ZUR
THEORIE DER LINIENTAXIERUNG

VON

J. W. LINDEBERG

HELSINKI 1926

HELSINKI 1926,
DRUCKEREI DER FINNISCHEN LITERATURGESELLSCHAFT

Zur Theorie der Linientaxierung.

Von
J. W. LINDEBERG.

1. Im Jahre 1924 habe ich in Acta forestalia fennica 25 eine Arbeit über die Berechnung des Mittelfehlers des Resultates einer Linientaxierung publiziert, worin die früher zu diesem Zwecke angewandten Methoden kritisiert werden und eine neue Formel vorgelegt wird. Im Anschluss an eine denselben Gegenstand behandelnde Arbeit des Herrn A. Langsaeter [1] will ich auf einige Punkte der Frage zurückkommen.

Ich habe u.a. dagegen Bedenken erhoben, dass bei der Berechnung des Mittelfehlers die Linien in eine so kleine Anzahl von Gruppen wie zehn zusammengefasst werden. Herr Langsaeter berührt diese Frage im Vorbeigehen, indem er an Beispielen zeigt, dass auch Mittelfehler, die aus einer sehr kleinen Anzahl von Gruppen berechnet worden sind, mit unbedeutenden Fehlern behaftet sein können. Da die Frage nach der Zuverlässigkeit solcher Mittelfehler sehr wichtig zu sein scheint, will ich etwas näher auf sie eingehen.

Es liege eine Anzahl von Beobachtungen $x_1, x_2, \ldots, x_n$ einer Grösse X vor, und es werde angenommen, dass der Mittelwert

$$M = \frac{1}{n} \sum x_\mu \tag{1}$$

und die Streuung

$$\sigma = \sqrt{\frac{1}{n} \sum (x_\mu - M)^2} \tag{2}$$

berechnet worden sind. Wenn angenommen werden kann, dass an den Beobachtungen keine systematischen Fehler haften, macht man sich in der Praxis eine Vorstellung von der Genauigkeit, womit M die Grösse X darstellt, indem man die Wahrscheinlichkeiten der Ungleichungen

[1] Om beregning av middelfeilen ved regelmässige linjetakseringer, Meddelelser fra det norske skogforsøksvesen, Oslo 1926.

$$|M - X| > 2\frac{\sigma}{\sqrt{n-1}} \quad \text{und} \quad |M - X| > 3\frac{\sigma}{\sqrt{n-1}}$$

heranzieht. Falls $\sigma : \sqrt{n-1}$ der richtige Mittelfehler von M wäre, hätten diese Wahrscheinlichkeiten die Werte 0.05 und 0.003, und wenn n gross ist, geben deshalb diese Zahlen eine gute Auffassung von den in Frage stehenden Wahrscheinlichkeiten. Für kleine Werte von n liegen aber die Sachen anders. Wir stellen uns als erste Aufgabe, die präzisen Werte dieser Wahrscheinlichkeiten zu berechnen, wenn n klein ist.

2. Indem wir der Einfachheit halber X gleich Null annehmen und mit α eine positive Zahl grösser als 1 bezeichnen, fragen wir allgemein nach der Wahrscheinlichkeit $P_{n,\alpha}$, dass

$$|M| > \alpha\frac{\sigma}{\sqrt{n-1}}$$

oder also

$$M^2 - \alpha^2\frac{\sigma^2}{n-1} > 0.$$

Führen wir hier für M den Ausdruck (1) ein und für σ den Ausdruck, den man erhält, wenn in (2) M durch (1) ersetzt wird, und dividieren wir mit $\frac{1-\alpha^2}{n^2}$, so ergibt sich die Ungleichung

$$\sum x_\mu^2 - 2\frac{\alpha^2 + n - 1}{(\alpha^2-1)(n-1)}\sum x_r x_s < 0. \tag{3}$$

$P_{n,\alpha}$ ist also die Wahrscheinlichkeit, dass die Grössen $x_1, x_2, \ldots, x_n$ so ausfallen, dass sie der Ungleichung (3) genügen.

Indem wir annehmen, dass alle Beobachtungen demselben normalen Wahrscheinlichkeitsgesetz unterliegen, sei $ae^{-h^2x^2}$ die diesem Gesetze zugehörige Wahrscheinlichkeitsfunktion. Die Wahrscheinlichkeit, dass die n Beobachtungen zwischen die respektiven Grenzen

$$x_1 \text{———} x_1 + dx_1,\ x_2 \text{———} x_2 + dx_2,\ \ldots,\ x_n \text{———} x_n + dx_n$$

fallen, ist dann gleich

$$a^n e^{-h^2\sum x_\mu^2} dx_1 dx_2 \ldots dx_n,$$

und $P_{n,\alpha}$ ist also gleich dem n-fachen Integral dieses Ausdruckes, über das durch die Ungleichung (3) definierte Gebiet des n-dimensionalen $x_1, x_2, \ldots, x_n$-Raumes erstreckt. Da der Integrand nur von dem Abstand vom Nullpunkt abhängt, findet man ohne Schwierigkeit, dass der Wert

dieses Integrals gleich dem Quotienten zwischen dem Volumen des durch die Ungleichung (3) definierten Doppelsektors der n-dimensionalen Einheitskugel $\Sigma x_\mu^2 = 1$ und dem Volumen dieser Kugel selber ist. Mit Benutzung bekannter Formeln ergibt sich aus dieser Bemerkung

$$P_{n,\alpha} = \frac{\int_0^\Theta \sin^{n-2}\varphi\, d\varphi}{\int_0^{\frac{\pi}{2}} \sin^{n-2}\varphi\, d\varphi},$$

wo Θ den durch die Gleichung

$$\cos\Theta = \frac{\alpha}{\sqrt{\alpha^2 + n - 1}}$$

definierten positiven spitzen Winkel bedeutet. Vermittels dieser Formeln berechnet sich für die wichtigsten Werte $P_{n,\alpha}$ die folgende Tafel:

n	$P_{n,2}$	$P_{n,3}$
3	0.184	0.095
4	0.139	0.058
5	0.116	0.040
6	0.102	0.030
7	0.092	0.023
8	0.086	0.020
9	0.081	0.017
10	0.076	0.015

Der Wert von $P_{10,3}$ ist, wie ersichtlich, ungefähr fünfmal so gross wie der Wert von $P_{\infty,3}$, wenn n gross ist. Die Zahl 0.015 ist aber an sich noch klein und zeigt, dass die in meiner vorigen Arbeit ausgesprochenen Bedenken gegen die Anwendung eines auf Grund von nur zehn Beobachtungen berechneten Mittelfehlers übertrieben waren. Es scheint sogar, als ob die in den landwirtschaftlichen Untersuchungen oft auf Grund einer noch kleineren Anzahl von Beobachtungen berechneten Mittelfehler nicht ganz bedeutungslos wären.

Die Zahlen unserer Tafel setzen voraus, dass der in Frage stehende Beobachtungsfehler dem Gauss'schen Gesetze folgt. Dies kann nun immer bezweifelt werden. Eine Abweichung von diesem Gesetze kann aber bald vergrössernd, bald verkleinernd auf die fraglichen Zahlen einwirken, und da man im allgemeinen keine Basis für die Beurteilung der Art der Abweichung besitzt, scheint es nicht möglich zu sein, eine bessere Vorstellung von den gesuchten Wahrscheinlichkeiten zu erhalten als die oben gegebene.

3. Ein zweiter Punkt, worauf ich eingehen will, betrifft die von mir vorgeschlagene Formel zur Berechnung des Mittelfehlers des Taxierungsresultates. Es seien $l_1, l_2, \ldots, l_n$ die Längen der Linien und $p_1, p_2, \ldots, p_n$ die beobachteten Prozente. Ferner möge $L = \Sigma\, l_\mu$ gesetzt werden. Für die Berechnung des Mittelfehlers des Prozentes

$$P = \sum \frac{l_\mu}{L} p_\mu$$

habe ich die Formel

$$\varepsilon(P) = \sqrt{\frac{n}{2(n-1)} \sum \left(\frac{l_\mu + l_{\mu+1}}{2L}\right)^2 \left(p_\mu - p_{\mu+1}\right)^2} \tag{4}$$

vorgeschlagen. Herr Langsaeter hat nun in einer Reihe von zehnprozentigen Waldbodentaxierungen die durch diese Formel geleisteten Mittelfehler mit den wirklichen Fehlern verglichen und dabei gefunden, dass die vermittels der Formel erhaltenen Werte zu gross sind. Er stellt deshalb eine neue Formel auf, die statt der ersten Differenzen der Zahlen p_μ die zweiten Differenzen benutzt. Seine Formel lautet mit meinen Bezeichnungen

$$\varepsilon(P) = \sqrt{\frac{n}{1{,}5\cdot(n-2)} \sum_{2}^{n-1} \frac{(l_{\mu-1} + l_{\mu+1})\, l_\mu}{2L^2} \left(\frac{p_{\mu-1} + p_{\mu+1}}{2} - p_\mu\right)^2} \tag{5}$$

Bei der Prüfung dieser Formel an demselben Material, das bei der Prüfung der Formel (4) zur Anwendung kam, hat sich aber gezeigt, dass die neue Formel zwar bessere Werte als (4), aber noch keine befriedigenden Werte gibt. Herr Langsaeter gibt deshalb noch ein Verfahren an, die vermittels (5) erhaltenen Werte noch weiter zu korrigieren.

Ich möchte hierzu folgendes bemerken. Die Formel (4) ist eine Majorantenformel, die eher zu grosse als zu kleine Werte gibt, und da sich herausgestellt hat, dass sie bei hochprozentigen Taxierungen des Waldbodens kleiner Gebiete nicht akzeptable Mittelfehler leistet, müssen

natürlich in solchen Fällen andere Methoden angewandt werden. Im allgemeinen dürfte aber der in (4) enthaltene Sicherheitszuschlag sehr wohl vernachlässigt werden können.

Der Gedanke, dass bei Benutzung der zweiten Differenzen der berechnete Mittelfehler herabgedrückt werden könne, ist ohne Zweifel richtig. Es ist mir aber nicht verständlich, warum Herr Langsaeter in (5) als Koeffizient den Ausdruck $\frac{(l_{\mu-1} + l_{\mu+1})\, l_\mu}{2L^2}$ anstatt $\frac{l_\mu^2}{L^2}$ benutzt. Wenn er in (4) $\left(\frac{l_\mu + l_{\mu+1}}{2L}\right)^2$ durch $\frac{l_\mu l_{\mu+1}}{L^2}$ ersetzt, kommt doch eine kleine formale Vereinfachung zustande, und die Rechnung mit der so erhaltenen Formel kann sich bei gewisser Rechnungsweise etwas einfacher gestalten als bei Benutzung der richtigen Koeffizienten. Diese Gesichtspunkte können aber nicht zu Gunsten der Formel (5) herangezogen werden. Irgendein Missverständnis muss vorliegen, und ich mache deshalb darauf aufmerksam, dass, falls z.B. alle zweiten Linien sehr kurz sind, die Formel (5) einen entsprechend sehr kleinen Mittelfehler leistet, was nicht richtig sein kann.

Dem Umstand, dass noch (5) in gewissen Fällen zu grosse Mittelfehler leistet, wird nicht damit abgeholfen, dass $\frac{l_\mu^2}{L^2}$ als Gewichtskoeffizient eingeführt wird. Mir scheint aber, dass man, bevor zu anderen Massnahmen geschritten wird, den schon mit der Aufstellung der Formel (5) eingeschlagenen Weg noch weiter zu verfolgen hat.

Indem wir die Bezeichnungen

$$\begin{aligned}
d^{(1)}_{\mu,\mu+1} &= p_{\mu+1} - p_\mu \\
d^{(2)}_{\mu,\mu+2} &= d^{(1)}_{\mu+1,\mu+2} - d^{(1)}_{\mu,\mu+1} \\
d^{(3)}_{\mu,\mu+3} &= d^{(2)}_{\mu+1,\mu+3} - d^{(2)}_{\mu,\mu+2} \\
d^{(4)}_{\mu,\mu+4} &= d^{(3)}_{\mu+1,\mu+4} - d^{(3)}_{\mu,\mu+3}
\end{aligned}$$

einführen, denken wir uns die folgende Differenzentafel aufgestellt:

$$\begin{array}{ccccc}
p_1 & & & & \\
 & d^{(1)}_{1,2} & & & \\
p_2 & & d^{(2)}_{1,3} & & \\
 & d^{(1)}_{2,3} & & d^{(3)}_{1,4} & \\
p_3 & & d^{(2)}_{2,4} & & d^{(4)}_{1,5} \\
 & d^{(1)}_{3,4} & & d^{(3)}_{2,5} & \\
p_4 & & d^{(2)}_{3,5} & & \cdot \\
 & d^{(1)}_{4,5} & & \cdot & \\
p_5 & & \cdot & & \cdot \\
\cdot & \cdot & & \cdot & \\
\cdot & \cdot & \cdot & & \cdot \\
 & \cdot & & \cdot & \\
\cdot & \cdot & \cdot & & d^{(4)}_{n-4,n} \\
 & & & d^{(3)}_{n-3,n} & \\
\cdot & & d^{(2)}_{n-2,n} & & \\
 & d^{(1)}_{n-1,n} & & & \\
p_n & & & &
\end{array}$$

Wenn man die Berechnung des Mittelfehlers auf die zweiten Differenzen basieren will, so geschieht dies am besten durch die Formel

$$\varepsilon(P) = \sqrt{\frac{n}{6(n-2)} \sum_{\mu=2}^{n-1} \left(\frac{l_\mu d^{(2)}_{\mu-1,\mu+1}}{L}\right)^2}. \tag{6}$$

Daneben möchte ich als praktisch brauchbare Formeln, von welchen erwartet werden darf, dass sie den Wert des Mittelfehlers noch weiter herabdrücken, die Formeln

$$\varepsilon(P) = \sqrt{\frac{n}{20(n-3)} \sum_{\mu=2}^{n-2} \left(\frac{l_\mu + l_{\mu+1}}{2L} \cdot d^{(3)}_{\mu-1,\mu+2}\right)^2} \tag{7}$$

und

$$\varepsilon(P) = \sqrt{\frac{n}{70(n-4)} \sum_{\mu=3}^{n-2} \left(\frac{l_\mu d^{(4)}_{\mu-2,\mu+2}}{L}\right)^2} \tag{8}$$

in Vorschlag bringen.

Um z.B. die Formel (8) herzuleiten, hat man folgendes zu beachten. Es ist erstens

$$d^{(4)}_{\mu-2,\mu+2} = p_{\mu-2} - 4p_{\mu-1} + 6p_\mu - 4p_{\mu+1} + p_{\mu+2}.$$

Wenn p'_μ den wahrscheinlichen Wert und σ_μ den Mittelfehler von p_μ bezeichnet, und die Grössen p_μ voneinander unabhängig vorausgesetzt werden, findet man weiter, dass der wahrscheinliche Wert der Grösse $(d^{(4)}_{\mu-2,\mu+2})^2$ gleich

$$(\sigma^2_{\mu-2} + 16\sigma^2_{\mu-1} + 36\sigma^2_\mu + 16\sigma^2_{\mu+1} + \sigma^2_{\mu+2}) +$$
$$+ (p'_{\mu-2} - 4p'_{\mu-1} + 6p'_\mu - 4p'_{\mu+1} + p'_{\mu+2})^2$$

ist. Ist das zweite Glied hier für jedes μ klein, so wird also der wahrscheinliche Wert des Ausdruckes unter der Quadratwurzel in (8) annäherungsweise gleich $\sum \left(\frac{l_\mu}{L}\right)^2 \sigma^2_\mu$. Zwar ist hierbei zu bemerken, dass, auch wenn das genannte Glied immer genau gleich Null wäre, die Gleichheit nicht exakt bestände. Wie die Linienlängen auch variieren, ist aber kein Grund vorhanden, eine positive Abweichung eher als eine negative, oder umgekehrt, a priori zu erwarten, und ich nenne deshalb die Formel (8) theoretisch richtig. Unsere oben betreffs der Formel (5) gemachte Bemerkung besagt, dass diese Formel nicht theoretisch richtig ist.

4. Bei der Anwendung der Wahrscheinlichkeitsrechnung auf praktische Fragen ist es meistens eine sehr schwierige Sache, die Berechtigung der Anwendung darzulegen, und dies scheint leider davon herzurühren, dass es überhaupt keinen unanfechtbaren Grund für solche Anwendungen gibt. Jedes praktische Resultat hat deshalb in gewissem Grade den Charakter einer subjektiven Schätzung. Die Formeln (4), (6), (7) und (8) sind nun in der Hinsicht ziemlich befriedigend, weil sie keine Voraussetzungen spezieller Art brauchen. Wenn Herr Langsaeter auf Seite 33 seiner Arbeit sagt, dass ich in meiner Ableitung der Formel (4) stillschweigend eine von ihm näher angegebene spezielle Voraussetzung mache, muss dies auf einem Missverständnis beruhen.