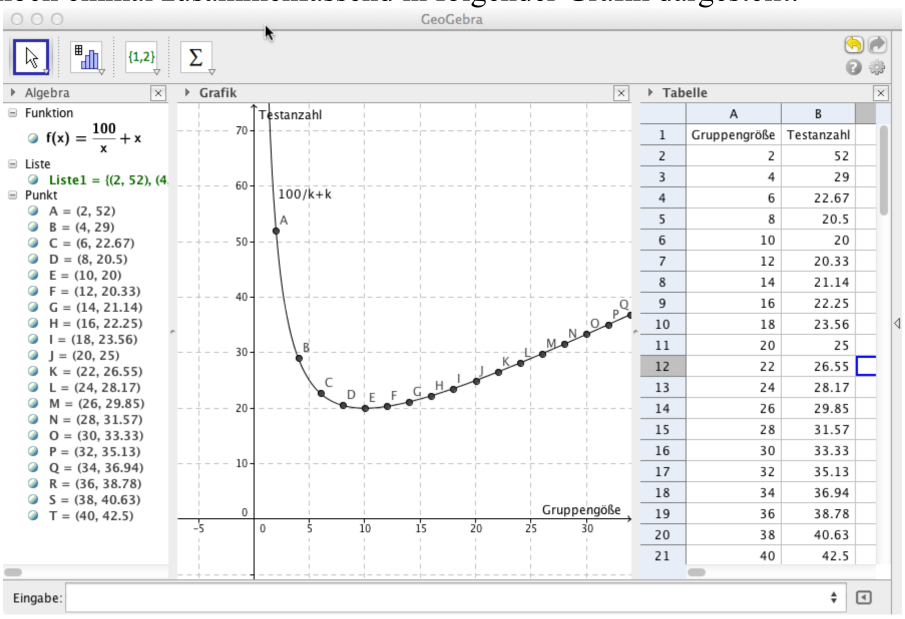
Zu 1)
Dem Problem kann man sich gut experimentell (TR) oder mit Hilfe einer Wertetabelle nähern. Man ordnet der Anzahl an Soldaten, deren Blut man mischt die jeweils erforderliche Testanzahl zu. Die zugrundeliegende Arithmetik ergibt sich aus weiteren Beispielrechnungen: Mischt man das Blut von z.B. jeweils 3 Soldaten, so ergeben sich 100/3≈33 Tests für die Gruppen und 3 Einzeltests für die „infizierte" Gruppe. Also insgesamt 100/3 + 3 Tests. Mischt man 4er Gruppen, so ergeben sich 100/4+4 Tests (bei nicht ganzzahligen Ergebnissen von 100/k könnte man noch darüber diskutieren, inwiefern man dies mathematisch präziser/ stimmiger/ korrekter modellieren kann). Grob genähert ergibt sich die Folge $t(k) = \frac{100}{k} + k$ mit dem man die Anzahl der insgesamt erforderlichen Tests t in Abhängigkeit von der Gruppengröße k berechnen kann.
Nimmt man k als reellwertige Variable an, so kann man mit Hilfe der Differentialrechnung das gesuchte Minimum von t berechnen. Es gilt

$$t'(k) = -\frac{100}{k^2} + 1$$

Die Nullstelle von t'(k) liegt offensichtlich bei k=10. Da $t''(k) = \frac{200}{k^3}$ bei k=10 positiv ist, liegt dort auch das lokale Minimum der Funktion. Mit Hilfe von DGS oder GTR sind diese Ergebnisse noch einmal zusammenfassend in folgender Grafik dargestellt:



Zu 2)
Bei einer Prävalenz (so nennt man den Anteil der Verbreitung einer Krankheit in einer Population) von 2% wird die Situation komplexer, da die Art der stochastischen Modellierung davon abhängt, ob man davon ausgeht, dass das die Eigenschaft erkrankt zu sein eine abhängige oder unabhängige Zufallsgröße ist.

Nimmt man Unabhängigkeit an, d.h. jeder Soldat ist mit einer konstanten Wkt. infiziert (praktisch würde das bedeuten, dass sich die Soldaten z.B. nicht gegenseitig anstecken können), dann kann man die Wkt., dass eine Blutmischung von 2 Soldaten nicht infiziert ist mit

$$0{,}98^2 \approx 0{,}9604$$

berechnen. Dann ist genau 1 Test erforderlich. Die Wkt., dass eine Gruppe infiziertes Blut enthält ist demzufolge $1 - 0{,}98^2$. Dann sind genau 3 Tests erforderlich (einen für die Gruppe und 2 für die beiden Soldaten). Pro Gruppe sind also $1 \cdot 0{,}98^2 + 3 \cdot (1 - 0{,}98^2) \approx 1{,}08$ Tests zu erwarten. Insgesamt sind also in etwa $50 \cdot 1{,}08 \approx 54$ zu erwarten.

Nimmt man keine Unabhängigkeit an, d.h. die Soldat sind mit unterschiedlichen Wkt.'en infiziert (das könnte z.B. bedeuten, dass sich Soldaten untereinander oder bei einer gemeinsam besuchten Prostituierten infizieren könnten), dann kann man die Wkt., dass eine Blutmischung von 2 Soldaten nicht infiziert ist mit Ziehen ohne Zurücklegen aus einer Urne simulieren. In einer Urne befinden sich 100 Kugeln, von denen zwei markiert (infiziert) sind. Die Wkt., dass eine Blutmischung nicht infiziert ist, kann man mit

$$\frac{\binom{98}{2}\binom{2}{0}}{\binom{100}{2}} \approx 0{,}9602$$

berechnet werden. Dann ist genau 1 Test erforderlich. Die Wkt., dass eine Gruppe infiziertes Blut enthält ist demzufolge 1 - 0,9602. Dann sind genau 3 Tests erforderlich. Pro Gruppe sind $1 \cdot 0{,}9602 + 3 \cdot (1 - 0{,}9602) \approx 1{,}08$ Tests zu erwarten, also insgesamt ebenfalls 54, wie im Modell zuvor.

Zu 3)

Bei der Mischungs des Bluts von jeweils 3 Soldaten verändert sich die Formel zur Berechnung der insgesamt zu erwartenden Tests zu

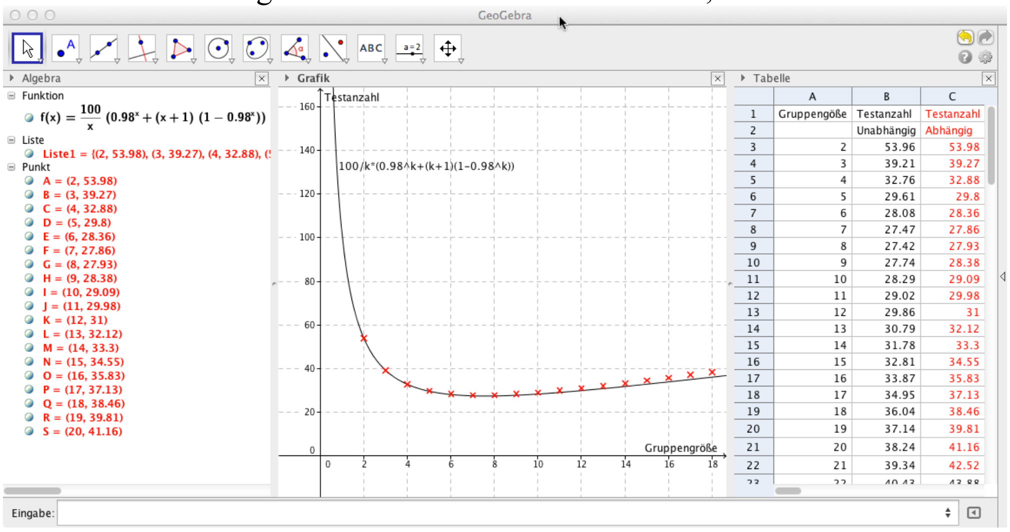
$$\frac{100}{3} \cdot (1 \cdot 0{,}98^3 + 4 \cdot (1 - 0{,}98^3) \approx 39$$

Man untersucht etwa 100 / 3 Gruppen, die Wkt., dass eine Gruppe nicht infiziertes Blut enthält beträgt $0{,}98^3$, wodurch nur 1 Test erforderlich ist. 4 Test (für die Blutmischung und die 3 Einzelproben) sind hingegen mit einer Wkt. von $1\text{-}0{,}98^3$ erforderlich.

Im anderen Modell ergibt sich die gleiche Anzahl an zu erwartenden Tests durch

$$\frac{100}{3} \cdot \left(1 \cdot \frac{\binom{98}{3}\binom{2}{0}}{\binom{100}{3}} + 4 \cdot \left(1 - \frac{\binom{98}{3}\binom{2}{0}}{\binom{100}{3}}\right)\right) \approx 39$$

Für die weiteren Berechnungen erscheint der Einsatz von DGS, CAS oder ein GTR nützlich:



Betrachten wir k wiederum reellwertig, so kann wie zuvor die Funktion

$$t(k) = \frac{100}{k}(0.98^k + (k+1)(1 - 0.98^k))$$

(mit oder ohne Tabellenkalkulation) analysiert werden oder man betrachtet die folgende Folge, die rot in der Abbildung oben zu sehen ist:

$$t(k) = \frac{100}{k}\left(1 \cdot \frac{\binom{98}{k}\binom{2}{0}}{\binom{100}{k}} + (k+1) \cdot \left(1 - \frac{\binom{98}{k}\binom{2}{0}}{\binom{100}{k}}\right)\right)$$

Offensichtlich sind die Abweichungen nur geringfügig und kaum bemerkenswert. Insofern stellt sich die Frage, ob man nicht das „abhängige" Modell vernachlässigt, da im Folgenden auch die Prävalenz verallgemeinert wird.

Zu 4)
Unsere zuletzt betrachtete „unabhängige" Funktion t(k) hängt nun von den Variablen k, n und p ab und wird mit t(k, n, p) bezeichnet. Es gilt

$$t(k,n,p) = \frac{n}{k}((1-p)^k + (k+1)(1-(1-p)^k))$$

Vereinfacht gilt

$$t(k,n,p) = n\left(1-(1-p)^k + \frac{1}{k}\right)$$

Zur unabhängigen Variablen k wird das Minimum von t gesucht. Berechnet wird deshalb

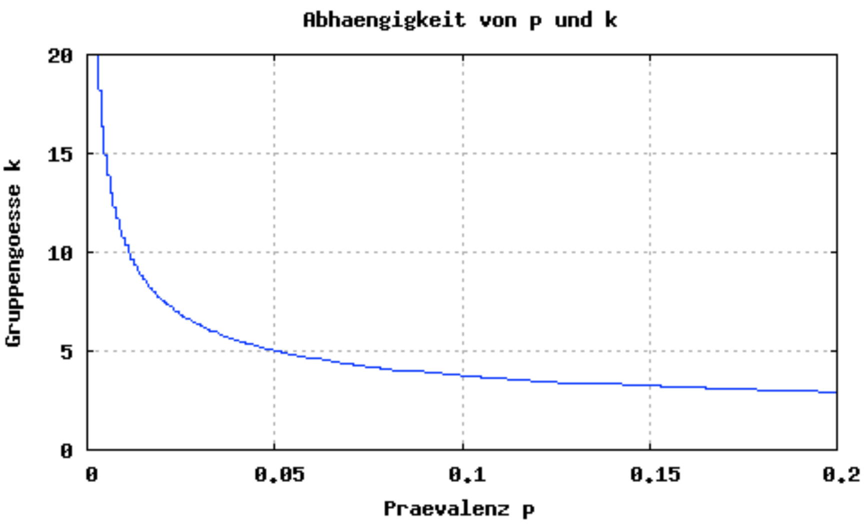
$$t'_{n,p}(k) = n \cdot \left(-\ln(1-p)\cdot(1-p)^k - \frac{1}{k^2}\right)$$

Die Nullstelle dieser Funktion folgt aus

$$-\ln(1-p)\cdot(1-p)^k - \frac{1}{k^2} = 0$$

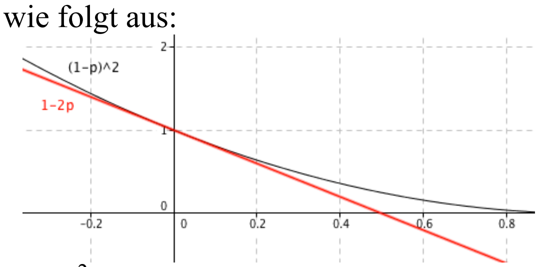
Es bieten sich nun wenigstens 3 Möglichkeiten an, um diese Gleichung zu lösen („lösen" bedeutet hier eine Abhängigkeit zwischen p und k herauszufinden):

- Man experimentiert mit Zahlenbeispielen (wenig zufriedenstellend!)
- Man nutzt ein CAS/GTR/DGS, welches implizite Gleichungen lösen kann:



- Man versucht die Gleichung so zu modifizieren, dass sie geschlossen nach k lösbar ist:

Der „unangenehme" Teil der Gleichung ist der Term $(1\text{-}p)^k$.
Aus der Grafik zuvor ist offensichtlich, dass sich das Screening Verfahren nur für kleine Prävalenzen eignet, also geht es uns im Wesentlichen nur um „kleine" p (wenn eine Krankheit mit 20% in einer Bevölkerung verbreitet ist liegt ohnehin eine Epedemie vor!).
Für k=2 sieht der Term wie folgt aus:



Offensichtlich lässt sich $(1\text{-}p)^2$ im Intervall [0,0.2] gut durch die Tangente im Punkt (0,1) approximieren. Das klappt auch für k>2 „anschaulich" gut (GTR-Einsatz!). Insofern kann die Funktion

$$t(k,n,p) = n\left(1 - (1-kp) + \frac{1}{k}\right) = n\left(kp + \frac{1}{k}\right)$$

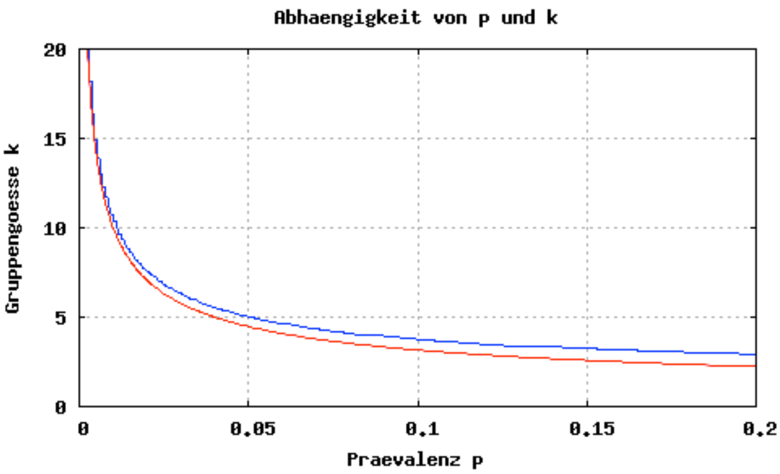
betrachtet werden. Es gilt

$$t'_{n,p}(k) = n\left(p - \frac{1}{k^2}\right) = 0$$

wenn

$$k = \frac{1}{\sqrt{p}}$$

(nur die positive Lösung der Gleichung macht Sinn!). Jetzt kann man rückblickend beide Lösungen noch einmal vergleichen:



Der blaue Graph zeigt die implizite Gleichung, der rote Graph die Näherungslösung $1/\sqrt{p}$

Zu $k = \frac{1}{\sqrt{p}}$ erhält man mit

$$t(n,p) = n\left(\frac{1}{\sqrt{p}}p + \sqrt{p}\right) = 2n\sqrt{p}$$

also eine relativ einfache Funktion, der „etwas“ zu kleine Gruppengrößen zugrundeliegen (was eine größere Testanzahl zur Folge hat) dafür aber aber sehr praktikabel erscheint.