STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

ZW 1954 - 005

Voordracht in de serie Actualiteiten

H.J. Prins

zaterdag 27 februari 1954

Toetsingsmethoden en toepassingen van de Poisson-verdeling

1954

Voordracht door de Heer H.J.Prins in de serie Actualiteiten op Zaterdag 27 Februari 1954.

Toetsingsmethoden en toepassingen van de Poisson-verdeling.

Verdelingsfuncties.

In de statistiek worden experimenten beschouwd, waarbij onder uiterlijk dezelfde omstandigheden verschillende uitkomsten $\underline{x}$ worden gevonden voor een meting of telling. Bij herhaling van eenzelfde experiment is het dus mogelijk dat verschillende waarden van $\underline{x}$ gevonden worden. Wij noemen zo'n variabele een stochastische variabele en geven het stochastische karakter van een grootheid door onderstreping aan.

Als wiskundig model voor een dergelijk verschijnsel wordt de waarschijnlijkheidsverdeling gebruikt, d.i. een functie die aangeeft met welke kansen de stochastische variabele $\underline{x}$ bepaalde waarden aanneemt of in bepaalde intervallen van waarden ligt. Bij tellingen kunnen wij de kansen beschouwen, waarvoor $\underline{x}$ bepaalde waarden aanneemt, bij metingen zijn deze kansen 0 en slechts de kans dat $\underline{x}$ in een bepaald interval van waarden ligt wordt beschouwd. In het eerste geval spreken wij van een discrete verdeling, in het laatste geval van een continue verdeling. In fig. 1 zijn beide typen weergegeven.

a. continu   b. discreet

Fig. 1.

In fig. 1a geeft het gearceerde oppervlak de kans aan dat $\underline{x}$ in het interval (x_1, x_2) ligt; in fig. 1b geeft de hoogte van ieder staafje de kans aan dat die waarde door $\underline{x}$ aangenomen wordt bij een experiment. De kansen worden gegeven als fracties. Kans 1 wil zeggen volledige zekerheid. Dus het totale oppervlak onder de kromme in 1a moet 1 zijn en de som van de lengten van de staafjes in 1b eveneens. We noteren deze kansen als volgt: voor het eerste geval

$$P\{x_1 < \underline{x} < x_2\} \overset{\text{def}}{=} \int_{x_1}^{x_2} f(x).dx$$

voor het laatste geval

$$P\{\underline{x} = i\} \qquad i = 1, \ldots, 5$$

en hebben dus tevens de betrekkingen

$$\mathcal{P}\{-\infty < \underline{x} < +\infty\} = \int_{-\infty}^{+\infty} f(x).\,dx = 1$$

en

$$\sum_{1}^{5} i \;\; \mathcal{P}\{\underline{x} = i\} = 1$$

Een voorbeeld van een meting met een stochastisch karakter is de bepaling van de elementaire ladingseenheid volgens Millikan.
Een voorbeeld van een dergelijke telling is het aantal bacteriën in een bloedmonster of het aantal verkeersongelukken in een bepaalde periode of het aantal keren kruis bij het opgooien van een muntstuk.

We willen ons hier beperken tot een bepaald type discrete verdeling: de Poisson-verdeling en houden ons dus bezig met experimenten, waarbij tellingen verricht worden.

Wanneer wij veel waarnemingen doen van een bepaalde discrete stochastische grootheid $\underline{x}$, dan zal de fractie van de gevallen waarin i optreedt zeer dicht bij de fractie liggen, die de kans op i aangeeft. De frequentiecurve voor een groot aantal waarnemingen zal dus steeds dichter naderen tot de wh-verdeling[1)] van $\underline{x}$.

In een bepaald geval van toepassing zal in het algemeen niet van te voren bekend zijn wat de wh-verdeling is; we willen immers uit de uitkomsten van het experiment conclusies trekken over de optredende grootheden (b.v. het aantal bacteriën in het bloed, waaruit een monster genomen is). Wanneer de wh-verdeling bekend is, zijn deze grootheden van te voren al bekend. Naar de aard van het probleem wordt dan ook van te voren een bepaald type wh-verdeling gekozen. Dit type is een functionele uitdrukking, die nog enkele onbekende grootheden bevat (de parameters).

Nu zijn er 3 vragen die zich bij dergelijke experimenten voordoen.
1. Is het gekozen type wh-verdeling juist? (we kunnen dit toetsen met als alternatieven andere modellen).
2. Wat is een goede schatting voor de parameters van de wh-verdeling? Of wat is een goede schatting voor de grootheid, ter bepaling waarvan het experiment is opgezet (deze grootheid is een functie van parameters), b.v. sterkte van een radioactief preparaat.
3. Komen verschillende waarnemingen uit éénzelfde wh-verdeling? (het vergelijken van experimenten).
Met dit laatste toetsingsprobleem zullen we ons in hoofdzaak bezighouden.

----

1) wh is per afkorting waarschijnlijkheid.

De Poisson-verdeling (van S. D. Poisson 1837) is een discrete verdeling, die aan alle natuurlijke aantallen een bepaalde kans toevoegt. De uitdrukking bevat één parameter. Dit is dus geen experimenteel realiseerbaar model, want bij ieder concreet probleem komen slechts begrensde aantallen voor. Het is dan ook een benadering die vaak goed voldoet en tot eenvoudige berekeningen leidt. Hoge aantallen hebben bij de Poisson-verdeling kleine kansen; immers moet gelden:

$$\sum_{0}^{\infty} {}^{i}\, P\{\underline{x} = i\} = 1$$

De functionele uitdrukking is

$$P\{\underline{x} = x\} = \frac{e^{-\mu}\mu^{x}}{x!} \qquad x \in \{0, 1, 2, \ldots\}$$

$\sum_{0}^{\infty} {}^{x}\, P\{\underline{x} = x\} = 1$, want $e^{\mu} = \sum_{0}^{\infty} {}^{x}\, \frac{\mu^{x}}{x!}$. De parameterwaarde μ hoeft niet noodzakelijk geheel te zijn.

<u>Voorbeelden</u> van een variabele met een Poisson-verdeling zijn: het aantal α-deeltjes uitgezonden door een radioactief preparaat in een bepaalde periode; het aantal aangevraagde telefoongesprekken binnen een bepaalde periode in een telefooncentrale, tellingen van het aantal bacteriën in bloed-monsters, e.a. Het klassieke voorbeeld is het aantal ongelukken van huzaren met paarden in het Duitse leger.

Voor het geval van radioactiviteit zullen wij in tabel I laten zien, dat het Poisson-model aanvaardbaar is; de waargenomen frequenties N_x wijken weinig af van de verwachte $N . P\{\underline{x} = x\}$.

x	N_x	$N . P\{\underline{x} = x\}$
0	57	54
1	203	211
2	383	407
3	525	525
4	532	508
5	408	394
6	273	254
7	139	140
8	45	68
9	27	29
10	16	17
totaal	2608	2608

<u>Tabel 1</u>

Wij kunnen statistisch toetsen of de afwijkingen die hier voorkomen toelaatbaar zijn voor dit aantal en vinden dan een overschrijdingskans van 0,17[2]. We concluderen dat het model goed is.

De uitdrukking voor de Poisson-verdeling is een benadering, zoals al opgemerkt werd. Het is ook langs wiskundige weg een benadering van een ander type verdelingsfunctie: de binomiale verdeling:

$$P\{\underline{x}=x\} = \binom{n}{x} p^x q^{n-x}; \qquad p+q=1$$

ook hier is

$$\sum_0^n{}^x P\{\underline{x}=x\} = \sum_0^n{}^x \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1$$

De binomiale verdeling geeft de kans aan dat onder n waarnemingen, die slechts 2 waarden met bepaalde kansen p en q kunnen aannemen, x waarnemingen voorkomen, die de kans p hebben.

Voorbeelden van de binomiaal verdeelde grootheid zijn: het aantal malen dat kruis optreedt bij n maal gooien met een munt, het aantal pareloesters, dat een parel bevat. Voor de binomiale verdeling geldt nu

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\mu}} \binom{n}{x} p^x q^{n-x} = \frac{e^{-\mu}\mu^x}{x!}$$

Het bewijs verloopt eenvoudig met behulp van het volgende aanschouwelijke model.

Een grote hoeveelheid deeltjes, N, bevindt zich in de ruimte V. Ieder deeltje heeft gelijke waarschijnlijkheid in ieder volume-element voor te komen. Kies nu een willekeurig volume-element D, dan is de kans dat een deeltje in D voorkomt D/V en de kans, dat er x in voorkomen

(1)
$$P\{\underline{x}=x\} = \binom{N}{x} \left(\frac{D}{V}\right)^x \left(\frac{V-D}{V}\right)^{N-x}$$

Dit is juist de binomiale verdeling. Passen we de limietovergang toe, dan moet dus $N\to\infty$, $V\to\infty$, terwijl $\frac{N}{V}\to d$. We vinden

$$\lim_{\substack{N\to\infty \\ V\to\infty \\ \frac{N}{V}\to d}} P\{\underline{x}=x\} = \lim_{\substack{N\to\infty \\ V\to\infty \\ N/V\to d}} \frac{(1-\frac{1}{N})\cdots(1-\frac{x-1}{N})(Dd)^x(1-\frac{Dd}{N})^{N-x}}{x!}$$

$$= \frac{1.(Dd)^x}{x!}\, e^{-Dd}$$

---

2) De overschrijdingskans geeft aan de kans op de gevonden of nog grotere afwijkingen van het gekozen model; is deze kans er klein (kleiner dan b.v. 0,01), dan concluderen we, dat het model niet juist is. Hieronder zal dit begrip nog nader worden omschreven.

De Poisson-verdeling wordt dus onder deze voorwaarden als benadering voor een binomiale verdeling gebruikt: kleine kans p en groot aantal n, of in het gebruikte model wanneer een groot aantal N deeltjes over een zeer groot volume V zijn verdeeld wordt de kans op x deeltjes in het volume-element D gegeven door de Poisson-verdeling.

Er is nog een andere wiskundige afleiding voor de uitdrukking

$$\frac{e^{-\mu}\mu^{x}}{x!}$$

Deze maakt gebruik van de theorie der stochastische processen en maakt een aantal andere gevallen van toepassing aanvaardbaar. We zullen hier niet verder op in gaan.

Toetsingsproblemen.

We zullen ons hier bezig houden met de toetsingsproblemen, daar de schattingsproblemen wiskundig eenvoudig en niet interessant zijn.

De toetsingsproblemen houden in het toetsen van bepaalde lineaire hypothesen betreffende de parameters van verschillende Poisson-verdelingen. Dit betekent dat we willen weten of b.v. een aantal waarnemingen uit éénzelfde Poisson-verdeling komt of dat ~~het~~ deze uit Poisson-verdelingen komen met verschillende parameterwaarden.

Een practisch voorbeeld is het volgende: men wil weten of verschillende telefooncentrales even druk bezet zijn of men wil weten of het aantal verkeersongelukken over een bepaalde periode constant is gebleven.

Wiskundig ziet het probleem er als volgt uit. Gegeven n onderling onafhankelijke waarnemingen $\underline{x}_1, \ldots, \underline{x}_n$ uit Poisson-verdelingen met parameters resp. $\mu_1, \ldots, \mu_n$. We willen weten of $\mu_1 = \mu_2 = \ldots = \mu_n$; wij noemen dit de nulhypothese H_0.

De hypothese is een lineaire vorm, vandaar de naam lineaire hypothese. We willen nu dus H_0 toetsen met als alternatief de hypothesen, waarvoor de μ's onderling ongelijk zijn.

Om dit te bereiken kiezen wij een geschikte toetsingsgrootheid, die aan een aantal eigenschappen moet voldoen.

1. De toetsingsgrootheid moet alleen een functie zijn van de waarnemingen en is dus stochastisch en bezit dus ook een wh-verdeling (dat de toetsingsgrootheid niet de parameterwaarden mag bevatten is duidelijk, omdat deze van te voren onbekend zijn).
2. De toetsingsgrootheid moet zo gekozen zijn, dat wanneer de waarnemingen niet onder de nulhypothese gedaan zijn, maar onder een der beschouwde alternatieve hypothesen, de afwijkingen van de toetsingsgrootheid van zijn gemiddelde groot zal zijn.

Met een dergelijke toetsingsgrootheid hebben wen een middel om uit te maken of de hypothese H_0 "verworpen" kan worden. Immers onder de nulhypothese hebben de $\underline{x}$'n bepaalde verdelingsfuncties en dus ook de toetsingsgrootheid $\underline{z}$, onder een der alternatieve hypothesen H heeft de toetsingsgrootheid ook een (andere) verdelingsfunctie (zie fig. 2).

We bepalen nu een grenspunt s , zodat $P\{\underline{z} > s\} = \varepsilon$ onder H_0, waarbij ε klein gekozen wordt (ε heet de <u>onbetrouwbaarheid</u>), dus de kans op waarnemingen groter dan is onder H_0 zeer klein.

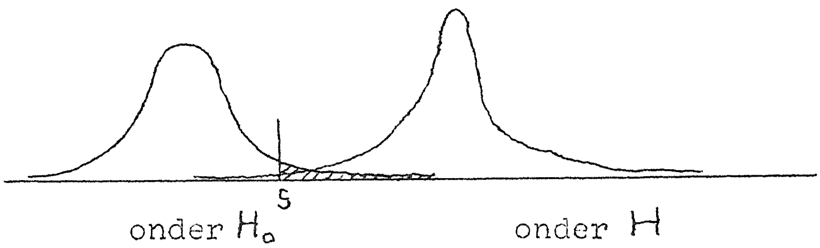

Fig. 2. Verdelingsfunctie van de toetsingsgrootheid onder twee hypothesen.

Vinden we toch een $\underline{z}$ in dit gebied, dan concluderen we, dat de nulhypothese niet juist was en verwerpen deze. In fig. 2 ziet men, dat het veel aannemelijker is, dat $\underline{z}$ gevonden wordt onder de hypothese H. , als $\underline{z} > s$. Is $\underline{z} > s$, dan noemen we de uitkomst <u>significant</u>. De kans $P\{\underline{z} > \cancel{s}\, z\}$ wordt de overschrijdingskans genoemd, als z een waargenomen waarde van $\underline{z}$ is.

Voor het beschouwde geval van de Poisson-verdeling willen we toetsen of voor n trekkingen $\underline{x}_1, \ldots, \underline{x}_n$ uit Poisson-verdelingen geldt.

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_n$$

De toetsingsgrootheid is hier

$$\underline{z}_{n,1} = \sum_1^n {}^i \frac{(\underline{x}_i - \bar{\underline{x}})^2}{\bar{\underline{x}}}$$

en heeft asymptotisch (voor grote μ en grote n) de bekende, getabelleerde wh-verdeling van χ^2 met $(n-1)$ vrijheidsgraden. De exacte verdeling van $\underline{z}_{n,1}$ is moeilijk te bepalen, echter is voor $\mu > 1$ en $n > 6$ de χ^2 -benadering al zeer goed.

Dat $\underline{z}_{n,1}$ gevoelig is voor afwijkingen van de nulhypothese is aan de uitdrukking te zien. Immers de meeste waarnemingen $\underline{x}_i$ vallen niet ver van het gemiddelde; zijn alle gemiddelden μ_i gelijk, dan is $|\underline{x}_i - \bar{\underline{x}}|$, de afwijking van het steekproefgemiddelde, in het algemeen kleiner dan wanneer de μ_i ongelijk zijn.

We zullen nu een algemenere nulhypothese beschouwen H_0^*:

$$\mu_1 : \mu_2 : \ldots : \mu_n = \nu_1 : \nu_2 : \ldots \nu_n$$

De ν's zijn hierin gegeven grootheden. Voor de eenvoud van de berekeningen zullen we onderstellen, dat $\sum_1^n {}^i \nu_i = 1$. De toetsingsgrootheid is

$$z'_{n,1} = \sum_1^n {}^i \frac{(\underline{x}_i - \nu_i . n . \bar{\underline{x}})^2}{\nu_i . n . \bar{\underline{x}}}$$

en heeft onder de nulhypothese weer asymptotisch voor grote μ's en grote n een χ^2-verdeling met $(n-1)$ vrijheidsgraden.

De alternatieve hypothesen zijn: "andere onderlingen verhoudingen der parameters". Voor $\nu_i = \frac{1}{n}$ $(i = 1, \ldots, n)$ vinden we de voorgaande toetsingsgrootheid als speciaal geval terug.

Hieronder zullen wij nog een toepassing van deze toets geven.

De nulhypothese en de vorm van de toetsingsgrootheden zijn in analogie met die der variantie-analyse, waar dergelijke nulhypothesen met dergelijke toetsingsgrootheden getoetst worden. De wh-verdeling der $\underline{x}$'n is dan echter niet de Poisson-verdeling, maar de verdeling van Gauss, ook de normale verdeling genoemd.

<u>Berekening van de verdeling van $\underline{z}'_{n,1}$</u>.

We beschouwen

$$P\{\underline{x}_i = x_i\} = \prod_{1}^{n} \frac{e^{-\nu_i\mu}(\nu_i\mu)^{x_i}}{x_i!}$$

hierin is $\mu = \sum_1^n \mu_i$.

Immers geldt steeds

$$P(A \text{ en } B) = P(A|B).P(B)$$

Als A en B onafhankelijke gebeurtenissen zijn, geldt echter

$$P(A \text{ en } B) = P(A).P(B)$$

Volgens $P(A \text{ en } B) = P(A|B).P(B)$ kunnen we schrijven

$$P\{\underline{x}_i = x_i \ (i=1,\ldots,n)\} = P\{\underline{x}_i = x_i \ (i=1,\ldots,n) | \textstyle\sum_1^n \underline{x}_i = X\}.P\{\textstyle\sum_1^n \underline{x}_i = X\}$$

of

$$P\{\underline{x}_i = x_i \ (i=1,\ldots,n)\} = \frac{\prod_1^n \frac{e^{-\nu_i.\mu}(\nu_i.\mu)^{x_i}}{x_i!}}{\frac{e^{-\mu}\mu^X}{X!}} \cdot \frac{e^{-\mu}\mu^X}{X!}$$

(2) $$= \frac{X!}{\prod_1^n x_i!}\prod_1^n (\nu_i)^{x_i} \cdot \frac{e^{-\mu}\mu^X}{X!}$$

asymptotisch voor $X \to \infty$ en $n \to \infty$ geldt

(3) $$\frac{X!}{\prod_1^n x_i!}\prod_1^n (\nu_i)^{x_i} \approx K.\, e^{-\frac{1}{2}\sum_1^n \frac{(x_i - \nu_i.n\bar{x})^2}{\nu_i.n\bar{x}}} \, dx_1 \ldots\ldots dx_n$$

waarin K een constante is.

Geldt (3), dan geeft het rechterlid van (3) voor $\underline{z}'_{n,1}$ na transformatie juist de verdelingsfunctie van χ^2 met $(n-1)$ vrijheidsgraden aan.

Dus

$$P\{\underline{x}_i = x_i \ (i = 1,\ldots,n)\} \approx P\{z < \underline{z}'_{n,1} < z + dz\} \cdot P\{\sum_1^n \underline{x}_i = X\}$$

$\underline{z}'_{n,1}$ en $\sum_1^n \underline{x}_i$ zijn dus onafhankelijk verdeeld en de verdeling van $\underline{z}'_{n,1}$ is als χ^2 met $(n-1)$ vrijheidsgraden.

Voorbeeld van toetsing met $\underline{z}'_{n,1}$.
Ongelukken van arbeiders in een fabriek.

De bedrijfsleiding van een fabriek wil onderzoeken of het aantal ongelukken van arbeiders over een lange periode, L, afgezien van toevallige fluctuaties, constant gebleven is.

Aangezien het aantal ongelukken per tijdseenheid steeds klein is in vergelijking met het totaal aantal arbeiders, is hier als benadering voor de binomiale verdeling de Poisson-verdeling beschouwd (de binomiale verdeling is een juist model voor dit geval, omdat we het aantal verongelukte arbeiders per n arbeiders beschouwen en onderstellen, dat de kans op een ongeluk voor een arbeider, p, constant is). De Poisson-verdeling heeft dus de parameterwaarde $\mu = n.p$. (De complicatie, dat bij een ongeluk meerdere arbeiders betrokken kunnen zijn, laten wij terwille van de eenvoud van het voorbeeld buiten beschouwing.)

Wanneer het totaal aantal arbeiders gedurende de periode L fluctueert, wat in het algemeen het geval zal zijn, dan verdelen we de periode L in m gelijke deelperioden $L_1, \ldots, L_m$, zodat voor iedere deelperiode het totaal aantal arbeiders constant is, resp. $n_1, \ldots, n_m$. Voor iedere deelperiode hebben we een waarneming x_i $(i = 1, \ldots, n)$ van het aantal verongelukte arbeiders. Volgens het bovenstaande mogen wij onderstellen, dat de waarneming x_i komt uit een Poisson-verdeling met parameter $\mu_i = n_i p$ (we onderstellen steeds, dat de kans p op een ongeluk voor een arbeider constant is voor alle deelperioden). Nu willen we de hypothese toetsen, dat deze kansen inderdaad gelijk zijn, dus dat voor de Poisson-verdelingen geldt:

$$H_0 : \quad \mu_1 : \mu_2 : \ldots : \mu_m = n_1 : n_2 : \ldots : n_m$$

Zorgen we er voor, dat $\sum_1^n \nu_i = 1$, dan vinden we $\nu_i = \dfrac{n_i}{\sum n_i}$.

We kunnen nu toetsen met $\underline{z}'_{n,1}$ of de nulhypothese verworpen kan worden.

De grootte en het aantal der deelperioden wordt bepaald door de fluctuatie van het aantal arbeiders, maar ook door het effect, dat men verwacht te vinden als de nulhypothese niet juist is. Veronderstel b.v., dat er een zekere periodiciteit voorkomt in het aantal ongelukken, doordat de arbeiders in bepaalde perioden slordiger werken. Wanneer de deelperioden groot zijn t.o.v. deze periode, dan zal men bij toetsing niets merken. Is echter de deelperiode ongeveer van de helft van de lengte van de periode, dan zal dit effect gemakkelijker leiden tot verwerping van de nulhypothese. In voorkomende gevallen is het dus zaak nauwkeurig te bepalen hoe de onderverdeling geschiedt. In tabel 2 is een numeriek geval gegeven.

deelperiode	L_1	L_2	L_3	L_4	L_5	L_6	L_7	L_8	L_9	L_{10}	L_{11}	L_{12}	L_{13}	L_{14}	L_{15}
aantal ongelukken	2	0	1	2	0	1	0	3	2	0	4	3	5	0	5

tabel 2

In de periode $L_8 - L_{15}$ was de fabriek uitgebreid en het aantal arbeiders het viervoud geworden van het aantal in de periode $L_1 - L_7$.

Voor $\underline{z}'_{n,1}$ vinden wij

$$\underline{z}'_{n,1} = 16.6$$

en voor de overschrijdingskans 0,25. Dit is geen significant resultaat als de onbetrouwbaarheid drempel ~~0,05 of~~ 0,01 is, zoals meestal het geval is.

Een ander voorbeeld.

Ook bij het bepalen van de benodigde apparatuur voor een automatische telefooncentrale wordt gebruik gemaakt van een variabele met een Poisson-verdeling voor het aantal telefoongesprekken, dat binnenkomt in de centrale of het aantal gesprekken, dat over een kabel verbonden moet worden. Men wil weten hoeveel aders de kabel moet bevatten wanneer een bepaald klein percentage der gesprekken mag uitvallen wegens volledige bezetting van de kabel. Door langdurige waarneming is het gemiddelde aantal telefoongesprekken bekend. Het aantal telefoongesprekken over de kabel gedurende een bepaalde periode heeft een Poisson-verdeling, waarvan het gemiddelde μ dus bekend is. Men wil nu het aantal aders c zo bepalen, dat de kans dat meer dan c gesprekken voorkomen een bepaalde zeer kleine waarde ε heeft. Dus c is bepaald door

$$P\{\underline{x} > c\} = \varepsilon$$

Het algemene toetsingsschema.

We kunnen een algemeen schema geven voor toetsing van lineaire hypothesen betreffende de parameters van variabelen, die een Poisson-verdeling bezitten.

Gegeven zijn de variabelen

$$\begin{array}{c}\underline{x}_{11}, \dots, \underline{x}_{1n_1} \\ \vdots \qquad\qquad \vdots \\ \underline{x}_{i1}, \dots, \underline{x}_{in_i} \\ \vdots \qquad\qquad \vdots \\ \underline{x}_{m1}, \dots, \underline{x}_{m,n_m}\end{array}$$

deze komen uit verschillende Poisson-verdelingen met parameters

$$\begin{matrix} \mu_{11}, & \cdots, & \mu_{1n_1} \\ \vdots & & \vdots \\ \mu_{i1}, & \cdots, & \mu_{in_i} \\ \vdots & & \vdots \\ \mu_{m1}, & \cdots, & \mu_{m,n_m} \end{matrix}$$

We toetsen nu verschillende lineaire nulhypothesen:

H_0 : alle μ 's zijn onderling gelijk

H_0' : alle μ 's van iedere rij zijn onderling gelijk.

Wanneer we onderstellen, dat H_0' waar is toetsen we de nulhypothese:

H_0'': de μ 's van de rijen zijn alle onderling gelijk.

De toetsingsgrootheden zijn voor

H_0 :
$$\underline{z}_{n,1} = \sum_{ij} \frac{(x_{ij} - \bar{x})^2}{\bar{x}}$$

verdeeld als χ^2_{n-1} 3) (dit is de al besproken toetsingsgrootheid), voor

H_0' :
$$\underline{z}_{n,m} = \sum_{ij} \frac{(x_{ij} - \bar{x}_i)^2}{\bar{x}_i}$$

verdeeld als χ^2_{n-m},
voor

H_0'' :
$$\underline{z}_{m,1} = \sum_{1}^{m} {}_i \frac{n_i(\bar{x}_i - \bar{x})^2}{\bar{x}}$$

verdeeld als χ^2_{m-1}.

Hierin is $\bar{x}$ het gemiddelde van alle waarnemingen en $\bar{x}_i$ het gemiddelde van alle waarnemingen uit de i^e rij.

Dit toetsingsschema stemt overeen met dat van de variantie-analyse met dit verschil, dat in de variantie-analyse alleen de hypothese H_0'' getoetst kan worden. Het verschil tussen beide methoden ligt hoofdzakelijk daarin, dat de Poisson-verdeling 1 parameter heeft, terwijl de normale verdeling er twee heeft.

We kunnen ook de toetsingsgrootheden $\underline{z}_{n,m}$ en $\underline{z}_{m,1}$ nog uitbreiden tot gevallen, waarin niet beschouwd wordt dat de parameterwaarden gelijk zijn, maar een bepaalde onderlinge verhouding hebben. Zo vonden we al dat $\underline{z}_{n,1}$ een speciaal geval was van $\underline{z}'_{n,1}$.

---

3) χ^2_{n-1} staat voor: χ^2 met $(n-1)$ vrijheidsgraden.