Wilfried Rohm wrohm@aon.at

Stichprobenanweisungen n-c : Simulation und Berechnungen

INHALT

1) Prinzip einer n-c-Stichprobenanweisung

zurück zum INHALT

Bei Stichprobensystemen handelt es sich um eine leicht verständliche Methode der statistischen Qualitätssicherung. Stichproben ersetzen in vielen Fällen 100-%-Prüfungen, insbesondere bei zerstörenden Prüfungen oder Prüfungen an nicht allzu kritischen Teilen (d.h. ein gewisser durchschlüpfender Fehleranteil kann verkraftet werden.) bzw. Prüfungen, die sonst nicht wirtschaftlich durchführbar sind.

In der Praxis wird eine Vielzahl unterschiedlicher Stichprobensysteme verwendet. Hier beschränken wir uns auf den einfachsten und bekanntesten Fall einer sogenannten **Einfachstichprobenanweisung für qualitative Merkmale** beschränken. "Qualitativ" soll heißen, daß die Prüfung eines Teiles nur auf der Basis gut oder schlecht (bzw. fehlerfrei oder fehlerhaft) erfolgt.

Beispiel :
Monatlich werden 20.000 Spritzgußteile bestimmter Spezifikation an eine Firma geliefert. Der Lieferant hat zugesichert, daß höchstens 1% dieser Spritzgußteile unbrauchbar sind. Damit erklärt sich die Firma einverstanden, jedoch soll diese Angabe des Lieferanten im Wareneingang stichprobenartig überprüft werden. Aus der Norm DIN ISO 2859-1 (AQL-Stichprobensystem) wurde folgende Stichprobenanweisung herausgelesen:

$$n - c = 315 - 7$$

Was bedeutet diese Stichprobenanweisung ?

Aus dem Losumfang N=20.000 soll eine Stichprobe von n=315 Spritzgußteilen entnommen und geprüft werden. Die Zahl c wird Annahmezahl genannt, das bedeutet: Sind bis zu 7 fehlerhafte Teile unter den 315 geprüften Teilen, so gilt das Los im obigen Sinne als "o.k." und wird angenommen. Werden mehr als 7 fehlerhafte Teile gefunden, so wird das Los abgelehnt bzw. rückgewiesen. Was dann zu tun ist, muß vertraglich geregelt sein (z.B.: Reklamation / Rücksendung an den Lieferanten).

Fragestellungen
Offensichtlich ist eine Stichprobenentnahme mit einem bestimmten Risiko verbunden.

So können etwa "zufälliger Weise" mehr als 7 schlechte unter den 315 untersuchten Teilen sein, obwohl in der gesamten Lieferung der Ausschußprozentsatz kleiner als 1% ist.

Auf der anderen Seite könnte es sein, daß "zufälliger Weise" die Anzahl der gefundenen Schlecht-stücke ≤ 7 ist, obwohl der Ausschußprozentsatz in der gesamten Lieferung größer als 1% ist.

Zur Abschätzuung dieser Risiken müssen etwa folgende Fragen beantwortet werden können:

a) Wie groß ist beispielsweise die Wahrscheinlichkeit, daß die Lieferung angenommen wird, wenn die Lieferung p gerade 2% fehlerhafte Teile enthält ?
Umgekehrt: Bei welchem Fehlerprozentsatz beträgt die Annahmewahrscheinlichkeit genau 90% ?

b) Wie groß ist die Anzahl der durchschlüpfenden fehlerhaften Teile für einen bestimmten vorgegebenen Ausschußprozentsatz p ?

c) Bei welchem Ausschußprozentsatz p ist dieser Durchschlupf maximal ?

Bei der Beantwortung dieser Fragen stößt man sehr schnell auf Grenzen, was die praktische Berechenbarkeit ohne Computerprogramm oder CAS-Programm betrifft. Daher gibt es auch umfangreiche Tabellen. Gerade deshalb ist es aber sehr reizvoll, diese (rein gedanklich) nicht allzu schweren Fragen mittels Mathcad zu lösen.

2) Simulation der Stichprobenentnahme, Ermittlung der OC

zurück zum INHALT

Wir wollen die Stichprobennahme einer Lieferung simulieren, die einen konstanten "Schlechtanteil" p aufweist und die mittels einer n-c-Anweisung auf Annahme überprüft werden soll.
(n Teile werden geprüft, Annahme der Lieferung, wenn höchstens c schlechte darunter sind)
Indem wir anschließend die Größe p variieren, können wir ermitteln, wie häufig es zu einer ANNAHME bzw. RÜCKWEISUNG kommt und dies schließlich mit den theoretischen Wahrscheinlichkeiten vergleichen.

Die Versuchsentnahme entpsricht (wegen dem konstanten Schlechtanteil p) dem Versuchsschema der Binomialverteilung. Daher können wir mittels der Funktion rbinom die Simulation durchführen.
(Hinweis: rbinom liefert einen Vektor!)

$n := 50$ $c := 2$ $p := 0.02$ $N := 100$ Anzahl der Simulationen

$\text{simulation} := \text{rbinom}(N, n, p)$

$\text{simulation}^T =$

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0	0	1	1	1	0	0	1	0	1	0	1	2	2	4	1	1	0	2	1	2

Man experimientiere mit unrterschiedlichen Werten von p!

Wir wollen nun die relative Häufigkeit der Annahme der Lieferung in einem Diagramm für verschiedene Werte von p (punktweise) darstellen und mit der "theoretischen" Annahmewahrscheinlichkeit in Abhängigkeit vom Ausschußprozentsatz p vergleichen - diese Kurve wird **OPERATIONSCHARTAKTERISTIK (OC)** genannt.
Die Berechnung erfolgt mittels der Verteilungsfunktione G(x) der Binomialverteilung.

Zunächst die reine Simulation (Mathcad-Programm)

```
annahmehäufigkeit(N, n, c, p) := | sum ← 0
                                 | simulation ← rbinom(N, n, p)
                                 | for  i ∈ 0.. N − 1
                                 |   sum ← sum + 1  if  simulation_i ≤ c
                                 | sum
                                 | ───
                                 |  N
```

N Simulationen durchführen und im Feld "simulation" speichern

Die Annahmen zählen ...

Rückgabe der relativen Häufigkeit für die Annahme

annahmehäufigkeit(100, 50, 2, 0.05) = 0.56 — Beispiel für den Aufruf

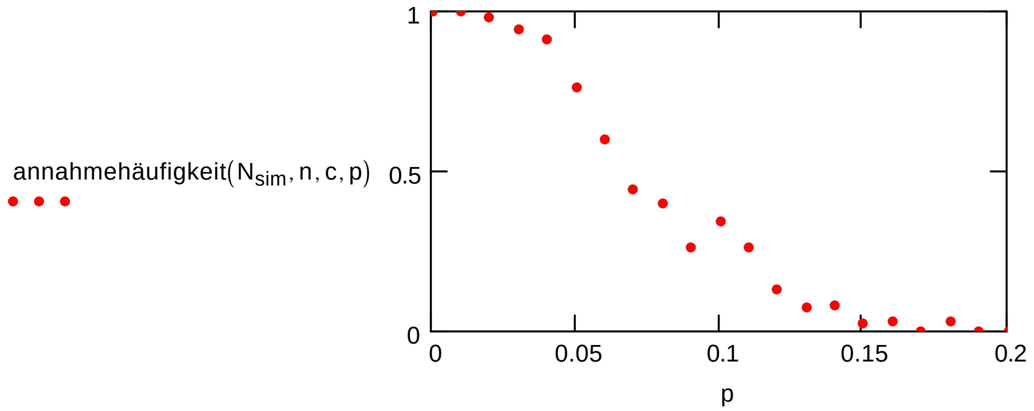
Simulation:

$N_{sim} := 100$ — Anzahl der Simulationen (für jedes p)

$n := 50$ — Simulation der n-c-Anweisung (für nochmalige Simulation das Grafikfenster anklicken und F9 drücken!)

$c := 3$

$p := 0, 0.01 .. 0.20$



Vergleich der Simulation mit rechnerischen Werten

Definition der Binomialverteilung

$g_{bi}(x, n, p) := \mathrm{dbinom}(x, n, p)$ — Wahrscheinlichkeit für genau x schlechte in Stichprobe (n)

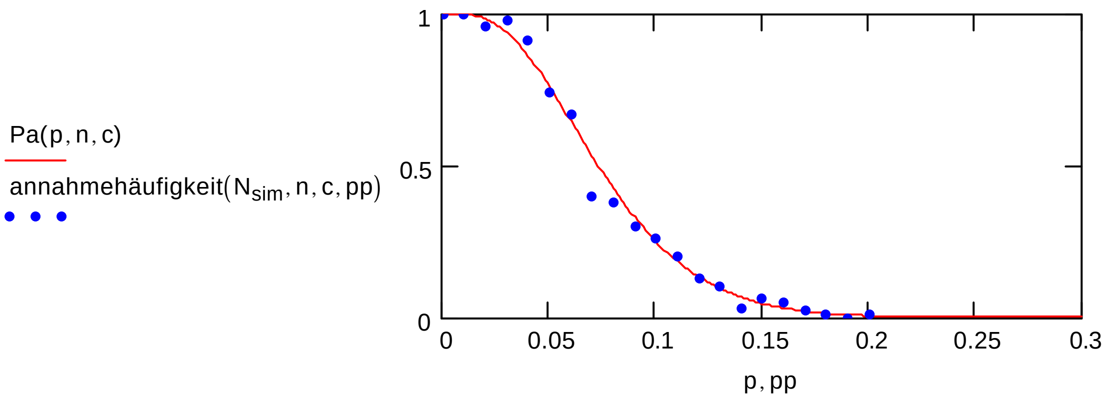
$G_{bi}(x, n, p) := \mathrm{pbinom}(x, n, p)$ — Wahrscheinlichkeit für höchstens x schlechte in Stichprobe(n)

Für die Annahmewahrscheinlichkeit Pa= $P(x \le c)$ gilt: $Pa(p, n, c) := \mathrm{pbinom}(c, n, p)$

$N_{sim} := 100$ $n := 50$ $c := 3$ — Festzulegende Werte (n-c-Anweisung) und N (Anzahl der Simulationen)

$p := 0, 0.001 .. 0.30$ — Werte für die berechnete Kurve

$pp := 0, 0.01 .. 0.20$ — Werte in größeren Schritten für die Simuilation

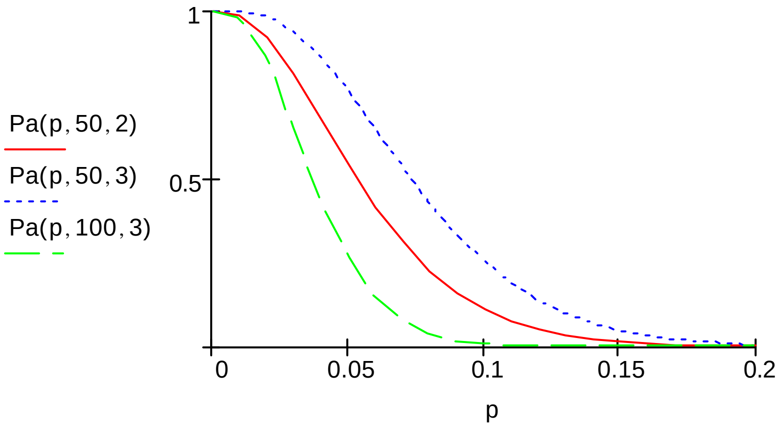


3) Wie hängt der Verlauf der OC von den Parametern n und c ab?

zurück zum INHALT

OC=Annahmewahrscheinlichkeit der Einfachstichprobe n-c

$Pa(p, n, c) := G_{bi}(c, n, p)$ $\quad p := 0, 0.01 .. 0.20$



4) Berechnung der charakteristischen p90 / p10 - Werte

zurück zum INHALT

In der Praxis werden zur Charakteristik einer Stichprobenanweisung häufig 2 Punkte angegeben, die in Kürze den Verlauf der OC beschreiben können:

p_{90} ... Das ist jener Ausschussprozentsatz in der Grundgesamtheit, bei dem die Annahmewahrscheinlichkeit 90% beträgt.

p_{10} ... Das ist jener Ausschussprozentsatz in der Grundgesamtheit, bei dem die Annahmewahrscheinlichkeit 10% eträgt.

Anmerkung:
Die Berechnung dieser Größen gelingt im allgemeinen nur numerisch, da z.B. bei der Stichprobenanweisung 315-7 dazu bereits Gleichungen 315.Grades zu lösen sind!! Dies kann man allerdings nur dann demonstrieren, wenn man für die Binomialverteilung eigene Definitionen verwendet, etwa so:

$$g_{bi}(x, n, p) := \frac{n!}{x! \cdot (n-x)!} \cdot p^{x} \cdot (1-p)^{n-x}$$

$$Pa(p, n, c) := \sum_{k=0}^{c} g_{bi}(k, n, p)$$

$$Pa(p, 315, 7) = 0.9 \rightarrow (1-p)^{315} + 315 \cdot p \cdot (1-p)^{314} + 49455 \cdot p^{2} \cdot (1-p)^{313} + 51$$

Variante zur Berechnung mit numerischen Lösungsblock: $n := 50$ $c := 2$

$p := 0.01$ Startwert für die numerische Suche nach der Lösung

Vorgabe

$Pa(p, n, c) = 0.90$

$p_{90} := \text{suchen}(p)$ $p_{90} = 0.022$

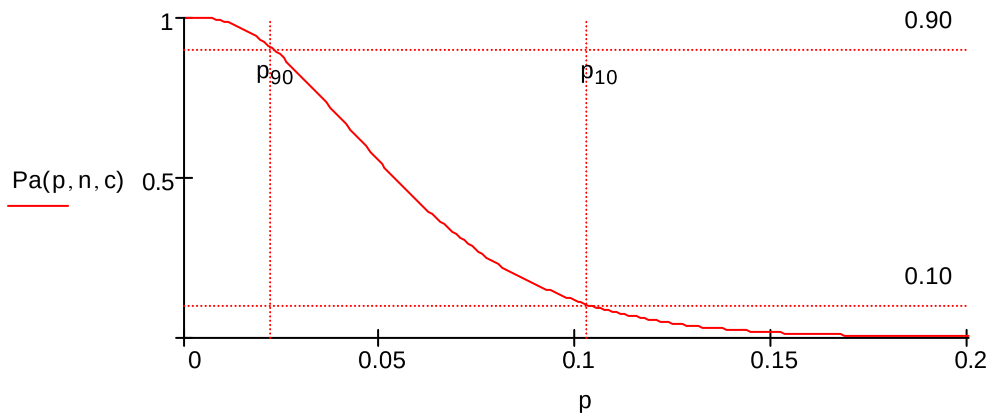
$p := 0.05$

Vorgabe

$Pa(p, n, c) = 0.10$

$p_{10} := \text{suchen}(p)$ $p_{10} = 0.103$

$p := 0, 0.001 .. 0.20$



5) Prüfaufwand

zurück zum INHALT

Prüfaufwand: wieviel Stück werden im Mittel bei bestimmtem Ausschußprozentsatz p geprüft
= Erwartungswert des Stichprobenumfanges

N ... Losumfang ; n ... Stichprobenumfang

$N := 1000$ $n = 50$ $c = 2$

$n_{mittel}(p) := Pa(p, n, c) \cdot n + N \cdot (1 - Pa(p, n, c))$

Begründung: Bei Annahme des Loses werden nur n Stück geprüft, bei Rückweisung erfolgt eine Sortierprüfung (alle N Stück werden geprüft)

Einige Werte in Abhängigkeit vom Ausschußprozentsatz p zeigen: bei schlechter Qualität wird die Prüfung teuer!!

$n_{mittel}(0) = 50$ $n_{mittel}(0.02) = 124.506$

$n_{mittel}(0.01) = 63.126$ $n_{mittel}(0.05) = 486.494$

6) Durchschlupf, Durchschlupfkurve und maximaler Durchschlupf

zurück zum INHALT

Durchschlupf D:

Darunter versteht man den mittleren ANTEIL fehlerhafter Einheiten in den eingehenden Losen (also jener Schlechtanteil, der trotz Stichprobenprüfung "durchschlüpft").
Bei der folgenden Berechnungsformel wird vorausgesetzt, dass rückgewiesene Lose sortiergeprüft werden.

$$D(p,N) := p \cdot Pa(p,n,c) \cdot \frac{N-n}{N}$$

N ... Losumfang (Grundgesamtheit)
n ... Stichprobenumfang
c ... Annahmezahl der n-c-Anweisung
p ... Anteil der Schlechtstücke im Los

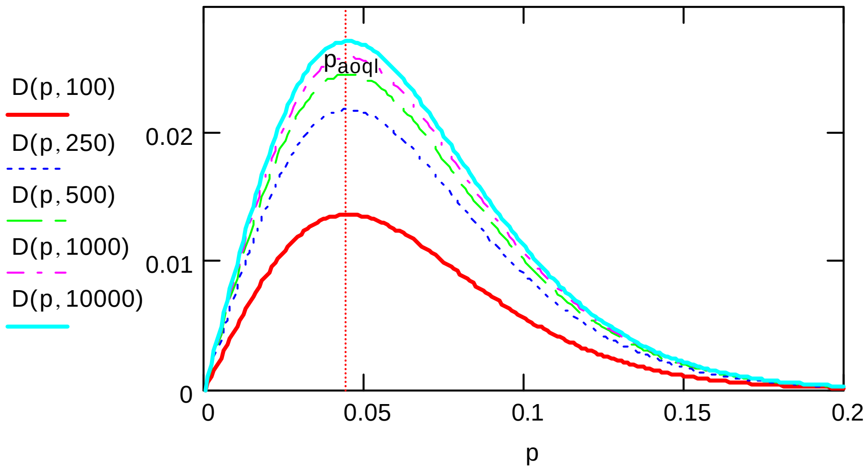
Erläuterung:
p*Pa(p) ... W., daß die Lieferung angenommen wird UND dass ein Stück schlecht ist.
N-n ... die ANZAHL der nicht geprüften Stücke (bei Annahme des Loses)
(N-n)/N ... der ANTEIL der nicht geprüften Stücke im Vergleich zum Losumfang

An welcher Stelle wird der Durchschlupf maximal (Extremwertaufgabe): **p_{AOQL}-Wert**

$n := 50$ $c := 2$ $p := 0.03$ Schätzwert für die Lösung

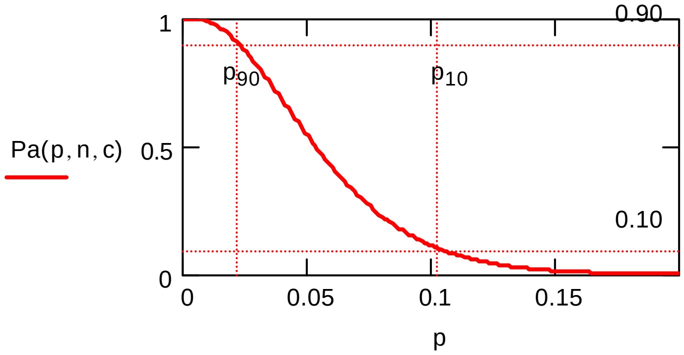
$$p_{aoql} := \mathrm{wurzel}\left(\frac{d}{dp}D(p,100),p\right) \qquad p_{aoql} = 0.04469$$

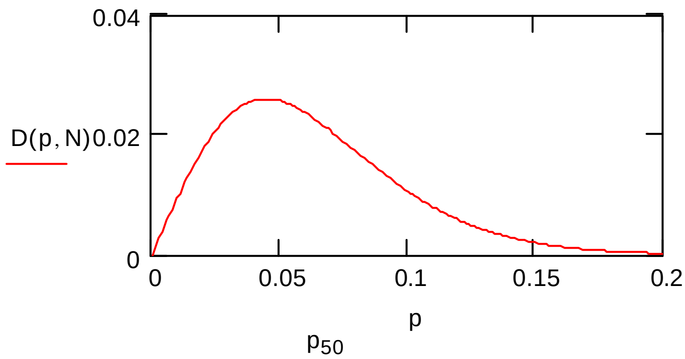
$$p := 0, 0.001 .. 0.2$$



Interpretation:
* Die Stelle des Maximums ist vom Losumfang N unabhängig (mathematisch logisch!)
* Für großes N gilt : D ist ungefähr Pa(p)*p

Gegenüberstellung zwischen OC und Durchschlupfkurve





7) Vergleich der Berechnung mit Binomial- und Poissonverteilung

zurück zum INHALT

In der Praxis wird speziell im Bereich der Stichprobensysteme gerne mit der einfacher handzuhabenden Poissonverteilung statt der Binomialverteilung gerechnet. Allgemein kann die Poissonverteilung als Näherung der Binomialverteilung für kleines p und "nicht zu kleines" n sehr gut verwendet werden. Da diese Bedingungen gerade bei Stichprobensystemen im normalen Anwendungsfall erfüllt sind, sind etwaige Ungenauigkeiten vernachlässigbar. [*Es sei ergänzt, dass derartige Näherungen im Zeitalter der Computeralgebrasysteme nicht mehr so bedeutend sind (wie früher).*]

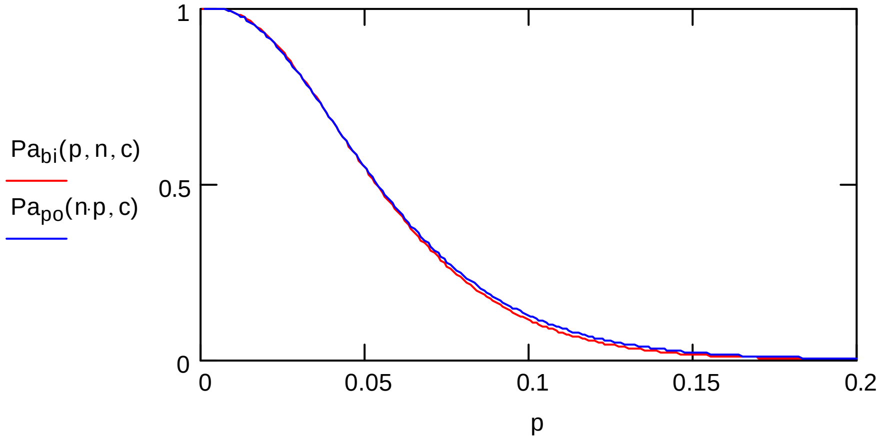
Wir demonstrieren dies hier an Hand des OC-Verlaufes, einmal mit der Binomial- , das andere Mal mit der Poissionverteilung berechnet.

$G_{bi}(x, n, p) := \mathrm{pbinom}(x, n, p)$ $\qquad$ $G_{po}(x, \mu) := \mathrm{ppois}(x, \mu)$

$Pa_{bi}(p, n, c) := G_{bi}(c, n, p)$ $\qquad$ $Pa_{po}(\mu, c) := G_{po}(c, \mu)$

Man experimentiere mit verschiedenen Werten von n und p und interpretiere das Ergebnis:

$n := 50$ $\qquad$ $c := 3$



8) Der AQL-Wert: Bedeutung und Interpretation

zurück zum INHALT

Der Praktiker verwendet für die Charakterisierung der statistischen Risiken einer Stichprobenanweisung (die exakt durch die Operationscharakteristik = OC erfolgt), gerne den sogenannten AQL-Wert.

AQL = Accepatable Quality Limit

Die genormten Stichprobenvorschriften sind nach derartigen AQL-Werten geordnet, z.B:

AQL = 0,25 | 0,40 | 0,65 | 1,0 | 1,5 | 2,5 | 4,0 | 6.5

Eine Stichprobenvorschrift, die bei AQL = 1,0 zu finden ist, wurde so festgelegt, dass ***bis zu* 1% Anteil fehlerhafter Einheiten im Los gemäß dieser Stichprobenvorschrift eine hohe Annahmewahrscheinlichkeit (in der Regel über 90%) gegben ist.**

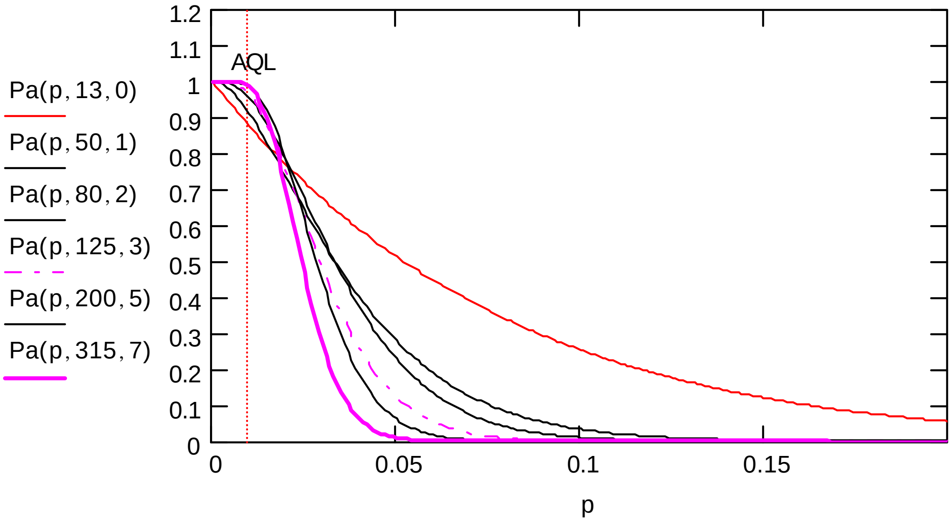
In der Norm ISO 2859 findet man u.a. folgende **n-c-Prüfanweisungen bei AQL=1,0**

n-c =	13 - 0
	50 - 1
	80 - 2
	125 - 2
	200 - 5
	315 - 7

Die spezielle Wahl hängt übrigens in erster Linie vom Losumfang N ab

$Pa(p, n, c) := G_{bi}(c, n, p)$ AQL := 0.01

$p := 0, 0.001 .. 0.2$



Wir sehen aus diesem Schaubild:
* Der AQL-Wert soll eine ungefähre Ahnung vom Verlauf der OC vermitteln.
* Bis zum AQL-Wert (in Prozent) wird "eine hohe Annahmewahrscheinlichkeit" garantiert!

Im speziellen erhalten wir folgende Annahmewahrscheinlichkeiten beim AQL-Wert 1%

$Pa(0.01, 13, 0) = 0.878$ $Pa(0.01, 80, 2) = 0.953$ $Pa(0.01, 200, 5) = 0.984$

$Pa(0.01, 50, 1) = 0.911$ $Pa(0.01, 125, 3) = 0.963$ $Pa(0.01, 315, 7) = 0.985$

$$.59805 \cdot p^3 \cdot (1-p)^{312} + 402464790 \cdot p^4 \cdot (1-p)^{311} + 25033309938 \cdot p^5 \cdot (1-p)^{310} + 1293387680130 \cdot p^6 \cdot$$

$(1-p)^{309} + 57093827594310 \cdot p^7 \cdot (1-p)^{308} = .9$