenden Parameter bestimmt werden (für eine Anleitung s. [8]).

4 Erstellung der Probability Plots

Probability Plots sind im Wesentlichen Streudiagramme von Zeiten gegen die geschätzten Perzentile der jeweiligen Datenpunkte in der Gesamtpopulation der Testteilnehmenden. [4] bietet ein Excel-Spreadsheet zum Download an, mit dem die Berechnungen ausgeführt werden können, sowie eine praxisorientierte Anleitung. [7] gibt eine ausführliche Einführung.

In der Zerfallskurve in Abb. 1a zeigt die vertikale Achse den verbleibenden Anteil der Substanz an. Analog zeigt Abb. 1b den Anteil „in der Aufgabe verbleibender“, d.h. noch daran arbeitender Teilnehmender.

Dieser Anteil ist besonders bei kleinen Stichproben nicht trivial zu schätzen. Angenommen, man hätte nur 2 Teilnehmende, welches Perzentil würde jeder von ihnen repräsentieren? Wenn einer die Aufgabe nicht löst, was bedeutet dies für die Schätzung? Glücklicherweise ist die Schätzung solcher Perzentile in der statistischen Zuverlässigkeitsanalyse gut untersucht und verstanden. Eine bewährte Schätzmethode ist der sog. „modifizierte Kaplan-Meier-Schätzer“ (mKM, [6,9]), der die Schätzung der Perzentile auch bei unvollständigen Erfolgsquoten erlaubt. [7] gibt eine ausführliche Anleitung zur Berechnung im Fall von Bearbeitungszeiten, daher wird hier nur eine kurze Übersicht gegeben.

Der mKM-Algorithmus verwendet Rangdaten zur Ermittlung der Perzentile. Die Bearbeitungszeiten werden einfach der Größe nach sortiert, und die Rangziffern werden zur Perzentilberechnung herangezogen. Im Fall unvollständiger Erfolgsquoten sind weitere Aspekte zu beachten. Dafür ist wesentlich, ob Teilnehmende aus zufälligem Grund an der Aufgabe scheiterten (z.B. durch eine Unterbrechung), oder aber aus inhaltlichen Gründen, d.h. aufgaben, eine falsche Lösung nannten oder am Zeitlimit scheiterten.

Im ersteren Fall werden die Zeitpunkte, an denen das Scheitern beobachtet wurde, einfach „einsortiert“, die Zeiten aber nicht geplottet. Im letzteren Fall können wir annehmen, dass die Teilnehmenden sehr lange für die richtige Lösung gebraucht hätten. Da

auch hier die „wahren“ Zeiten nicht geplottet werden, interessieren sie nicht – sie werden lediglich am Ende einsortiert.

Es ist leicht zu sehen, warum das in der Praxis funktioniert. Angenommen, die Lösungsquote liegt bei 80%. Die Perzentile der erfolgreichen Teilnehmenden müssen hier nicht auf 100, sondern nur 80% aufgeteilt werden. Zufällig scheiternde Teilnehmende dagegen hätten die Aufgabe vielleicht durchaus lösen können, müssen also innerhalb der Gruppe der erfolgreichen Teilnehmenden einsortiert werden.

5 Analyseschritte

In der Praxis benötigt man die beobachteten Zeiten sowie eine 0/1-Kennung, ob die Aufgabe gelöst wurde oder nicht. Eine Teilnehmenden-ID hilft, auffällige Datenpunkte nachzuverfolgen, z.B. um Ausreißer-Kandidaten zu untersuchen.

Die Berechnung der Perzentile nach dem mKM-Algorithmus und die Erstellung der Plots lässt sich in einem Tabellenkalkulationsprogramm vorbereiten [zum Download in 4,7,8]. Das Einkopieren und Sortieren ist dann eine Sache von Sekunden.

In der Regel wird man zunächst den exponentiellen Probability Plot auf Auffälligkeiten inspizieren und die entsprechenden Datenpunkte genauer untersuchen. Falls die Datenpunkte hier bereits eine gerade Linie ergeben, ist die Parametrisierung durch „click time“ und „think time“ ausreichend.

Falls die Datenpunkte im Plot einen kurvilinearen Verlauf nehmen, werden die Daten einfach in das Kalkulationsblatt für die Weibull-Verteilung eingekopiert. Da die „Click time“ in diesem Plot nicht automatisch ermittelt werden kann, muss sie manuell eingegeben werden. Die Bestimmung aus dem Exponentialplot ist in der Regel jedoch ohne weiteres möglich. [8] beschreibt eine Methode, diesen Parameter so zu berechnen, dass die Gesamt-Modellpassung optimiert wird.

Diese Modelle werden in etwa 90% der Fälle ausreichen [8]. In den verbleibenden Fällen können andere Verteilungsmodelle in Betracht gezogen werden [7] – falls eine quantitative Modellierung überhaupt sinnvoll ist (vgl. Abb.3, wo sie es nicht ist).

5 Zusammenfassung und weiterführende Hinweise

Die Visualisierung von Bearbeitungszeiten in Probability Plots ist ein effektives Werkzeug, um Verteilungsmodelle zu identifizieren [7]. Im Fall von Bearbeitungszeiten in Gebrauchstauglichkeitstests erlaubt besonders der Exponentialverteilungs-Plot eine schnelle Identifikation von Ausreißern und anderen Auffälligkeiten, mit der die Sinnhaftigkeit quantitativer Modellierung schnell beurteilt werden kann [4]. Die Exponentialverteilung und als Erweiterung die Weibull-Verteilung decken einen großen Anwendungsbereich ab und sind in Bezug auf Gebrauchstauglichkeitsprobleme leicht zu interpretieren [8].

Erweiterungen dieser Methodik sind aus Sicht des Autors in zweierlei Richtung sinnvoll. Da ihre Wurzeln in der technischen Zuverlässigkeits- und medizinischen Survival-Analyse liegen, ist die Auseinandersetzung mit entsprechender Literatur ausgesprochen

lohnend (für eine Einführung s. [9]), zumal hier gut validierte Modellierungsmethoden auch für komplexe Sachverhalte zur Verfügung stehen. Zum anderen sind Zeitverteilungen und die Prozesse, die sie generieren, nicht auf die Familie der Exponentialverteilungen beschränkt (s. z.B. [10]). Die Absicht des vorliegenden Papers ist, Interesse für diese Literatur zu wecken und einen praxisorientierten Einstieg zu bieten.

LITERATUR

- [1] Coursaris, C., & Kim, D. (2011). A meta-analytical review of empirical mobile usability. *Journal of Usability Studies*, 6(3), 117–171.
- [2] Sauro, J., & Lewis, J. R. (2009). Correlations among prototypical usability metrics: Evidence for the construct of usability. *Proc. CHI 2009*, ACM Press.
- [3] Sauro, J., & Lewis, J. R. (2010). Average task times in usability tests: What to report? *CHI '10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2347–2350). New York, NY: ACM Press.
- [4] Rummel, B. (2020). About Time: A Practitioner's Guide to Task Completion Time Analysis. *Journal of Usability Studies* 15(3), 124-134.
- [5] Rummel, B. (2017b). Predicting Post-Task User Satisfaction With Weibull Analysis of Task Completion Times. *Journal of Usability Studies* 13(1), 5–16.
- [6] NIST/SEMATECH (2012). Probability Plotting. In: *E-handbook of statistical methods*. National Institute of Standards and Technology. <https://www.itl.nist.gov/div898/handbook/apr/section2/apr221.htm>.
- [7] Rummel, B. (2014). Probability plotting: A tool for analyzing task completion times. *Journal of Usability Studies* 9(4), 152–172.
- [8] Rummel, B. (2017a). Beyond average: Weibull analysis of task completion times. *Journal of Usability Studies* 12(2), 56–72.
- [9] Tobias, P. A., & Trindade, D. C. (2012). *Applied reliability* (3rd ed.). Boca Raton, FL: CRC Press.
- [10] Lindeløv, Jonas K. (2019). Reaction time distributions: an interactive overview. <https://lindelov.shinyapps.io/shiny-rt/>