STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

S 47 (M 16)

Normaliteitstoetsen van Geary en Pearson

MATHEMATISCH CENTRUM,
2de Boerhaavestr. 49,
Amsterdam-O.

Statistische Afdeling
S 47 (M 16)

Normaliteitstoetsen van Geary en Pearson[1].

Om de hypothese H_0 te toetsen, dat een reeks waarnemingen $x_1, x_2, \ldots x_n$ beschouwd kan worden als een steekproef uit een normale verdeling, d.i. een verdeling met verdelingsdichtheid

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

zijn door Geary en Pearson de verdelingen, onder hypothese H_0, berekend van enkele statistische grootheden, welke betrekking hebben op de vorm van de kromme $f(x)$.

De eerste grootheid is de volgende:

$$(1) \quad \underline{a} = \frac{\sum_{i=1}^{n} |\underline{x}_i - \underline{\bar{x}}|}{\sqrt{n \sum_{i=1}^{n} (\underline{x}_i - \underline{\bar{x}})^2}} \qquad \text{waarin} \quad \underline{\bar{x}} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{is.}$$

Deze grootheid a is een maat voor de "kurtosis" ("platheid" of "slankheid" der kromme $f(x)$). In figuur 1 zijn 3 verdelingsdichtheden $f(x)$, $f_1(x)$ en $f_2(x)$ getekend waarvan $f(x)$ een normale is, terwijl $f_1(x)$ te slank is (en te dik in de staarten), en $f_2(x)$ te plat (en te dun in de staarten). Pearson noemt $f_1(x)$ "leptocurtic" en $f_2(x)$ "platycurtic".

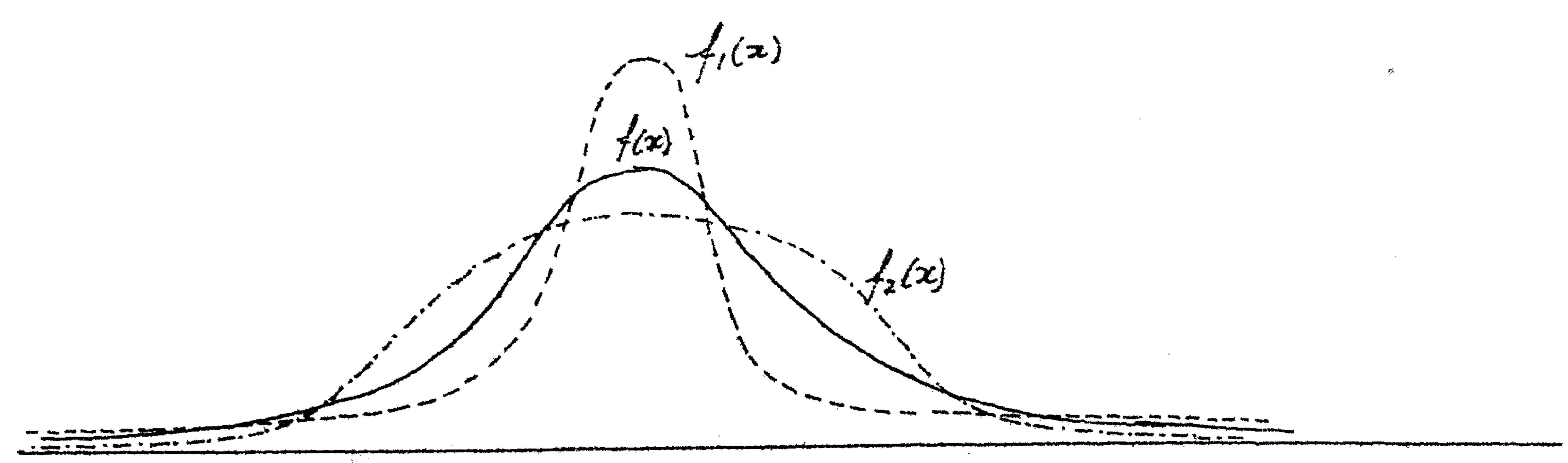

fig. 1

De kritieke zône voor $\underline{a}$ bestaat uit grote en kleine waarden van a.

---

1) Dit memorandum is slechts bedoeld ter oriëntering en streeft niet naar volledigheid of volledige exactheid.

Als tweede toetsingsgrootheid wordt gebruikt:

(2) $$\sqrt{b_1} = \sqrt{n}\,\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^{\frac{3}{2}}}$$

Deze grootheid is een maat voor de "scheefheid" (Engels: "skewness") van de verdeling. In figuur 2 zijn 3 verdelingsdichtheden getekend, waarvan $f(x)$ symmetrisch is, terwijl $f_1(x)$ "positief scheef" (omdat de "dikke staart" rechts ligt) en $f_2(x)$ "negatief scheef" is.

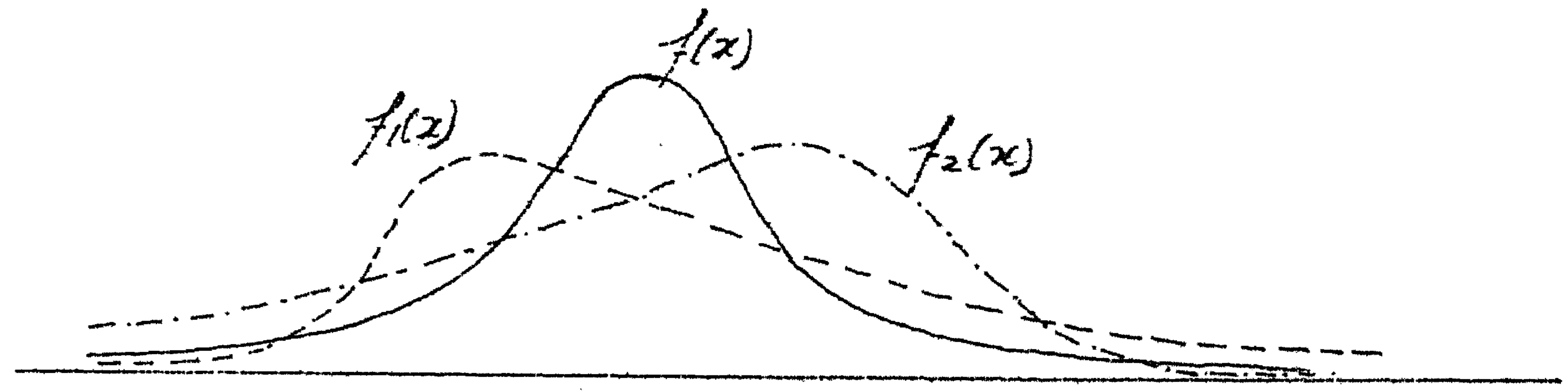

fig. 2

De kritieke zône voor de grootheid $\sqrt{b_1}$, bestaat uit grote en kleine waarden van $\sqrt{b_1}$, (daarbij wordt gewoonlijk ook voor negatieve waarden de notatie $\sqrt{b_1}$ gebruikt).

De derde toetsingsgrootheid

(3) $$b_2 = \frac{n\sum_{i=1}^{n}(x_i-\bar{x})^4}{\left\{\sum_{i=1}^{n}(x_i-\bar{x})^2\right\}^2}$$

is weer een maat voor de "kurtosis" der kromme en wordt alleen voor zeer grote steekproeven berekend. Voor kleinere steekproeven is de maat a voldoende.

De kritieke zône bestaat uit grote en kleine waarden van b_2.

<u>Nomogrammen der 3 grootheden:</u>

1e. Voor de grootheid a zijn onder de hypothese H_0 nomogrammen en tabellen berekend voor de onbetrouwbaarheidsdrempels 0,01; 0,05 en 0,10, waarbij $10 \leqq n \leqq 1000$.

2e. Voor de grootheid $\sqrt{b_2}$ zijn nomogrammen berekend voor de onbetrouwbaarheidsdrempels 0,01 en 0,05, $25 \leqq n \leqq 1000$.

3e. Voor de grootheid $\sqrt{b_2}$ zijn nomogrammen berekend voor de onbetrouwbaarheidsdrempels 0,01 en 0,05. Hierbij is $100 \leqq n \leqq 1000$.

Literatuur:

R.C.Geary, Moments of the ratio of the mean deviation to the standard deviation for normal samples, Biometrika 28 (1936) p.295.

R.C.Geary and E.S.Pearson, Tests of Normality, Biometrika Office, London 1938.

Bovenstaande nomogrammen en tabellen komen in deze beide publicaties voor.
Alle daarin genoteerde overschrijdingskansen zijn éénzijdig.