BEMERKUNGEN ZUM STIRLING'SCHEN AUFZEICHNUNGSBLATT FÜR EXPERIMENTE ZUR WAHRSCHEINLICHKEIT NACH G. GILES

B. WOLLRING

Die folgenden Bemerkungen beziehen sich auf das von *G. Giles* vorgestellte Arbeitsblatt zum schnellen Aufzeichnen relativer Häufigkeiten bei Bernoulli-Experimenten. Das "Aufzeichnungsblatt", wie wir es im folgenden nennen wollen, ist meiner Meinung nach ein derart schönes und praktisches Hilfsmittel, daß mir einige Bemerkungen zur Ergänzung des Textes von *G. Giles* (siehe *Giles*) lohnend erscheinen.

Die Bilder 1 und 2 können als Kopiervorlage* des Aufzeichnungsblattes genommen werden, wobei Bild 2 gegenüber Bild 1 zusätzlich die 90 %-Vertrauensintervalle für p bei k = n/2 enthält. Gegenüber dem Original von *Giles* wurden einige Ablesehilfen hinzugefügt. Ferner ist unter der Skala für den Gewinnanteil ein "Magazin" eingezeichnet, in dem die Ergebnisse mehrerer Versuchsserien wie in einem Histogramm gesammelt werden können. Wir kommen weiter unten darauf zurück. Ist das Magazin nicht erforderlich, so kann es beim Kopieren abgedeckt werden.

Offenbar beabsichtigt *Giles*, das Aufzeichnungsblatt den Schülern zunächst nur mit einer Gebrauchsanweisung in die Hand zu geben und baut darauf, daß sie die Konstruktion im Laufe der prakti-

* Siehe Beilage zu diesem Heft

schen Arbeit immer besser durchschauen. Hier scheint es mir, speziell angesichts der deutschen Curricula, sinnvoll, die Schüler auf die Analogie des Aufzeichnungsblattes zum *Baumdiagramm* hinzuweisen. Daß das Aufzeichnungsblatt ein verzerrtes Baumdiagramm ist, kann der Schüler an einem kurzen Abschnitt beider Strukturen feststellen, indem er sie Punkt für Punkt und Weg für Weg vergleicht. Dazu genügen zwei Abschnitte bis n = 5, die er zum Vergleich auch selbst zeichnen sollte (siehe Bild 3).

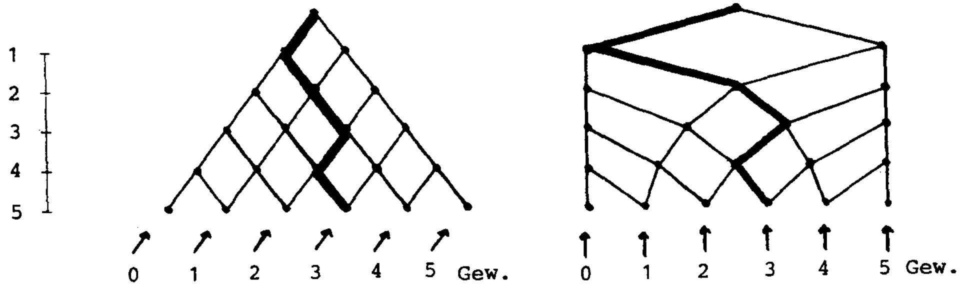

Bild 3 : Wahrscheinlichkeitsbaum in Dreieck-Darstellung und in Giles-Darstellung

Das Identifizieren analoger Punkte und Wege sollte in beiden Richtungen geübt werden. Bei kombinatorischen Überlegungen, wie sie weiter unten folgen, ist der Rückgriff auf die Dreieck-Darstellung oft hilfreich.

Giles weist darauf hin, daß durch den verständigen Gebrauch des Aufzeichnungsblattes die Erarbeitung des Pascal'schen Dreiecks bereits vorbereitet wird. Auf die Binomialkoeffizienten $\binom{n}{k}$ stößt man, wenn man nach der Zahl der Wege vom Start bis zum Punkt

(n; k) fragt. Allerdings erscheint mir wohl einige Lehrerhilfe nötig, bis die Schüler einsehen, daß

(1) $$\binom{n}{k} = \frac{n\cdot(n-1)\cdot\_\_\_\_\cdot(n-k-1)}{1\cdot 2\cdot\_\_\_\_\cdot k} = \frac{n!}{k!\,(n-k)!}$$

die Zahl der Wege vom Start bis zum Punkt (n; k) ist. Meiner Ansicht nach kann diese Formel durch Aktivitäten am Aufzeichnungsblatt allein nicht gefunden werden. Wohl findet der Schüler, wenn er mit $\binom{n}{k}$ die Zahl der Wege vom Start bis (n; k) bezeichnet, die Formeln

(2)

Randformeln: $\binom{n}{0} = 1 \quad \binom{n}{n} = 1$

Schrittformel: $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$

Aber damit ist die Struktur der Formel (1) noch nicht gefunden. Mit einem Beweis durch Induktion über n kann sie wohl bestätigt werden, aber dazu muß ihre Struktur bereits bekannt sein.

Da die Formel (1) jedoch bei der weiteren Diskussion des Aufzeichnungsblattes vonnöten ist, soll sie mit einer einfachen *kombinatorischen Überlegung* erarbeitet werden, bei der das Aufzeichnungsblatt zur Veranschaulichung herangezogen wird, und Wege darin mit Losfolgen einer Ziehung in Zusammenhang gebracht werden.

Wir nehmen an, daß n! als Anzahl der Permutationen von n Objekten bekannt ist. Nun ziehen wir n Lose aus einer Urne, davon k Gewinne, und numerieren die Lose beim Ziehen durch. Der gezogenen Losfolge entspricht dann ein Weg im Aufzeichnungsblatt (siehe Bild 4).

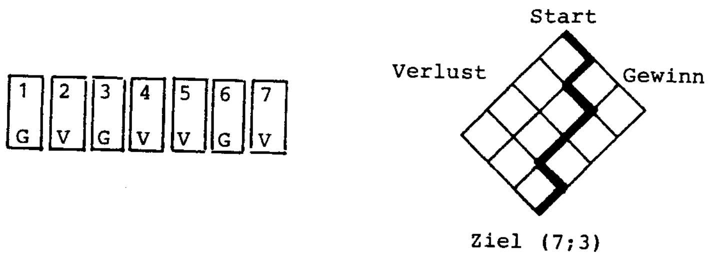

Bild 4: Losfolge mit zugehörigem Weg

Insgesamt erzeugen n! Losfolgen Wege mit dem Ziel (n; k); denn jedes Los kann an jeder Stelle auftreten, ohne daß das Wegziel sich ändert. Aber mehrere Losfolgen erzeugen denselben Weg (siehe Bild 5).

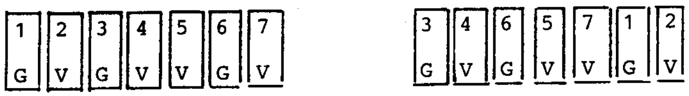

Bild 5 : Losfolgen mit demselben Weg

Vertauscht man die k Gewinnlose untereinander und ebenso die n - k Verlustlose untereinander, so ändert sich der erzeugte Weg nicht. Und das geht auf k! (n - k)! Arten. Also ist $\binom{n}{k}$ gemäß (1) die Zahl der Wege mit Ziel (n; k).

Giles benutzt nun offensichtlich die Formel (1), um die Wahrscheinlichkeit p (n; k) von k Gewinnen unter n Versuchen zu bestimmen:

(3)
$$p\,(n;\,k) = \frac{\text{Zahl der Wege mit Ziel } (n;\,k)}{\text{Zahl der Wege mit n Schritten}}$$
$$= \frac{n!}{k!\,(n-k)!} \cdot \frac{1}{2^{k}}$$

Das ist die Wahrscheinlichkeit, einen bestimmten Punkt (n; k) "auf der Höhe n" zu erreichen, denn alle Wege der Länge n sind bei p = 1/2 ja gleichwahrscheinlich.

Es wäre schön, auch zu dieser Wahrscheinlichkeit einen experimentellen Zugang mit Hilfe des Aufzeichnungsblattes zu haben. Dazu dient das *Magazin* unter der Gewinnskala, in dem die Schüler die Ergebnisse mehrerer Versuchsserien sammeln, die sie entweder durch Wiederholen oder Zusammenfassen der Ergebnisse aus der ganzen Klasse finden. Die Notierung von 30 Versuchsserien zeigt Bild 6.

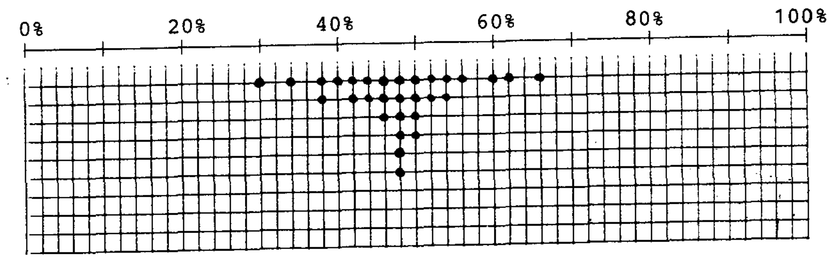

Bild 6 : Notieren der Ergebnisse von 30 Versuchsserien

Damit kann auch die folgende Diskussion der Vertrauensbereiche durch ein Experiment vorbereitet werden. Mit der "Magazin"-Technik erhält man auch bei p = 1/2 qualitativ brauchbare Ergebnisse.

Die 90 %-Vertrauensbereiche für die Anzahl k der Gewinne unter n Versuchen bei p = 1/2 kann man zwar, wie *Giles* bemerkt, "auf dem Aufzeichnungsblatt angeben", aber die Konstruktion der von *Giles* gezeichneten Kurven ist nicht wie die übrige Zeichnung auf Anhieb zu verstehen, jedenfalls nicht für Schüler.

Als Hintergrundinformation für den Lehrer möchte ich daher zunächst erklären, wie ich sie rekonstruiert habe, und anschließend Möglichkeiten nennen, sie mit Schülern zu diskutieren.

Betrachten wir den Anteil k/n der Gewinne bei n Versuchen als binomialverteilte Zufallsgröße A, so hat sie den Mittelwert $\mu = p$ und die Standardabweichung $\sigma = p\ (1 - p)/n$. Für hinreichend hohe Werte von n ist sie nach dem lokalen Grenzwertsatz annähernd normal verteilt. Einer Tabelle zur Normalverteilung entnehmen wir nun, daß für einen beobachteten Wert k/n von A folgende Gleichung gilt:

(4) $$P\ (\mu - 1{,}645\ \sigma \leq \frac{k}{n} \leq \mu + 1{,}645\ \sigma) = 90\ \%$$

Das bedeutet hier:

(5) $$P\ (p - 1{,}645\sqrt{\frac{p\ (1 - p)}{n}} \leq \frac{k}{n} \leq p + 1{,}645\sqrt{\frac{p\ (1 - p)}{n}}) = 90\ \%$$

Speziell bei p = 0,5 erhalten wir für k/n das 90 %-Vertrauensintervall $p_1; p_2$ mit Mittelpunkt 0,5 und den Grenzen

(6) $$p_1\ (n) = 0{,}5 - 1{,}645\ \frac{1}{2\sqrt{n}}$$
$$p_2\ (n) = 0{,}5 + 1{,}645\ \frac{1}{2\sqrt{n}}$$

Typisch ist, daß die Länge des Vertrauensbereiches ein Vielfaches von $1/\sqrt{n}$ ist, was nicht nur bei p = 1/2 gilt. Die Graphen der Funktionen $n \to p_2\ (n)$ sind für $n \geq 8$ in das Aufzeichnungsblatt eingetragen, sie stimmen mit den Kurven im Original von *Giles* überein. Die Annahme der Normalverteilung von A enthält eine Näherung, so daß man bei direkten Berechnungen geringfügig andere Ergebnisse erhält.

Bei der Arbeit mit Schülern bieten sich meines Erachtens zwei Wege zur Erläuterung und Bestätigung dieser Kurven an.

Der erste besteht darin, daß Vertrauensintervall für ein gegebenes n, etwa n = 30, exemplarisch zu bestimmen, wobei Formel (3) benutzt wird. Es wird jedoch aus Zeitgründen kaum möglich sein, diese Berechnung für mehrere, insbesondere für große n durchzuführen.

Gesucht sind also k_1 und k_2, so daß gilt:

(6) $$p\ (30;\ k_1) + p\ (30;\ k_1 + 1) + \ldots + p\ (30;\ k_2) \geq 90\ \%$$

wobei $(k_1;\ k_2)$ ein Intervall mit Mittelpunkt 15 sein soll. Dazu bestimmen wir k_1 so, daß gilt:

(7) $$\frac{30!}{0!(30 - 0)!} + \frac{30!}{1!(30 - 1)!} + \ldots + \frac{30!}{k_1!(30 - k_1)!} < 5\ \% \cdot 2^{30}$$

Damit vermeiden wir die großen Zahlen "in der Mitte" und nutzen die Symmetrie aus. Wir erhalten $5\ \% \cdot 2^{30} \approx 53\ 687\ 091$

und für die Binomialkoeffizienten

$$
(8) \qquad
\begin{aligned}
\binom{30}{0} &= & 1 \\
\binom{30}{1} &= & 30 \\
\binom{30}{2} &= \binom{30}{1}\frac{29}{2} = & 435 \\
\binom{30}{3} &= \binom{30}{2}\frac{28}{3} = & 4\,060 \\
&\quad\vdots & \\
\binom{30}{10} &= \binom{30}{9}\frac{23}{10} = & 32\,906\,445 \\
\binom{30}{11} &= \binom{30}{10}\frac{22}{11} = & 60\,812\,890
\end{aligned}
$$

Also erhalten wir $k_1 = 10$ und $k_2 = 20$ und damit

$$(9) \qquad p_1 = 0{,}333\overline{3} \quad \text{und} \quad p_2 = 0{,}666\overline{6}$$

als Grenzen des 90 %-Vertrauensintervalls für p. Man benötigt etwa 10 Minuten für diese Rechnung, wenn man einen Taschenrechner benutzt. Die Zeichnung bestätigt das Ergebnis. Zum Vergleich nennen wir die entsprechenden Ergebnisse, die man mit Formel (6) gewinnt:

$$(10) \qquad p_1 = 0{,}3499 \quad \text{und} \quad p_2 = 0{,}6501$$

Sie führen ebenfalls auf $k_1 = 10$ und $k_2 = 20$. Diese Methode hat den Nachteil, daß man nicht unmittelbar erkennt, wie die Länge des Vertrauensintervalls von der Zahl n der Versuche abhängt. Daher sei hier vorgeschlagen, das Magazin unter der Gewinnskala zur experimentellen Bestimmung des Vertrauensintervalls zu be-

nutzen.

Dazu läßt man mehrere Schüler, sagen wir 25, mit analogem Gerät gleichzeitig Versuchsserien durchführen und die Ergebnisse jeweils nach n = 10; 20; ...; 50 Versuchen austauschen. Dabei trägt jeder Schüler sämtliche Zwischenergebnisse in sein Magazin ein. Man sieht dann, daß die Verteilung der 25 Ergebnisse sich bei wachsender Länge n der Versuchsserien "zusammenzieht", wie es Bild 7 als Beispiel zeigt.

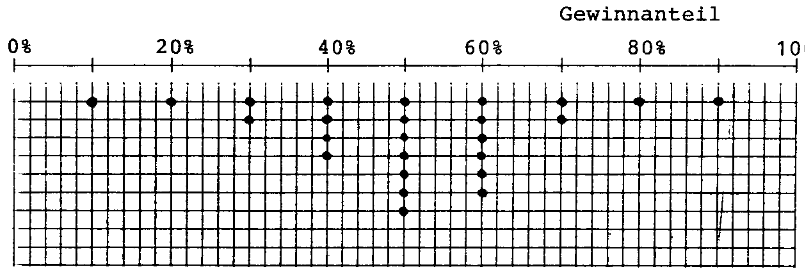

Bild 7a : Ergebnisse von 25 Versuchsserien der Länge 10

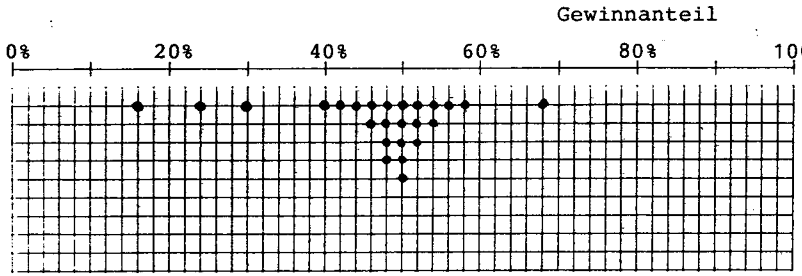

Bild 7b : Ergebnisse von 25 Versuchsserien der Länge 50

Die 90 %-Vertrauensintervalle sind nun jeweils durch Abzählen zu bestimmen.

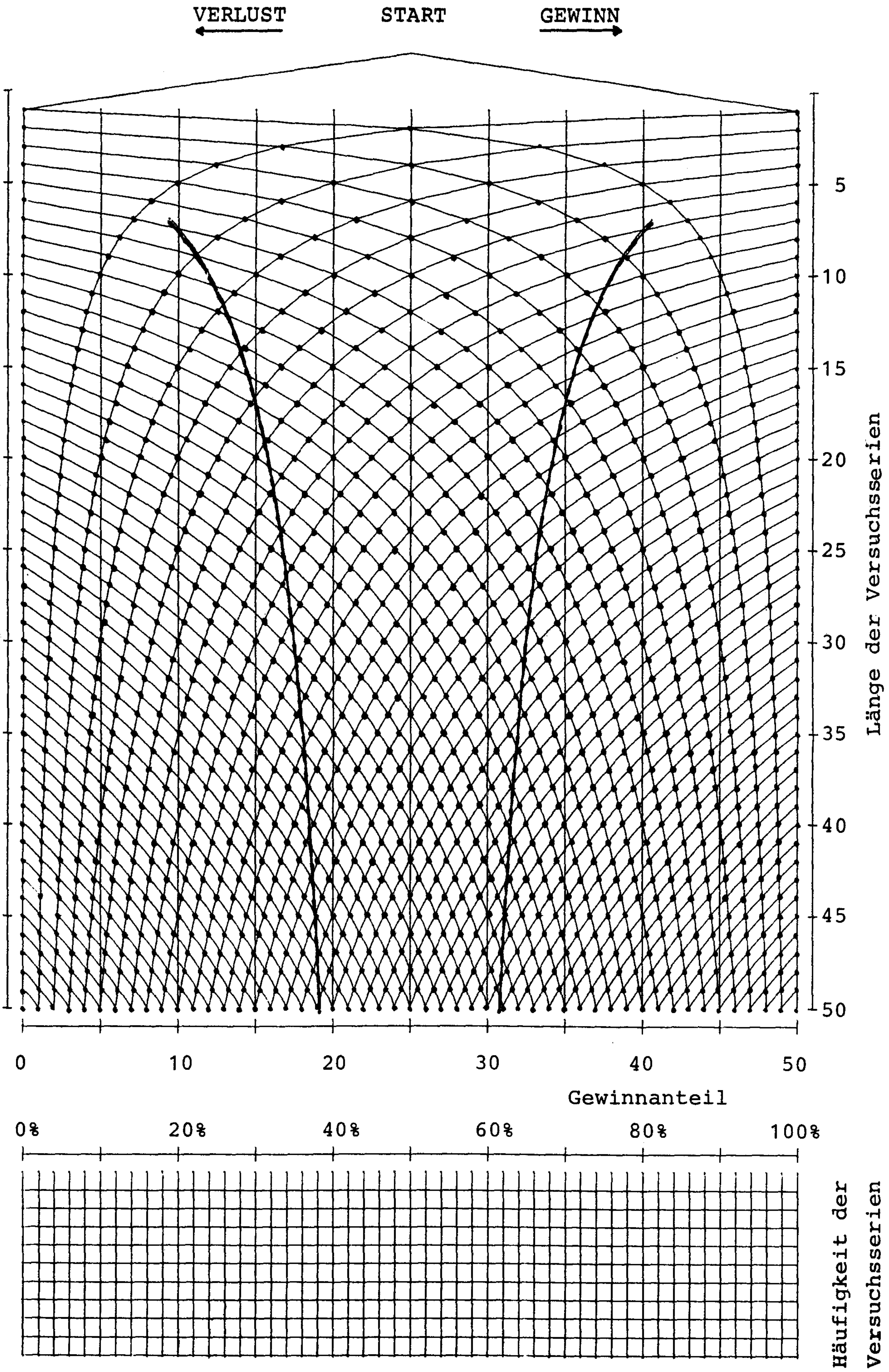

Bild 2 : Aufzeichnungsblatt mit 90%-Vertrauensintervall für p

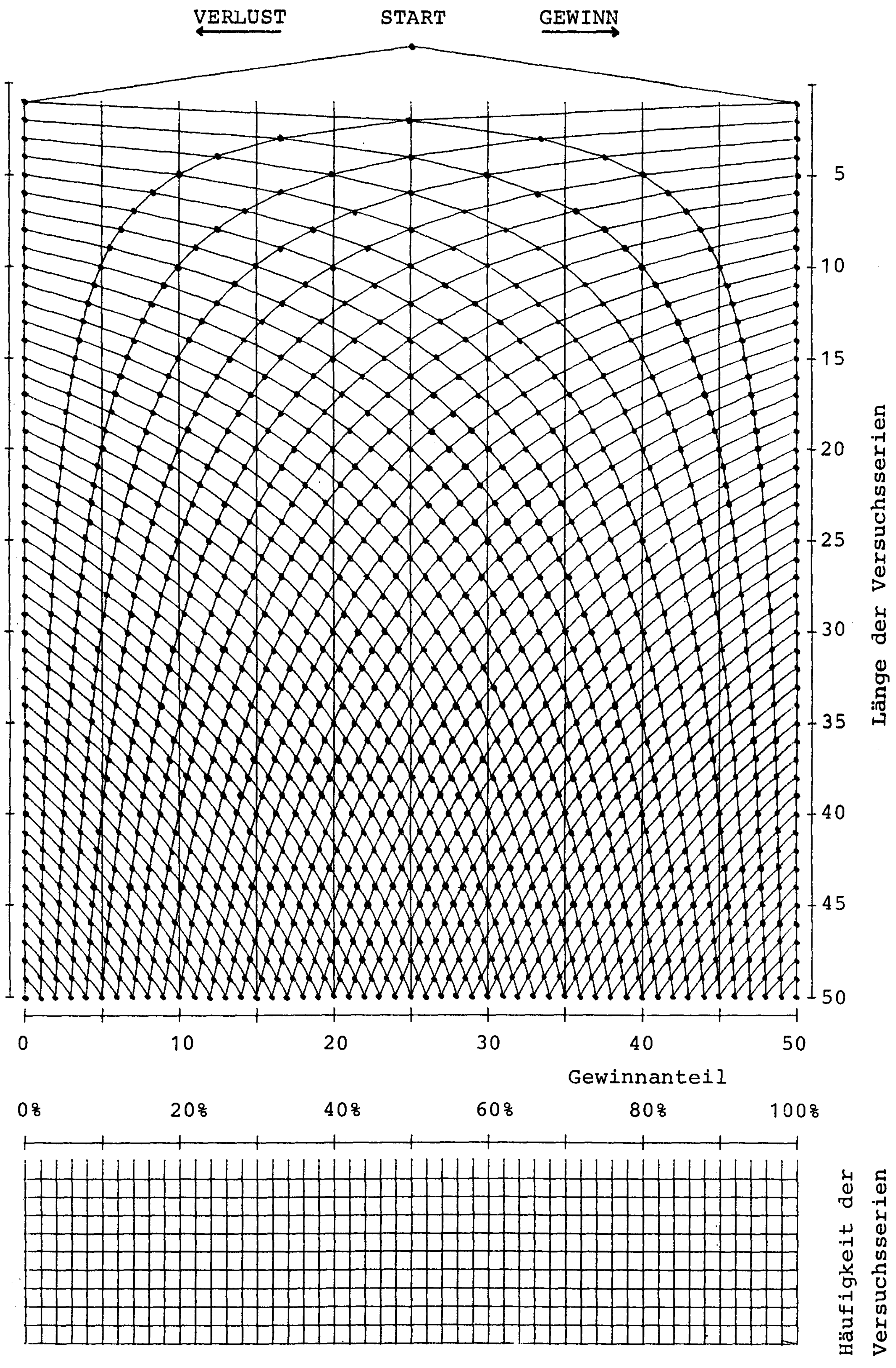

Bild 1 : Aufzeichnungsblatt