Was sagen uns medizinische Tests?

Volker Jentsch
http://www.volkerjentsch.de

April 2021, ergänzt April 2023

In der Medizin gibt es bekanntlich zahlreiche Testverfahren, um herauszufinden, ob jemand von einer bestimmten Krankheit befallen oder nicht befallen ist. Man erinnere sich (wenn auch mit Unbehagen) an Korona und die unaufhörlichen Teste unendlich vieler Menschen. Bei den Verfahren handelt es sich vorwiegend um Untersuchungen von Körperflüssigkeiten, also Blut, Urin, Schleim, Sperma etc. Besondere Bedeutung erlangt der Test als Prostata- Pränatal- Mammographie oder Herz/Kreislauf-Screeing, um anhand der Ergebnisse frühzeitig intervenieren zu können. Diese sind zumeist aber mehrdeutig. Ein positives Ergebnis bedeutet nicht zwangsläufig, dass die Person auch tatsächlich krank ist. Es gibt eben auch die Möglichkeit, dass sie fälschlich als krank bezeichnet wird. Wie also kann Klarheit in die Angelegenheit gebracht werden?

Man lese zum Beispiel G. Gigerenz[1], der sich in seinem Buch, unter anderem, mit diesem Thema befasst und beschreibt, wie nach einfacher Überlegung der Test aufhellt, also zum Beispiel in „richtig positive" und „falsch positive" Aussagen zerlegt werden kann; oder greife, falls an den Grundlagen interessiert, aus der riesigen Zahl von Lehrbüchern der Wahrscheinlichkeitsrechnung das von N.Henze[2] heraus. Ich will mit diesem Artikel anbieten, was bei den Zitierten fehlt: den nach erfolgtem Test Ratlosen ein kleines, interaktives Programm an die Hand geben, mit dessen Hilfe sie selbst ihren Test nach den Regeln der Kunst interpretieren können. Vorab müssen aber noch einige Grundlagen erläutert werden, die zum Verständnis des Ganzen unabdingbar sind.

Vor dem Test: Die Hersteller der Teste, egal ob Hard- oder Software, testen ihre Produkte in Studien, um herauszufinden, wie gut sie sind. Sie sind dann gut, wenn sie mit großer Wahrscheinlichkeit die kranke Person als krank und die gesunde Person als gesund erkennen, also die kranke von der gesunden Person sicher unterscheiden können. Dann ist $P(+|K)$ die Wahrscheinlichkeit, dass Ereignis + eintritt, unter der Bedingung, dass die Getesteten definitv krank sind. Diese (bedingte) Wahrscheinlichkeit nennen die Mediziner die Sensitvität p_{se} des Tests. Entsprechend ist $P(-|G)$ die Wahrscheinlichkeit, mit welcher der Test Personen, die erwiesenermaßen gesund sind, als gesund erkennt. Diesen

[1] Gerd Gigerenz, „Risiko", Bertelsmann Verlag 2014
[2] Norbert Henze, „Stochastik für Einsteiger", Springer Spektrum, 2017

Wert nennen die Mediziner die Spezifität p_{sp} des Tests. Wie wir später sehen werden, ist eine hohe Spezifität (in der Nähe von 1) geeignet, um den Zustand K mit großere Sicherheit zu identifizieren. Erinnern wir uns (wenn auch wiederum ungern) an Korona. Auf den Selbst-Test-Packungen sollten beide Werte angegeben sein, wenn nicht, kann man sie im Internet entdecken. Meist sind sie größer als > 0.9, die besseren liegen eher bei > 0.95

Nach dem Test: Wie leicht einzusehen, gibt es vier mögliche Ergebnisse des Tests. Davon zunächst die beiden wichtigen:
Getestet: positiv: Das ist für die Praxis der eigentlich wichtige Fall; es entsteht Unruhe und Besorgnis bei der Testperson, andererseits Neugier und Forscherdrang beim Medizin-Personal. Doch auch hier gibt es (glücklicherweise) zwei Möglichkeiten: krank und gesund. Welche wahrscheinlicher ist, hängt ab von genau drei Faktoren. Hier ist die Formel zur Berechnung von $P(K|+)$:

$$P(K|+) = \frac{P_{se}P(K)}{(1 - P_{sp})(1 - P(K)) + P_{se}P(K)} \tag{1}$$

Da es zwei Zustände gibt, ist die Gegenwahrscheinlichkeit

$$P(G|+) = 1 - P(K|+) \tag{2}$$

Es gibt zwei weitere, eher weniger wichtige Fälle: Test negativ, Patient gesund; dieser Fall wird mit $P(G|-) = P(G|-)$ notiert; in aller Regel kann man wohlgemut nach Hause gehen:

$$P(G|-) = \frac{P_{sp}(1 - P(K))}{(1 - P_{se})P(K) + P_{sp}(1 - P(K))} \tag{3}$$

Ein leichter Zweifel bleibt bestehen, denn auch hier gibt es die Gegenwahrscheinlichkeit, als *falsch negativ* notiert:

$$P(K|-) = 1 - P(G|-) \tag{4}$$

Die Formeln gehen auf Thomas Bayes zurück.[3] Eine verständliche Darstellung der Formeln (1) und (3) in Form von Pfad-Wahrscheinlichkeiten findet sich in zahlreichen Lehrbüchern; recht anschaulich in dem schon vorab zitierten Buch von G.Gigerenzer. Bleibt zu klären, was wir mit der Bayses Statistik in konkreten Fall gewinnen.

– Die Ausage des Tests läßt sich in falsch und richtig klassifizieren;

– Die Einschätzung des Zustands der Testperson ist nach dem Test um ein Vielfaches sicherer geworden.

[3] Dieter Wickmann hat sein mit vielen originellen Anwendungen gespicktes Buch („Bayes Statistik“, Wissenschaftsverlag, 1990) nicht nur seiner Frau, sondern unausgesprochen auch diesem englischen Gelehrten aus dem achtzehnten Jahrhundert gewidmet.

Diese aus den Formeln ablesbaren Vorteile sollen anhand des folgenden Beispiels veranschaulicht werden.

Beispiel: Wir erinnern uns ein weiteres Mal an die Korona Pandemie. Die Prävalenz von den über Sechzigjährigen erreicht damals über längere Zeit Werte von 1000-10000 Kranken pro 100000 Einwohner. Somit ist für diese Bevölkerungsgruppe die Wahrscheinlichkeit, zu erkranken, im Durchschnitt $P(K) = 0.01$ bzw. $P(K) = 0.1$. Das ist die sogenannte Priori-Wahrscheinlichkeit. Gehen wir von $P(K) = 0.1$ aus. Geschätzt befinden sich also 10000 Erkrankte in ihrer Stadt. Unser Testpaar H und F gehört dieser Gruppe an und will, angesichts der allgemeinen Verbreitung von Furcht und Schrecken, über den jeweiligen Zustand Klarheit gewinnen, die auf gesichertem Wissen basiert.[4] Da sie die langen Schlangen vor den test-Zentren nicht ausstehen können, nehmen sie die überall angebotenen Selbsttests zu Hilfe. In der Apotheke reagiert man mit Unverständnis, als F nach der Sensitivität und Spezifität des Tests fragt. Auch die eilens herbeigerufene Apothekerin vermag die Sitaution nicht zu klären. Eher unwillig entscheidet sie sich zu der Aussage, dass danach noch nie jemand gefragt habe. „Stets gibt es ein erstes Mal“, gibt F behend zurück, während M für die Unkenntnis der Chefin ein gewisses Verständnis hat, ist ihm doch selbst nicht ganz klar, was die nachgefragten Daten zu bedeuten haben. Und warum sie das überhaupt wissen wollten? möchte die Apothekerin wissen. F nimmt die Schachtel in die Hand und entdeckt die fraglichen Parameter: $P_{se} = P_{sp} = 0.9$. Ob noch andere Test verfügbar seien? Die Parameter des angebotenen Ware seien nicht gerade erste Sahne. Erneutes Unverständnis. Sei's drum, sagen sich A und B schließlich, besser diese als keine.

Zu Hause angekommen, nehmen sie sich Abb. 1 des vorliegenden Artikels zur Hand. In dieser ist der Einfluß der verschiedenen Größen auf die mittels (1) errechnete Wahrscheinlichkeit illustriert. Die unterste Kurve iust für die Test-Wahrscheinlichkeiten ihres soeben erworbenen Selbst-Tests gerechnet. Ihre persönliche Vor-Wahrscheinlichkeit wäre folglich: wenn positiv, dann mit einer Wahrscheinlichkeit von 50% krank. Dass die Wahrscheinlichkeit nicht näher an 100% liegt, ist der Ursache geschuldet, dass sich unter den positiv Getesten nicht nur Kranke, sondern auch Gesunde tummeln, und zwar in gleichem Verhältnis. Immerhin ist die Nach-Wahrscheinlichkeit um den Faktor 5 besser als die Vor-Wahrscheinlichkeit, die von einem Anteil von 10% ausgeht. Wären die Vorwahrscheinlichkeiten näher an der eins, also z.B. die Spezifität 0.99 (wie z.B. beim HIV-Test) könnte F laut Abb. 1 mit einer Sicherheit von 90% rechnen. Das Ergebnis: H ist negativ, und F ist positiv. Ist letzteres glaubwürdig? F weiß Abhilfe. Sie hat im oben erwähnten Buch von *Henze* gelesen und studiert daraufhin Abb.2. Sie wiederholt den Test und erhält wiederum eine positive Antwort. Sieht in Abb. 2, dass sich damit ihre Infektion zu fast 90% bestätigt hat. Wäre sie positiv und dennoch gesund, gäbe es dafür nur eine Wahrscheinlichkeit von 10%. Ein dritter Test erübrigt sich somit. F ist gefaßt, denn sie ist schon mit ganz anderem fertig geworden. Wichtig ist ihr, dass jetzt Klarheit

[4] Nomenklatur: G=gesund, K=krank; H=Herr, F=Frau.

besteht.

DerClou: Das eigentliche Hightlight der Angelegenheit ist folglich die Wiederholung des Tests. Damit läßt sich die Vertrauenswürdigkeit des Ergebnisses, vorausgesetzt, die Tests sind voneinander unabhängig und haben auch im nächsten und den folgenden Versuchen das gleiche Ergebnis, dramatisch steigern. Es folgt die mit Anzahl (n) der Wiederholungen modifizierte Formel (1):

$$P(K|+,n) = \frac{(P_{se})^n P(K)}{(1-P_{sp})^n(1-P(K)) + (P_{se})^n P(K)} \tag{5}$$

Anekdote: Übrigens hatte ich dem jetzigen Gesundheitsminister Lauterbach zu Korona-Zeiten den Vorschlag übersandt, die Sicherheit der verabreichten oder selbst angewendeten Tests durch Wiederholung zu verbessern. Offenbar hat er die darin enthaltene Begründung, die wie oben ausgeführt auf den eheren Gesetzen der Stochastik basiert, nicht verstanden; jedenfalls habe ich weder Dank noch Antwort erhalten. Vielleicht hat er die Formel heimlich bei sich selbst ausprobiert, sich verrechnet und die Formel dem Papierkorb übergeben.
Deshalb, liebe Leute: Ihr selbst müßt nicht rechnen. Ihr müßt euch nur einige Parameter besorgen, und nach Eingabe derselben auf dem Bildschirm bekommen ihr in Windeseile die für euch passenden Ergebnise, egal, ob es sich um Covid, Prostata, Brustkrebs etc. handelt. Hier ist die URL:
http://volkerjentsch.de/Covid_Testgetestet.html

Faustregel: Je besser der Test gesunde Personen als gesund erkennt, je näher als die Spezifität P_{sp} an der *eins* liegt, umso größer die Wahrscheinlichkeit, bei positivem Test-Ergebnis die Anzahl der falsch-positiven Resultate zu mindern, bzw. umgekehrt definitiv Kranke auch als krank zu erkennen. Die fälschlich als krank Getesteten (siehe (2)) erweisen sich in Fig. 1 bei $P(K) = 0.001$ als nah bei *eins* und gehen gegen *null*, wenn $P(K) = 0.1$. Nehmen wir ein weiteres Beispiel, eines mit anhaltend großer Aktualität: Brustkrebs. Die Prävalenz über den Lebenszeitraum der Frau, so habe ich gelesen, liegt bei etwa 10%. Spezifität und Sensitivität des Tests entsprechen denen in Abb. 1. Frau geht zum Screening und bekommt ein positives Ergebnis. Wie groß ist die Wahrscheinlichkeit, dass sie tatsächlich Krebs hat? Aus Abb.2 ergibt sich eine Wahrscheinlichkeit von 50%: von zwei Frauen hat nur eine Brustkrebs. Vielleicht liegt die Wahrscheinlichkeit, an Brustkrebs zu erkranken, aber auch deutlich niedriger, etwa bei 5%. Dann sind laut Abb. 1 nur drei von 10 positiv getesteten Frauen tatsächlich erkrankt! Also, liebe Frauen, seien Sie optimistisch! Es gibt eine gute Chance, dass sie gesund sind.

Die Abbildungen: In Abb. 2 ist der als „Clou“ vorgestellte Effekt der Wiederholung illustriert. Schon bei geringer Prävalenz P(K) erreicht der Test beim 3. Versuch 100 %ige Sicherheit: Bei positivem Testergebnis ist die Testperson definitiv krank. In Abb. 3 geht es um den Gewinn, wenn das Testergebnis gemäß Bayes Statistik interpretiert wird. Schon bei einer Spezifität von 0.9 ist

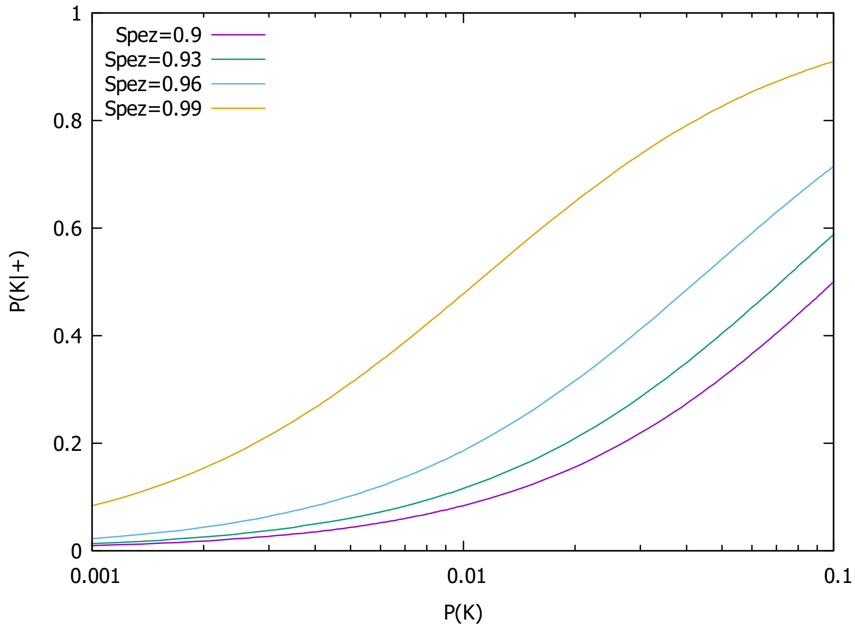

Abbildung 1: Posteriori-Wahrscheinlichkeit P(K|+) als Funktion der Priori-Wahrscheinlichkeit P(K), für verschiedene Spezifitäten.

die Posteriori-Wahrscheinlichkeit etwa zehnmal besser als die Priori. Das gilt vor allem für kleinen Prävalenzen, also bei geringer Wahrscheinlichkeit in der Bevölkerung, zu erkranken (wie z.B. an Aids). Anders als in Abb. 1 und Abb. 2 geht es hier um die Wahrscheinlichkeit P(G|-). Sie wird vor allem durch die Sensitivität beeinflußt. Für P(G)$>$ 0.9, entsprechend P(K)$<$0.1, konvergieren die Kurven gegen eins. Was deutlich macht, dass bei geringerem Krankheitsgeschehen die falsch positiven Testergebnisse an Bedetung gewinnen; falsch negative kommen so gut wie nicht vor. *Testfrage: wieso folgt das aus Abb.4?*

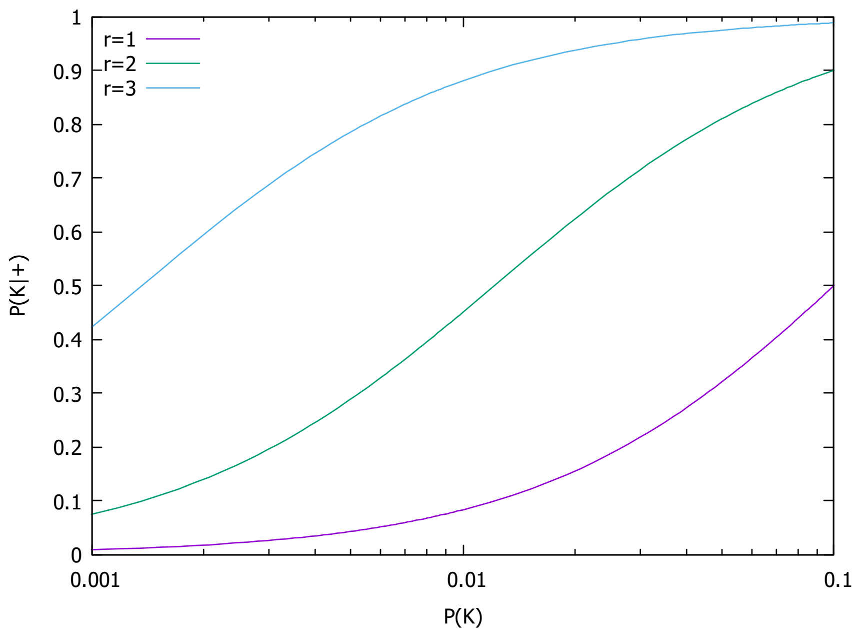

Abbildung 2: Posteriori-Wahrscheinlichkeit P(K|+) als Funktion der Priori-Wahrscheinlichkeit P(K), für aufeinanderfolgend ausgeführte Tests: Die Möglichkeit der Verbesserung eines Tests ist dessen Wiederholung.

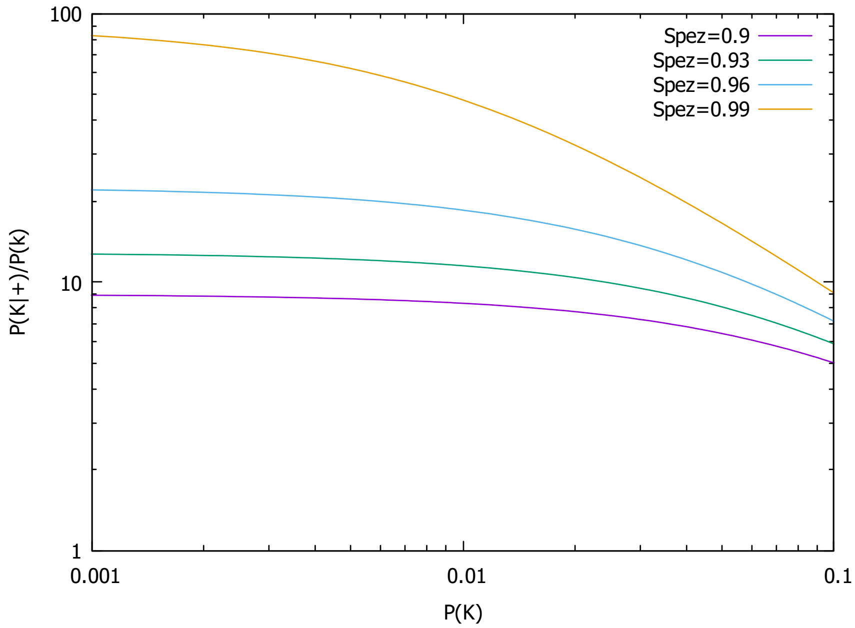

Abbildung 3: Verbesserung der Vorwahrscheinlichkeit durch Anwendung von Bayes: P(K|+) /P(K) als Funktion von P(K), für verschiedene Spezifitäten.

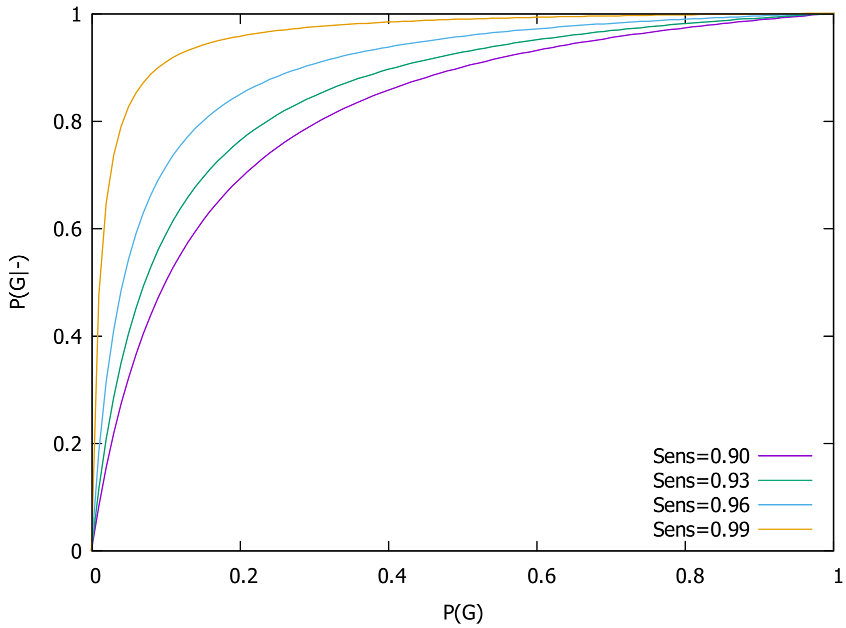

Abbildung 4: Posteriori-Wahrscheinlichkeit P(G|-) als Funktion der Priori-Wahrscheinlichkeit P(G), für verschiedene Sensitivitäten.