Zeitschrift: Mitteilungen / Vereinigung Schweizerischer Versicherungsmathematiker = Bulletin / Association des Actuaires Suisses = Bulletin / Association of Swiss Actuaries

Herausgeber: Vereinigung Schweizerischer Versicherungsmathematiker

Band: 65 (1965)

Artikel: Neuere Fortschritte in der Theorie der Bayesschen Statistik

Autor: Huber, Peter J.

DOI: https://doi.org/10.5169/seals-551083

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 09.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Neuere Fortschritte in der Theorie der Bayesschen Statistik

Von Peter J. Huber, Zürich

Zusammenfassung

Das vorliegende Referat führt in wichtige neuere Fortschritte auf dem Gebiet der Bayesschen Statistik ein, insbesondere in die Arbeiten von D. Freedman über das asymptotische Verhalten von Bayes-Schätzungen und in die empirischen Bayes-Methoden von H. Robbins. Die versicherungsmathematische Relevanz dieser Resultate wird anhand des Tarifierungsproblems in der Motorfahrzeughaftpflicht-Versicherung illustriert.

1. Vorbemerkungen

Dieser Beitrag stellt eine leicht überarbeitete Fassung des Vortrags dar, den der Verfasser am ASTIN-Kolloquium in Luzern, 9.–12. Juni 1965, gehalten hat. Den eigentlichen Anstoss zur Wahl des Vortragsthemas hatte eine Arbeit von Bühlmann (1964) über Erfahrungstarifierung in der Motorfahrzeughaftpflicht-Versicherung gegeben. Es erscheint deshalb angebracht, im nächsten Abschnitt aufzuzeigen, wie dieser Problemkreis auf eine natürliche Weise zu Bayes-Methoden führt.

2. Das Tarifierungsproblem

Es ist klar, dass die Prämien einer Motorfahrzeughaftpflicht-Versicherung so festgesetzt werden müssen, dass die totalen Prämieneinnahmen und Schadenauszahlungen der Versicherungsgesellschaft sich etwa die Waage halten. Aber auch für den einzelnen Versicherten sollten sich Prämien und Schadenauszahlungen «im Mittel» ungefähr ausgleichen. Die zweite Forderung kann vielleicht am zwingendsten mit einer spieltheoretischen Überlegung begründet werden: unter freien Wettbewerbsbedingungen verhindert diese Forderung nämlich, dass sich eine Versicherungsgesellschaft auf Kosten der anderen bereichern kann, indem

sie ausgewählte niedrige Risiken durch Angebot eines Spezialrabatts an sich zieht. Was soll aber der Ausdruck «im Mittel» bedeuten?

Nehmen wir einmal an, jede einzelne Versicherungspolice habe eine gewisse, durch eine Zahl p charakterisierte Schadenanfälligkeit: Zum Beispiel sei p die Wahrscheinlichkeit, dass die Police im nächsten Kalenderjahr mindestens einen Schadenfall hat; p könnte an sich zeitlich veränderlich sein, doch werden wir dies ausser acht lassen. Allgemeiner könnte p für eine endliche oder unendliche Folge von Zahlen $p_0, p_1, \ldots$ stehen, wo p_k etwa die Wahrscheinlichkeit bezeichnet, dass im nächsten Jahr genau k Schadenfälle auftreten, oder man könnte zwischen leichten und schweren Unfällen unterscheiden, oder p könnte für die erwartete Schadenhöhe stehen, usw. Die Zahl p ist in jedem einzelnen Fall unbekannt, aber innerhalb der Gesamtpopulation (oder innerhalb der etwas homogeneren Untergruppen einer Stratifikation) kann man die Verteilung von p genähert schätzen, indem man den vergangenen Schadenverlauf einer genügend grossen Zahl von Policen untersucht.

Die Schadenanfälligkeit einer neuen Police könnte durch den Mittelwert von p in der Gesamtpopulation geschätzt werden. In den folgenden Jahren kann man diese Schätzung von p auf Grund des beobachteten Schadenverlaufs sukzessive verbessern.

Wenn man nun anstrebt, dass jeder Versicherungsnehmer diejenige Prämie zahlen soll, die seinem individuellen p angemessen ist, führt die sukzessive Verbesserung der Schätzung von p von selbst auf variable Prämiensysteme, wie sie nunmehr in verschiedenen Ländern in Gebrauch sind.

Ich möchte das hier auftretende Schätzproblem herausgreifen. Es ist ein typisches sogenanntes Bayes-Problem: man hat a priori Information über p in Form einer Wahrscheinlichkeitsverteilung mit nachfolgenden Beobachtungen zu kombinieren.

In der Tat wurde denn auch das in Gebrauch stehende schweizerische Prämienstufensystem mit Hilfe von Bayes-Methoden konstruiert (*Bichsel* 1964).

3. Asymptotische Eigenschaften von Bayes-Schätzungen

Wir ersetzen jetzt das versicherungstechnische Problem durch ein Urnenmodell. Nehmen wir an, wir hätten eine mit verfälschten m-seitigen Würfeln gefüllte Urne vor uns; wir werden uns auf den Fall $m = 3$

beschränken, da er bereits alles Wesentliche zeigt. Jeder solche Würfel ist durch drei nicht-negative Zahlen mit Summe $p_1 + p_2 + p_3 = 1$ charakterisiert, wo p_k die Wahrscheinlichkeit ist, k Augen zu würfeln, oder, was dasselbe ist, durch einen Punkt p eines Dreiecks (nämlich durch den Schwerpunkt von drei Massen p_1, p_2, p_3, die in den Eckpunkten des Dreiecks angebracht sind). Den Vektor (p_1, p_2, p_3) bezeichnen wir mit p.

Wenn wir einen solchen Würfel eine grosse Zahl n von Malen werfen, dann ist wohlbekannt, dass die relative Häufigkeit $h_k = n_k/n$ des Auftretens von k Augen mit Wahrscheinlichkeit 1 gegen p_k konvergiert, und dass die Verteilung des Vektors $n^{\frac{1}{2}}(h_1 - p_1, h_2 - p_2, h_3 - p_3)$ gegen eine zweidimensionale Normalverteilung strebt.

Nehmen wir jetzt an, die Verteilung der Verfälschtheit der Würfel in unserer Urne sei bekannt. Diese sogenannte a priori Verteilung entspricht einer gewissen Wahrscheinlichkeitsverteilung auf dem Dreieck; der Einfachheit halber nehmen wir an, sie besitze eine Dichte $f(p)$.

Wir greifen jetzt einen Würfel zufällig aus der Urne heraus und werfen ihn n Male, wobei n_k Male die Augenzahl k auftritt ($k = 1, 2, 3$). Wir möchten nun das unserem Würfel eigene p schätzen. Im Anschluss an *Bayes* (1763) geschieht das folgendermassen.

Zuerst berechnet man die sogenannte a posteriori Dichte $f_n(p)$, d.h. die bedingte Wahrscheinlichkeitsdichte der Zufallsgrösse p, gegeben dass (n_1, n_2, n_3) beobachtet wurde. Diese Dichte ist

$$f_n(p) = \frac{p_1^{n_1} p_2^{n_2} p_3^{n_3} f(p)}{\int p_1^{n_1} p_2^{n_2} p_3^{n_3} f(p)\, dp},$$

wobei dp das Flächenelement des Dreiecks bezeichnet.

Dann schätzen wir den wahren Wert unseres p durch den Erwartungswert $\bar{p}$ von p unter der a posteriori Verteilung:

$$\bar{p} = \int p\, f_n(p)\, dp.$$

Diese sogenannte Bayes-Schätzung $\bar{p}$ hat folgende Eigenschaften. Sie minimalisiert die bedingte Erwartung des quadratischen Fehlers $(\bar{p} - p)^2$, gegeben dass (n_1, n_2, n_2) beobachtet worden ist, wie man leicht sieht. Da sie dies aber für alle möglichen Werte der n_k tut, folgt sofort, dass die Bayes-Schätzung unter allen möglichen Schätzungen von p den kleinsten (unbedingten) Erwartungswert des Fehlerquadrats $(\bar{p} - p)^2$ liefert.

Wie benimmt sich die Bayes-Schätzung für grosse n? Einige der im folgenden erwähnten Sätze besitzen ein ehrwürdiges Alter, indem sie schon von Laplace um 1820 in Spezialfällen bewiesen wurden. Die Beweise für den allgemeinen Fall sind jedoch recht neu. Natürlich kann ich die Allgemeinheit der Sätze in dieser kurzen Übersicht nicht ausloten, und ich werde mich damit begnügen, sie an unserem speziellen Urnenmodell zu erläutern. Mit $p^0 = (p_1^0, p_2^0, p_3^0)$ bezeichnen wir im folgenden die (unbekannten) Konstanten, welche den tatsächlich gezogenen Würfel charakterisieren.

Das allgemeinste Resultat über Bayes-Schätzungen ist dasjenige von

Doob (1949). *Für fast alle Werte von* p^0 *(in bezug auf die a priori Verteilung) konvergiert die a posteriori Verteilung gegen die Punktmasse* 1 *in* p^0, *und* $\bar{\mathrm{p}}$ *konvergiert stochastisch gegen* p^0, *wenn* n *gegen* ∞ *strebt.*

Man sagt, eine Schätzung sei konsistent, wenn sie stochastisch gegen den wahren Wert konvergiert. *Doobs* Satz kann also etwas kürzer so ausgesprochen werden: Die Bayes-Schätzung ist für fast alle p^0 konsistent.

Dieser Satz ist unter äusserst allgemeinen Voraussetzungen gültig und ist vollkommen zufriedenstellend, falls $f(p)$ die wahre a priori Verteilung wiedergibt. Wenn aber $f(p)$ bloss eine Annäherung an die wahre Verteilung ist, oder wenn man die Bayes-Methode bloss als bequemes Schätzverfahren benützt, hilft dieser Satz nicht viel. Es wäre dann nämlich an sich denkbar, dass alle Würfel in unserer Urne zur Ausnahmemenge vom Mass 0 gehören, wo die Schätzung nicht zu konvergieren braucht. – *Doobs* Satz ist übrigens in allerneuester Zeit von *Lorraine Schwartz* (1965) weiter verallgemeinert worden.

In unserem Fall ist jedoch ein schärferes Ergebnis richtig:

Freedman (1963). *Die Bayes-Schätzung* $\bar{\mathrm{p}}$ *ist konsistent für alle* p^0 *im Träger der a priori Verteilung, und nur für diese.*

Ein Punkt p^0 gehört genau dann zum Träger einer Wahrscheinlichkeitsverteilung, wenn jede beliebig kleine Umgebung von p^0 eine von 0 verschiedene Wahrscheinlichkeit besitzt.

Freedman (1963). *Wenn die a priori Verteilung in* p^0 *eine stetige, von* 0 *verschiedene Dichte besitzt, dann ist die a posteriori Verteilung von* $\mathrm{n}^{\frac{1}{2}}(\mathrm{p}-\mathrm{h})$ *asymptotisch normal, und die Bayes-Schätzung* $\bar{\mathrm{p}}$ *fällt asymptotisch mit*

der Maximum-Likelihood-Schätzung $\mathrm{h} = (\mathrm{n}_1/\mathrm{n}, \mathrm{n}_2/\mathrm{n}, \mathrm{n}_3/\mathrm{n})$ *zusammen, d.h.* $\mathrm{n}^{\frac{1}{2}}(\overline{\mathrm{p}} - \mathrm{h}) \to 0$ *fast sicher.*

Somit verhalten sich Bayes-Schätzungen für grosse n unter recht schwachen Regularitätsbedingungen ganz ausgezeichnet und sind insbesondere von der Wahl der a priori Verteilung weitgehend unabhängig.

Ich werde jetzt Freedmans Beweise kurz andeuten. Sei

$$H(h,p) = -\sum_k h_k \log p_k$$

die Entropie von p relativ zu h (der Begriff ist aus der Informationstheorie entlehnt). Für festes h ist $H(h,p)$ eine konvexe Funktion von p, mit Minimum in $p = h$. Die a posteriori Dichte kann geschrieben werden

$$f_n(p) = c\, e^{-nH(h,p)} f(p).$$

Für wachsendes n hat der Exponentialausdruck ein immer ausgeprägteres Maximum in $p = h$. Nach dem Gesetz der grossen Zahlen konvergiert h gegen p^0. Wenn nun p^0 zum Träger der a priori Verteilung gehört, kann man daraus schliessen, dass die a posteriori Verteilung sich näher und näher bei p^0 konzentriert.

Dass die a posteriori Verteilung asymptotisch normal ist, sieht man ein, indem man $H(h,p)$ um $p = h$ in eine Taylorreihe entwickelt, nach den quadratischen Termen abbricht und die Restglieder abschätzt.

Die oben erwähnten Sätze lassen sich noch etwas verfeinern, wie folgt.

Wenn man Würfel mit abzählbar unendlich vielen Seiten zulässt, gilt Doobs Satz ungeändert; Freedmans Sätze bleiben richtig unter der zusätzlichen Bedingung, dass der tatsächlich gezogene Würfel zufällig nur endlich viele Seiten hat (d.h. $p_k^0 = 0$ ausser für endlich viele k).

Wenn aber $p_k^0 > 0$ für unendlich viele k, stellt es sich überraschenderweise heraus, dass die Bayes-Schätzung für die meisten Paare, bestehend aus einer a priori Verteilung und einem p^0, nicht konsistent ist (Freedman 1965)!

Die Sätze bleiben jedoch gültig für a priori Verteilungen, welche «tail-free» sind; diese von Freedman eingeführte hinreichende Bedingung besagt, dass unter der a priori Verteilung die geeignet normierten p_k mit grossem k von den p_k mit kleinem k unabhängig sein sollen.

Man kann sogar «Würfel» mit kontinuierlich vielen Seitenflächen zulassen; Konsistenz bleibt erhalten, falls die a priori Verteilung «tail-free» ist (*Fabius* 1964).

Bis jetzt bin ich der Frage der sogenannten *persönlichen Wahrscheinlichkeiten* bewusst ausgewichen. Die a priori Verteilungen sind nur sehr selten genau bekannt, aber häufig hat man recht genaue subjektive Meinungen über die relativen Chancen. Wenn zum Beispiel jemand gewillt ist, beide Seiten einer 4:1 Wette auf Eintreffen eines gewissen Ereignisses (eines Rennsieges durch XY o.ä.) einzugehen, könnte man sagen, er gebe diesem Ereignis eine persönliche Wahrscheinlichkeit von 0.8.

Die Befürworter des Gebrauchs von persönlichen Wahrscheinlichkeiten (z.B. *Savage* 1954) behaupten, eine (ideale) vollständig rationale Person würde ihren persönlichen a priori Glauben gemäss der Formel von Bayes mit Beobachtungsresultaten kombinieren, um so eine persönliche a posteriori Wahrscheinlichkeit zu erhalten. Ob man nun die Einführung solcher idealisierter Personen für sinnvoll erachtet oder nicht, Freedmans Sätze zeigen immerhin, dass jeder persönliche a priori Glauben wenigstens zu ausgezeichneten asymptotischen Resultaten führt – sofern man vorsichtig genug ist, diesen Glauben hinreichend glatt, «tail-free» und mit genügend grossem Träger zu wählen!

4. Empirische Bayes-Methoden

In Abschnitt 2 sagte ich etwas voreilig, man könne «die Verteilung von p genähert schätzen, indem man den vergangenen Schadenverlauf einer genügend grossen Zahl von Policen» untersuche. Diese Behauptung ist nur teilweise richtig. Jeder Autofahrer fährt nur eine sehr beschränkte Anzahl von Jahren unter annähernd konstanten Bedingungen, und deshalb kann seine persönliche Unfallanfälligkeit nicht besonders genau geschätzt werden. Diese Ungenauigkeit geht in die Bestimmung der a priori Verteilung von p ein und kann im allgemeinen nicht eliminiert werden, indem man die Anzahl der untersuchten Policen vergrössert.

Es gibt jedoch Ausnahmen. In gewissen Fällen kann man nicht nur die a priori Verteilung schätzen, sondern man kann sogar statistische Verfahren konstruieren, welche die Kenntnis der a priori Verteilung nicht voraussetzen, aber asymptotisch dasselbe leisten, wie wenn man diese Kenntnis gehabt hätte. Diese Einsicht ist Herbert Robbins zu verdanken.

Um die Idee zu fixieren, nehmen wir an, die Anzahl Schadenfälle pro Versicherungspolice und pro Kalenderjahr habe eine Poissonverteilung mit Parameter λ:

$$p_k(\lambda) = e^{-\lambda} \frac{\lambda^k}{k!}, \qquad (k = 0, 1, \ldots),$$

und jeder einzelnen Versicherung komme ein individueller und unbekannter Wert von λ zu. Wir nehmen an, die Werte von λ seien in der totalen Population gemäss einer unbekannten a priori Verteilung G verteilt. Wenn man somit eine Police aufs Geratewohl herausgreift, dann ist die Wahrscheinlichkeit, dass sie im nächsten Jahr genau k Schadenfälle hat, gleich

$$p_k(G) = \int p_k(\lambda)\, dG(\lambda).$$

Nun sei eine grosse Zahl n von Policen während, sagen wir, einem Jahr beobachtet worden, und es sei $h_k = n_k/n$ die relative Häufigkeit der Policen mit genau k Schadenfällen. Die Zahl h_k kann als Schätzung für $p_k(G)$ gebraucht werden. Robbins definiert nun eine Folge von Schätzungen G_n von G, welche mit wachsendem n gegen G konvergiert. Damit man eine solche Folge konstruieren kann, ist offensichtlich notwendig, dass die folgende Identifizierbarkeitsbedingung erfüllt ist: wenn $p_k(G) = p_k(G')$ für alle k, dann ist $G = G'$. Diese Bedingung ist für die zusammengesetzte Poisson-, jedoch nicht für die zusammengesetzte Binomialverteilung erfüllt. Der einfachste (aber nicht-konstruktive) Weg zur Beschreibung einer solchen Folge G_n ist die Methode der minimalen Distanz von Wolfowitz: unter allen Kandidaten für die a priori Verteilung von λ wählt man dasjenige G_n, für welches die $p_k(G_n)$ am nächsten bei den beobachten h_k liegen. Es ist dabei weitgehend beliebig, wie man die Distanz zwischen den $p_k(G_n)$ und den h_k definiert.

Wenn man nicht an G selber, sondern nur an einem beschränkteren Vorhaben interessiert ist, z. B. an der Schätzung der Schadenfälligkeit λ einer individuellen Police, könnte man zwar die Bayes-Schätzung unter Zugrundelegung der a priori Verteilung G_n benützen, doch gibt es auch viel einfachere Methoden. So hat *M. V. Johns* (1956) gezeigt, dass man λ durch

$$(k+1)\frac{h_{k+1}}{h_k}$$

schätzen kann, falls die betreffende Police k Schadenfälle gehabt hat,

und dass diese Schätzung asymptotisch gleich gut ist wie die Bayes-Schätzung bei Benützung der wahren a priori Verteilung G. Für Einzelheiten und für weitere Literaturangaben verweise ich auf *Robbins* (1964). – Die Theorie ist noch im Entwicklungsstadium, und man weiss noch kaum, wie sich diese Schätzungen für kleine und mässig grosse Werte von n verhalten.

Literaturverzeichnis

Bayes, Th. (1763). Abgedruckt in: Biometrika *46* (1958), 293–315.

Bichsel, F. (1964). Erfahrungs-Tarifierung in der Motorfahrzeughaftpflicht-Versicherung. Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker *64*, 119–130.

Bühlmann, H. (1964). Optimale Prämienstufensysteme. Mitteilungen der Vereinigung schweizerischer Versicherungsmathematiker *64*, 193–214.

Doob, J.L. (1949). Application of the theory of martingales. Coll. Int. du CNRS, Paris.

Fabius, J. (1964). Asymptotic Behavior of Bayes' Estimates. Ann. Math. Statist. *35*, 846–856.

Freedman, D. (1963). On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case. Ann. Math. Statist. *34*, 1386–1403.

Freedman, D. (1965). On the Asymptotic Behavior of Bayes Estimates in the Discrete Case II. Ann. Math. Statist. *36*, 454–456.

Johns, M.V. (1956). Contributions to the theory of non-parametric empirical Bayes procedures in statistics. Columbia Univ. Diss.

Laplace, P.S. (1820). Théorie Analytique des Probabilités. Paris.

Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. Ann. Math. Statist. *35*, 1–20.

Savage, L.J. (1954). The Foundations of Statistics. Wiley, New York.

Schwartz, Lorraine (1965). On Bayes Procedures. Z. f. Wahrscheinlichkeitstheorie *4*, 10–26.

Résumé

Cet exposé nous introduit dans de nouveaux et importants progrès dans le domaine de la statistique de Bayes, en particulier dans les travaux de D. Freedman sur le comportement asymptotique des estimations de Bayes et dans les méthodes empiriques de Bayes, publiées par H. Robbins. L'importance de ces résultats pour la science actuarielle est illustrée par le problème de la tarification en assurance responsabilité civile des véhicules automobiles.

Summary

The present review forms an introduction to the important recent advances made in the field of Bayesian statistics, in particular to the work of D. Freedman on asymptotical relationship of Bayesian estimates and to the empirical Bayesian methods of H. Robbins. The actuarial relevance of these results is illustrated by problems of rating in motor car third party liability insurance.

Riassunto

La presente relazione introduce nei recenti e importanti progressi nell'ambito della statistica di Bayes, in particolare nei lavori di D. Freedman in merito al comportamento asintotico delle stime di Bayes e nei metodi empirici di Bayes di H. Robbins. La rilevanza attuariale di questi risultati viene illustrata sulla base del problema di tariffazione nell'assicurazione di responsabilità civile per le autovetture.