Über die Resultate medizinischer Teste

Volker Jentsch
http://www.volkerjentsch.de

April 2021, ergänzt April 2023 und März 2025

Selbsttest. Wir erinnern uns, wenn wohl auch eher unwillig, als Anfang der zwanziger Jahre der Korona-Virus grassierte. Überall wurden mit Hilfe zarter Stäbchen Proben aus den Schleimhäuten der Nase gezogen und mit frei verkäuflicher Testware auf krank oder gesund geprüft. Doch konnte man den Ergebnissen glauben?

Test-Eigenschaften. In der Medizin gibt es bekanntlich zahlreiche Testverfahren, um herauszufinden, ob jemand von einer bestimmten Krankheit befallen oder nicht befallen ist.
Bei den Verfahren handelt es sich vorwiegend um Untersuchungen von Körperflüssigkeiten, also Blut, Urin, Schleim oder Sperma; Gewebeproben, Funktionsprüfungen der Körperorgane, Sinnesorgane wie Hör-, Seh- und Gleichgewichtstest, sowie bildgebende Verfahren. Besondere Bedeutung haben Teste im Bereich Prostata, Brust und Herz/Kreislauf. Sofern die Krankheit nicht bereits manifest ist, geht es darum, anhand der Test-Ergebnisse Klarheit über das Leiden zu verschaffen oder zu intervenieren, um Schlimmeres zu verhindern. Die Ergebnisse solcher Teste sind aber keinesfalls so eindeutig, wie von (unkundigen) Ärzten behauptet. 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. Es ist also wichtig, Klarheit in die Angelegenheit zu bringen.

Gert Giggerenzer, einst Direktor des MPI für Bildungsforschung, hat mit seinem Buch „Risiko"[1] dazu wichtige Beiträge geleistet. Dort werden, unter anderem, die Geheimnisse des Tests auf krank oder gesund diskutiert, aufgehellt und leicht nachvollziehbar in „richtig positive“ und „falsch positive“ Aussagen unterteilt. Natürlich gibt es auch die zahlreichen Lehrbüchern der Wahrscheinlichkeitsrechnung, so zum Beispiel das im Teil 2 des Artikels erwähnte Buch von N.Henze, in denen das Anliegen in einem größeren Kontext behandelt wird. Ich bin mit diesem Teil meines Artikels zwischen den beiden Möglichkeiten. Insbesondere werde ich die Darstellung mit einer Zugabe versehen. Bei dieser handelt es sich um ein interaktives Programm, mit dessen Hilfe in Windeseile die eigenen Testergebnisse, wie zum Beispiel bei wiederaufflammendem Covid,

[1] Gerd Gigerenz, „Risiko“, Bertelsmann Verlag 2014

zu Hause in aller Ruhe am Computer nach den Regeln der Kunst interpretiert werden können.

Vor dem Test. Die Hersteller der Teste, egal ob es dabei um Hard- oder Software handelt, testen ihre Produkte in Studien, um herauszufinden, wie gut sie sind. Sie sind dann gut, wenn sie mit großer Wahrscheinlichkeit (typischerweise oberhalb der 90%) die kranke Person als „krank" erkennt. Das unumstößliche Wissen über „krank" oder „nicht krank" ist mittels anderer Verfahren, zum Beispiel durch chirurgischen Eingriff gesichert worden. Dann ist $P(+|K)$ die Wahrscheinlichkeit, dass Ereignis + (hier gleichbedeutend mit positiv=krank) eintritt, unter der Bedingung, dass die Getesteten krank sind. Diese (bedingte) Wahrscheinlichkeit nennen die Mediziner die *Sensitivität* des Tests, hier abgekürzt durch P_{se}. Entsprechend ist $P(-|G)$ die Wahrscheinlichkeit, mit welcher der Test Personen, die erwiesenermaßen gesund sind, zumindest nicht an der Krankheit leiden, auf die getestet wird, auch als gesund erkennt. Sie wird *Spezifität* genannt und hier mit P_{sp} gekennzeichnet. Je größer Sensitivität und Spezifität, umso sicherer ist die Aussage „krank" oder „gesund". In beiden Fällen bleibt aber eine, wenn auch zumeist geringe Wahrscheinlichkeit, dass das Gegenteil zutrifft, also „gesund" bzw." krank". Der Sinn, den Test anzuwenden, besteht in der Absicht, sich darüber Klarheit zu verschaffen, oder wissenschaftlich ausgedrückt, die sogenannte *a priori Wahrscheinlichkeit* durch die *a posteriori Wahrscheinlichkeit* zu verbessern, sie ggf. sogar durch diese zu ersetzen. Ziel ist, $P(K|+)$ herauszufinden – das ist die Zahl, die mir sagt, mit welcher Wahrscheinlichkeit ich krank bin, nachdem der Test positiv verlaufen ist.

Nach dem Test. Erinnern wir uns (wenn auch wiederum ungern) an Covid. Auf den Selbst-Test-Packungen sollten Sensitivität und Spezifität angegeben sein, wenn nicht, kann man sie im Internet entdecken (was zugleich vorweg eine Art Test bedeutet: ist diese Information nicht vorhanden, sollte das Material dem Verkäufer zurückgegeben werden; siehe unten). Meist sind Sensitivität und Spezifität größer als 0.9, die guten liegen eher bei 0.95.
Außer diesen beiden Werten wird noch die Prävalenz $P(K)$ benötigt. Im ersten Teil des Artikels wird die Prävalenz als das durchschnittliche Erkrankungsrisiko definiert; dieses ist ggf. nach Alter, Ort und Geschlecht zu differenzieren. $P(K)$ kann im Robert Koch Institut erfragt werden, sollte es nicht im Internet zu finden sein. Fehlt $P(G)$; das ist die Gegenwahrscheinlichkeit von $P(K)$.

Wie leicht einzusehen, gibt es vier mögliche Ergebnisse des Tests: $P(K|+)$, $P(K|-)$ sowie $P(G|-)$ und $P(G|+)$.
Fall (I), *getestet „positiv"*: wie groß ist die Wahrscheinlichkeit, dass ich tatsächlich krank bin? Gemäß Wahrscheinlichkeitsrechnung (und die allein ist an dieser Stelle maßgeblich) errechnet sich diese gemäß

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

Da es zwei Zustände gibt, nämlich krank und nicht krank (letzteres optimistisch mit G gekennzeichnet) ruht die Hoffnung auf der Gegenwahrscheinlichkeit, die als falsch positiv oder Fall (Ia) bezeichnet wird:

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

Sie ist verschwindend gering, wenn $P(K|+) \approx 1$.
Der andere wichtige Ausgang ist Fall (II), *getestet „negativ"*: wie wahrscheinlich ist es, dass ich tatsächlich nicht krank bin? Hier ist die Antwort:

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

Zweifler wissen dass es auch hier ein Haar in der Suppe gibt, denn auch hier gibt es die Gegenwahrscheinlichkeit, als *falsch negativ* konnotiert:

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

Die Formeln gehen auf Thomas Bayes zurück.[2] Bevor es zu den Beispielen geht, zunächst noch eine etwas andere Sicht auf die Formeln (1) und (3). Dem dient die **Figur 1** ganz unten. Der Test auf den Gesundheitszustand ist in Form eines zweistufigen Experiments dargestellt. Vom Startpunkt (oberster Kreis) verzweigt sich der Baum zur 1. Stufe (K) und (G); die dazugehörenden Pfad-Wahrscheinlichkeiten sind mit $P(K)$ und $P(G)$ notiert. Die Ergebnisse sind (K) krank oder (G) gesund. Wie beim Start, gehen auch von den Zwischenergebnissen zwei Äste in Richtung krank und gesund. Folglich gibt es nun statt der zwei tatsächlich vier Ergebnisse. Interessant ist, dass (K) und (G) auf jeweils zwei unterschiedlichen Pfaden, wenn auch mit deutlich geringerer Wahrscheinlichkeit, erreicht werden können. Das liegt daran, dass sowohl $P(K) < 1$ als auch $P(G) < 1$. Die Tür in den komplementären Zustand bleibt also einen Spalt breit offen. Wie groß dieser ist, kann mit Hilfe von Bayes errechnet werden. Übrigens ist die Summe der Pfadwahrscheinlichkeiten (in der Horizontalen) stets 1; und die Summe $P(+)$ bzw. $P(-)$ ist die (totale) Wahrscheinlichkeit, dass (+) bzw. (−) eintritt. Bleibt zu illustrieren, was wir mit der Bayes Statistik in konkreten Fall gewinnen.

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 (die Prävalenz) , im Durchschnitt $P(K) = 0.01$ bzw. $P(K) = 0.1$.
Wir haben es mit zwei Personen, Mann und Frau, zu tun, die den über Sechzigjährigen angehören. Sie werden der Einfachheit halber mit M bzw. F oder $M\&F$ abgekürzt. Die beiden sind und wollen, angesichts der allgemeinen Verbreitung von Furcht und Schrecken, die über das Land gekommen ist, über ihren Zustand – ob krank oder gesund – Klarheit gewinnen. Das aber nur auf der Basis eines gesicherten Wissens. Und da sie beide mit Bayes und den Grundlagen

[2]Dieter Wickmann, ein Privatdozent aus Aachen, hat sein Steckenpferd „Bayes Statistik" (Wissenschaftsverlag, 1990) nicht nur seiner Frau Mercedes, sondern unausgesprochen auch seinem Helden, dem englischen Privatgelehrten aus dem achtzehnten Jahrhundert gewidmet.

der Stochastik wohlvertraut sind, gehen sie daran, sich die notwendigen Daten zu beschaffen, die laut (1) und (2) erforderlich sind, um ihr individuelles Krankheitsrisiko zu berechnen. Als erstes stattet das Paar dem Gesundheitsamt ihres Landkreises einen Besuch ab. Dort will es die aktuelle Erkrankungsrate $P(K)$ erfahren. Man hört, wenn auch erst nach langem Hin- und Her und nur unter der Versicherung, dass von den Zahlen kein unzulässiger Gebrauch gemacht werde, diese sozusagen ganz bei ihnen bleiben müssten, dass von den 50000 Einwohnern im Landkreis etwa 5000 erkrankt seien. Die Virologen würden daraus ein Erkrankungsrisiko von 10% machen und es als Prävalenz $P(K)$ bezeichnen. Das Paar erschrickt ob solch hoher Zahl, ist aber insoweit zufrieden, als diese Information für das weitere Vorgehen von großer Bedeutung ist. Nun muss ein Test her. Da $M\&F$ 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 Chef-Apothekerin vermag die Situation 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. Seis drum, sagen sich M und F schließlich, besser diese als keine.
Zu Hause angekommen, geht es sogleich ans Testen. Das Ergebnis: *M ist negativ, und F ist positiv.*
M ist mit seinem Ergebnis zufrieden, und da er nicht ausschließen kann, dass seine Befindlichkeit durch weniger günstige Ergebnisse der Rechenei in Frage gestellt wird, verzichtet er fürs erste auf weitere Nachforschung. F dagegen will es genauer wissen. Schließlich ist ihr Testergebnis positiv. Sie will wissen, wie „positiv“ sie denn nun sei, bevor sie daran geht, weitere Schritte zu ergreifen.
Sie fertigt eine Skizze an (siehe Fig.1). Ganz oben ist der Kreis der 50.000. Diese Zahl ist ihr zu groß, um besser damit umgehen zu können, teilt sie durch 100, erhält folglich $N = 500$. Von hier zur ersten Stufe des Experiments zeichnet sie zwei Äste, davon der linke die Wahrscheinlichkeit $P(K)$ der Erkrankten und der rechte die Wahrscheinlichkeit $P(G)$ der Gesunden angibt. Daneben sind die dazugehörenden absoluten Zahlen notiert. Von der Stufe 1 zur Stufe 2 führen wiederum jeweils 2 Äste, und es wird etwas komplizierter, denn jetzt kommt weitere Information mit ins Spiel, und zwar die Eigenschaften des gerade erworbenen Tests, $P_{se} = P_{sp} = 0.9$. Diese sind zugleich die Übergangswahrscheinlichkeiten von der ersten zur zweiten Stufe und sind jeweils links und rechts der Äste notiert, während die Ergebnisse (zum besseren Verständnis) sich als absolute Anteilen in den Kreise befinden. Rot sind die Beiträge zu der Gruppe der „Kranken“, blau zu den „Nicht-Kranken“. Die Zahl der Kranken speist sich somit aus zwei Quellen: die tatsächlich Kranken aus dem linken Teil der Skizze und den irrtümlich als krank erkannten aus der rechten Teil der Skizze. Der große

Teil der falsch Positiven ist Folge der schlechten Spezifität des Tests. Mithin ist es nicht erstaunlich, dass die Nachwahrscheinlichkeit: 45/90=0.5 beträchtlich von der Vorwahrscheinlichkeit abweicht.
Dagegen wird die Quote der Gesunden, nachdem die Bayes Prozedur durchlaufen ist, nicht merklich von den fälschlich negativ getesteten beeinflusst; die ist $405/410 = 0.95$. Insofern war M gut beraten, weitere Überlegungen fallen zu lassen. Aber was ist mit F? Gibt sie sich mit den 50% zufrieden?
Das tut sie nicht. Sie blättert im Buch von *Henze* und findet auf Seite 108 den Hinweis auf wiederholte Durchführung des Tests und auf Seite 372 die dazu gehörende Verallgemeinerung von Bayes (siehe (5)). M wird beauftragt, drei weitere Teste aus der Apotheke zu holen. Der zweite Test bestätigt den ersten. Gemäß Abb.2 springt „krank“ auf 90%. Sie rechnet nach. 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 besteht.

Der Clou: Das eigentliche Highlight 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. Formel (5) ist die Erweiterung von (1) für n- Wiederholungen.

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

Anekdote: Übrigens hatte ich dem vorigen und jetzigen Gesundheitsminister Lauterbach zu Korona-Zeiten den Vorschlag übersandt, die Sicherheit der verabreichten oder selbst angewendeten Tests durch Wiederholung zu verbessern. Ob er das schon wusste und deshalb eine Antwort für ungehörig empfand? Oder weil er davon nichts wusste und mit einer Antwort das hätte bekunden müssen? Für das Wahrscheinlichste halte ich, dass er meine Ausführungen und die darin enthaltene Anwendung der ehernen Gesetzen der Stochastik nicht verstanden hat. Jedenfalls habe ich weder Dank noch Antwort erhalten. Denkbar aber auch, dass er die Formel heimlich bei sich selbst ausprobiert hat, sich verrechnet und die Formel dem Papierkorb übergeben hat. Übrigens hat auch das Robert Koch Institut nicht reagiert. Einzig Gingerenzer hat sich gemeldet und eine Überprüfung in Aussicht gestellt. Wie diese ausgefallen sein mag, davon habe ich allerdings nichts vernommen.
Fazit – liebe Leute: Ihr selbst müsst nicht rechnen. Ihr müsst euch nur einige Parameter besorgen, und nach Eingabe derselben bekommt ihr auf dem Bildschirm die für euch passenden Ergebnisse, 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 also 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 eine Sicherheit von 100% : Bei positivem Testergebnis ist die Testperson definitiv krank. Voraussetzung: die Teste müssen hintereinander das gleiche Ergebnis erbringen! In Abb. 3 geht es um den Gewinn, wenn das Testergebnis gemäß Bayes Statistik interpretiert wird. Schon bei einer Spezifität von 0.9 ist die Posteriori-Wahrscheinlichkeit etwa zehnmal besser als die Priori. Das gilt für kleine Prävalenzen, also bei geringem Krankheitsrisiko (wie z.B. an Aids). In Abb. 4 geht es 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: wenn negativ getestet, bist du sicher nicht krank. Folgerung: Bei geringerem Krankheitsgeschehen gewinnen die falsch positiven Testergebnisse an Bedeutung; falsch negative kommen so gut wie nicht vor.

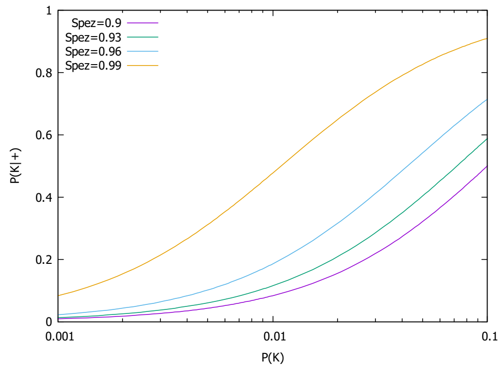


Abbildung 1: Posteriori-Wahrscheinlichkeit P(K|+) als Funktion der Prävalenz P(K), für verschiedene Spezifitäten. Die Prävalenz wird als individuelle „a priori" Wahrscheinlichkeit gesetzt.

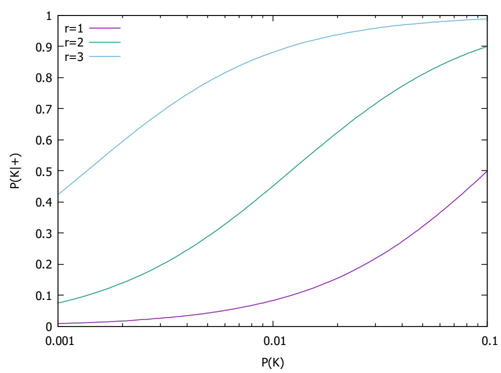


Abbildung 2: Posteriori-Wahrscheinlichkeit P(K|+) als Funktion der Prävalenz P(K), für aufeinanderfolgend ausgeführte Tests: Die Möglichkeit der Verbesserung eines Tests ist dessen Wiederholung.

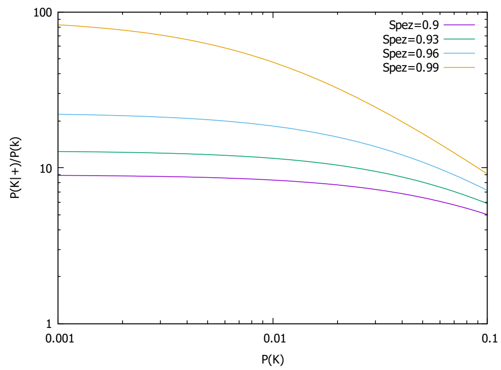


Abbildung 3: Verbesserung der Vorwahrscheinlichkeit durch Anwendung von Bayes: P(K|+) /P(K) als Funktion von P(K), für verschiedene Spezifitäten.

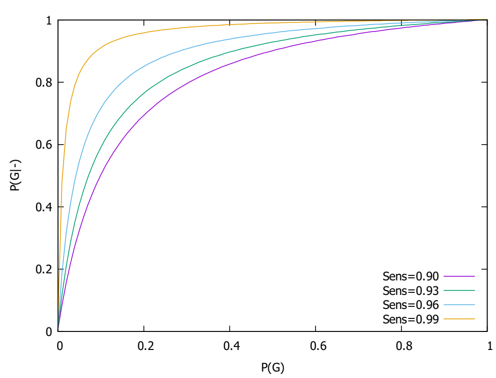
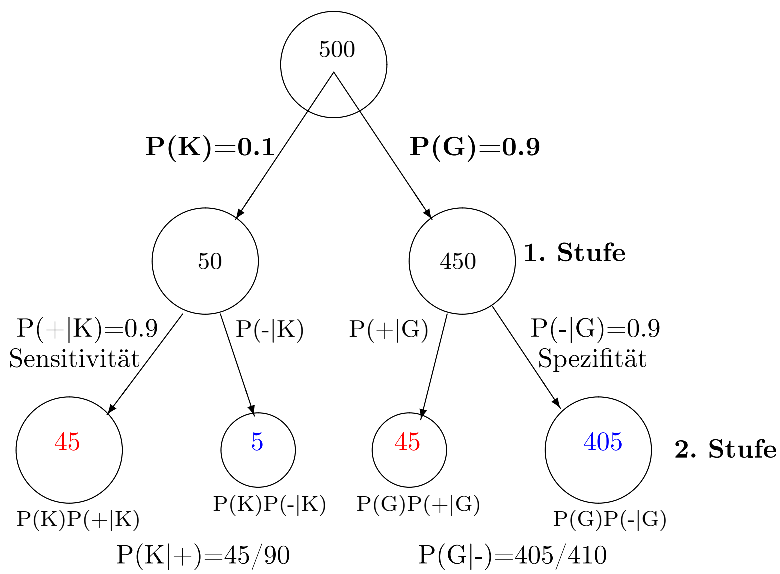


Abbildung 4: Posteriori-Wahrscheinlichkeit P(G|-) als Funktion der „inversen Prävalenz“ P(G)=1-P(K), für verschiedene Sensitivitäten.



Figur 1: Baumdiagramm auf krank oder gesund
Summe der Pfadwahrscheinlichkeiten für jede Verzweigung ist 1