DIE STATISTIEK EN WAARSKYNLIKHEIDSREKENING EN HUL TOEPASSING

REDE

BY DIE AANVAARDING VAN DIE PROFESSORAAT IN DIE STATISTIEK EN HANDELSWISKUNDE AAN DIE UNIVERSITEIT VAN PRETORIA,

UITGESPREEK OP 15 MAART 1939

deur

B. DE LOOR, M.Sc. (S.A.), Ph.D. (Amsterdam).

Pretoria

1939.

PUBLIKASIES VAN DIE UNIVERSITEIT VAN

PRETORIA.

Reeks IV. : Intreeredes No. 15.

Mnr. die Rektor,
Kollegas van die Staf,
Dames- en Here-studente,
Dames en Here.

Graag vra ek vir enkele oomblikke u welwillende aandag vir 'n paar grondprobleme in die toepassing van die statistiese metodes en die daarmee nou saamhangende waarskynlikheidsrekening.

Die steeds toenemende gebruik van statistiese metodes in wetenskaplike ondersoek in al sy vertakkings is vir ons tyd kentekenend. Nie langer word dit, soos vroeër die geval was, alleen gebruik vir beskrywing van die gebeurde nie, maar dit word veral vandag aangewend vir die hoofdoel van alle ondersoek: deur induksie te kom tot verdere kennis en wetmatighede.

In die 17e en 18e eeu word die statistiese metodes, in die destyds beperkte sin van die woord, alleen toegepas in die bevolkingsbeskrywing wat hoofsaaklik geboorte- en sterftesyfers behandel het. Die eerste belangrike publikasie was die van John Graunt: "Natural and Political Observations mentioned in a following Index, and made upon the Bills of Mortality." (1662). Hierin word al die vier belangrikste feite in die ontleding van die bevolkingsleer bewys. 'n Paar jare later verskyn die „Waerdye van Lyf-renten" van Johan de Witt waarin ons vir die eerste maal die toepassing van die waarskynlikheidsrekening vind op die begrip lewenskans in die versekeringswiskunde.

Die teenwoordige betekenis van die woord „statistiek" is nog geen 100 jaar oud nie. In die begin van die 19e eeu pas Quetelet in sy „Essai de physique sociale" die statistiese metode toe op allerhande sosiologiese verskynsels, op antropologie en kriminologie. Hy word daardeur die vader van die moderne statistiek en is sy werke die uitgangspunt van die verroweringstog van die statistiek. Dit kos nog baie stryd, wat in sommige gevalle nog voortduur, voordat die statistiese metodes 'n erkende hulpmiddel by wetenskaplike ondersoek word. Eers dring dit deur in die wetenskappe wat nou verwant is aan die sosiale: soos bv. biologie, erflikheidsleer en medisyne. Dan kom die meer eksakte wetenskappe aan die beurt: natuurkunde, skeikunde en selfs die sterrekunde, wat nog altyd as beste voorbeeld van die klassieke meganika gegeld het. In die natuurkunde het die statistiese metodes al volledige erkenning verkry en is daar vandag geen standaardwerk in die vak wat nie hiervan gebruik maak by die afleiding van formules nie. In die ekonomie is hierdie erkenning nog nie volledig nie. Dit hang saam met die metode-vraagstuk in hierdie vak. Belangrike rigtings soos bv. die Duitse histories-georiënteerde skool ken alleen waarde toe aan die statistiese metodes vir sover dit beskrywend is, maar ontken die waarde van enige verdere toepassing van die metodes vir die analise van die gegewens.

Die toepassing in die verskillende takke van die wetenskap is natuurlik verskillend al na die bepaalde behoeftes. So is die gebruik van die steekproef-metode van groot belang in die opstel van proewe, bv. in tuinbou of veeartseny. In die meteorologie wat in sy ondersoek van die weer te doen het met 'n faktor met eweveel onberekenbare nukke as die mens, is die behoeftes weer van 'n geheel ander aard. Maar al die onderliggende beginsels is een en dieselfde, selfs in sy toepassing op die taalkunde of op die ontleding van die Griekse prosa-ritme.

Maar laat ek, voor ek verder gaan in kort verduidelik wat onder statistiek verstaan word. Verryn Stuart definieer statistiek as „de methodische boekhouding der voor massale waarneming vatbare verschijnsels." Dit is sy eerste taak: meting en beskrywing van massa-verskynsels. Hierdie gedeelte is suiwer beskrywend. Grafiese en wiskundige metodes word ontwikkel waardeur die kenmerkende van 'n verskynsel uitgebring en weergegee kan word in 'n paar getalle, bv. ons meet die hoogte van 'n groot aantal mans en karakteriseer dit deur al die getalle in die vorm van 'n frekwensieverdeling weer te gee waarvoor dan 'n gemiddelde en strooiingsmaat bereken word wat die verdeling tipeer in twee eenvoudige getalle.

Of ons meet die hoogte en die gewig van elke man en druk deur een getal — die korrelasie-koëffisiënt — uit in hoever die twee verskynsels saam varieer.

Die eerste taak is deur sy meting en beskrywing suiwer tegnies van aard. Die tweede taak daarenteen is wetenskaplik. Dit bestaan uit die beoordeling van die resultate wat gevind is. Met behulp hiervan word dan gestreef om kennis uit te brei tot korresponderende karakteristieke wat nie gemeet is nie, bv. wat kan ons aflei omtrent die hoogte van al die mans wanneer ons die hoogte van 'n gegewe aantal meet?

Hierdie gedeelte wat gewoonlik die stochastiese gedeelte genoem word is nou verbind met, en 'n toepassing van, die wiskundige waarskynlikheids-rekening. Die naam stochastiek is afgelei uit die Griekse woord wat beteken: na die doel strewende, en is oorspronklik gebruik deur Jacob Bernoulli, een van die grondleggers van die waarskynlikheidsrekening, in verband met 'n stelling wat vandag bekend staan as die stelling van Bernoulli. Die stelling en die betekenis van die woord stochasties kan die beste deur 'n voorbeeld verduidelik word.

Veronderstel ons gooi 'n dobbelsteen en noteer die aantal male wat die ses bo kom. Volgens die wiskundige waarskynlikheidsrekening is die kans daarop een-sesde. Dus as ons ses maal gooi kan ons eenmaal ses verwag. Probeer ons dit dan sal dit allesbehalwe uitkom. Vir klein reekse is daar 'n groot onreëlmatigheid. Wanneer ons egter 600 maal agtermekaar gooi dan sal ons vind dat daar ongeveer 100 maal ses bo kom, d.i. die relatiewe frekwensie streef na die teoretiese waarskynlikheid van 1/6e en al hoe groter die aantal male is wat gegooi word, met soveel meer sekerheid kan ons sê dat die relatiewe frekwensie, wat 'n proefondervindelike resultaat is, naby die teoretiese waarskynlikheid van 1/6e val. Dit is die inhoud van die Bernoulli-stelling en verduidelik terselfdertyd die woord stochasties.

Om die toepassing van die statistiese metodes te verduidelik is dit miskien goed om uit te gaan van die voorbeeld van wiskundige afleiding uit veronderstellings. Uit sekere gegewens word op 'n streng logiese wyse die resultaat afgelei. Neem 'n eenvoudige voorbeeld: in die Euklidiese meetkunde word twee sye en 'n ingeslote hoek van 'n driehoek gegee: dan kan die driehoek slegs op een manier gekonstrueer word. Altyd kan, wanneer daar voldoende gegewens is die probleem op een bepaalde vaste wyse opgelos word deur logiese deduksie. Op identieke wyse los die klassieke meganika sy probleem op. Hier word weer bepaalde entiteitè gegee: partikels, massas, elektriese ladings, ens., en die kragvelde aangeneem net soos bv. in meetkunde die Euklidiese ruimte. Dan kan 'n probleem gestel word, bv. 'n partikel word met 'n gegewe snelheid in 'n gegewe rigting geprojekteer. As die swaartekrag 'n gegewe versnelling na onder veroorsaak wat is dan die baan van die projektiel en waar sal dit die horisontale vlak deur die projeksiepunt weer tref? Die oplossing van die vraag geskied op dieselfde manier as vir die meetkundige probleem: die resultaat is uniek en bepaald deur die logiese deduksie en kan gebruik word vir verdere ondersoek, bv. voorspelling van toekomstige resultate.

Die sterrekunde is wel die mooiste voorbeeld van die soort deduksie. Gegee die beginstand en snelheid van 'n planeet dan kan met behulp van eenvoudige wette die baan van die liggaam noukeurig bereken word en ook sy posisie op enige tydstip. Dink maar net daaraan hoe noukeurig sons- en maansverduisteringe voorspel kan word.

Hierdie beeld het Laplace voor oë gehad toe hy die skerpste formulering van die kausaliteitswet in die natuurwetenskappe gemaak het, deur die ideaal as volg te stel:

„'n Gees wat op 'n bepaalde tydstip al die in die natuur werkende kragte ken en die korresponderende posisie van al die dinge waaruit die wêreld bestaan, kan, wanneer dit alomvattend genoeg is, al die gegewens aan die Analise onderwerp en in een en dieselfde formule die bewegings van die grootste liggame van die heelal sowel as van die ligste atome saamvat; niks is vir die Gees onbepaald nie en in sy oë is die hele toekoms sowel as die verlede verteenwoordig.

Die menslike gees lewer in die volkomenheid wat hy aan die sterrekunde kon gee, 'n swakke weerspieëling van die Gees. Al sy pogings in die soek naar waarheid moet daarop gerig wees om voortdurend die Gees te benader wat ons geskets het, maar hy sal altyd daar oneindig ver van weg bly." Tot sover Laplace. Wiskundig geformuleer beteken dit dat vir elke partikel bewegingsvergelykings gegee kan word in die vorm van differensiaal-vergelykings van die 2e orde in terme van sy massa, koördinate, tyd en funksies van sy posisie t.o.v. die ander partikels.

'n Ideaal wat vir eens en altyd onuitvoerbaar is! Dit kan nog uitgevoer word in die astronomie waar 'n planetestelsel beskou kan word as 'n stelsel van partikels. Maar daar hou dit ook op. Sodra ons kom in die wêreld van die mikrokosmiese, van die atome, of van die kontinue, dan kan die natuurkundige in die verste verte nie eers droom aan 'n benadering van die ideaal nie. Nie meer die beweging en wetmatigheid van die enkele partikel nie, maar die wetmatigheid wat statisties uit die massabeeld te voorskyn kom word nou fundamenteel. Nie langer die determinisme en kousaliteit uit die Laplace-wêreld nie, maar 'n ander soort determinisme in die statisties-gemiddelde.

Die verskil tussen die wiskundige en statistiese afleidings word duidelik uit 'n voorbeeld wat Bertrand aangee: Beskou 'n naderende reënwolk. Die alomvattende gees van Laplace kan bereken waar en wanneer elke reëndruppel, hoe klein ook, ontstaan het; wat die baan van elke druppel is, en waar dit uiteindelik op die grond te reg kom. Niks is vir die gees onseker omtrent die verloop van die reën nie. So iets is egter onbereikbaar vir ons beperkte menslike verstand. Ons kan nie eers met sekerheid sê of 'n wolk ons reën sal bring nie. Al wat ons wel met sekerheid kan sê is dat, as daar 'n goeie bui reën uitsak, alles wat ongedek is gelykmatig nat sal word en elkeen wat hom na buite begeef sonder jas of sambreel 'n nat pak klere sal kry. Met ander woorde, ons kry hier uit die massabeeld 'n wetmatigheid wat ons wel kan begryp, sonder dat ons op die beweging van die enkele partikel hoef terug te gryp. Ons kennis van die gebeure neem dus 'n „statistiese" vorm aan in teenstelling met die wiskundig-deterministiese vorm.

Of om 'n ander beeld van Quetelet te gebruik: Wanneer ons met kryt op die swartbord 'n groot sirkel trek en 'n klein stukkie van naby beskou, bv. met 'n vergrootglas dan sal ons net 'n min of meer toevallig verspreide aantal wit stippe vind, hoe noukeurig en fyn die sirkel ook al geteken is. Wanneer ons egter al hoe verder van die swartbord wegstaan dan kom 'n al hoe groter aantal punte in die gesigsveld wat later as afsonderlike punte uit die gesig verloor word en plek maak vir 'n skynbaar gladde kromme wat die algemene (of statistiese) wet aangee van die verdeling van die punte. Hier verdwyn dan die enkeling in die massa en kry ons 'n wetmatigheid uit die massa-verskynsel waarin die enkele gevalle hulle, so te sê, verdig het.

Die statistiese standpunt wat dus beteken dat afgesien word van die verloop van die enkele verskynsel en alleen die massa-verskynsel betrag word, is gebaseer op die sogenaamde empiriese „Wet van groot getalle." Dit is die spil waar alles om draai en die beginsel waaraan die statistiese metode sy betekenis ontleen. Die wese van die wet bestaan hierin dat in die natuur en in maatskaplike verskynsels altyd gevind word dat wanneer die aantal waarnemings groter word daar 'n reëlmaat te voorskyn kom wat nie voorspel kon word uit 'n klein aantal waarnemings nie. Die Hollandse natuurkundige en filosoof s'Gravensande het in die 17e eeu van hierdie wet al 'n duidelike en skerp formulering gegee: „Saepe vero regularitas quae consideratis paucis effectibus nos fugit, ubi plures at examen vocantur, detergitur" (dit gebeur dikwels dat 'n reëlmatigheid, wat ons by die waarneming van enkele gevalle ontgaan, te voorskyn kom wanneer 'n groot aantal gevalle waargeneem word).

Hierdie reëlmaat in 'n groot aantal waarnemings is so sterk dat ons dit gewoonlik onwillekeurig as vanselfsprekend beskou. Dit is die grondslag van die berekenings van die lewensversekeringsmaatskappy en van die begroting van die Minister van Finansies. Dit tree orals op, sowel in natuurlike, maatskaplike of morele verskynsels. Die gemiddelde jaarlikse temperatuur of reënval is taamlik konstant, die verhouding van die aantal manlike tot vroulike geboortes bly konstant, die aantal huwelike en selfmoorde per 10,000 van die bevolking bly betreklik konstant. Die reëlmatigheid is die conditio sine qua non van ons teenwoordige maatskaplike gebou wat gegrond is op arbeidsverdeling en ruiling van produkte. Al die massa-verskynsels toon naas die reëlmaat in die massa 'n groot willekeurigheid vir die enkele verskynsel of vir 'n paar verskynsels. In die somer kan daar koue dae wees, in die winter warme dae maar die gemiddelde word daar byna of geheel nie deur beïnvloed nie. Wanneer 'n klein distrik van maand tot maand beskou word dan sal die verhouding tussen die aantal manlike en vroulike geboortes baie skommel maar vir die land as geheel is dit taamlik konstant naby 105:100. Die sterfregte sal van boedel tot boedel kan varieer van niks tot duisende ponde maar die opbrengs gedurende elke jaar vir die hele land is betreklik konstant. En so kan op elke gebied voorbeelde aangehaal word.

Hoe kan hierdie reëlmaat en wetmatigheid in die massa verklaar word naas die willekeur in die afwykings van die alleenstaande verskynsel?

Die eerste verskynsels waarin die reëlmaat reeds in die 17e eeu opgemerk is, was die geboorte- en sterftesyfers en dit lê voor die hand dat vir die verklaring 'n goddelike oorsprong gesoek is. So skryf Arbuthnot in 1710 oor "An argument in favour of the Divine Providence drawn from the constant regularity which one finds in the births of both sexes." Arbuthnot beweer dat wanneer toeval alleen die oorsaak is, so 'n reëlmatigheid ondenkbaar sou wees. Nicolaas Bernoulli ('n neef van die reeds genoemde Jacob Bernoulli), antwoord hierop dat, wanneer die toeval alleen as verklaring geneem word die reëlmaat, wat Arbuthnot in sy syfers van 82 jare gevind het, nog baie groter sou wees. In die middel van die 18e eeu verskyn die beroemde werk „Die göttliche Ordnung in den Veränderungen des menschlichen Geschlechts, aus der

Geburt, dem Tode und der Fortplanzung desselben erwiesen" van Pastor J. P. Süssmilch. Die titel van die werk spreek vir homself.

Later word 'n meer natuurlike verklaring gesoek in die saamwerking van die sistematiese en die toevallige, onafhanklike kragte of oorsake waardeur elke individuele verskynsel ontstaan. Die optree of wegbly van elke verskynsel is die gevolg van 'n groot aantal kragte waarvan sommige konstant en baie ander toevallig is.

Die afwykings in die individuele geval is nou 'n resultaat van die toevallige onafhanklike kragte wat eenmaal almal saamspan in die een rigting, 'n volgende maal weer in die ander rigting en die meeste male mekaar se invloed neutraliseer.

Wanneer nou 'n groot aantal gevalle geneem word dan verdwyn die resultate van die toevallige kragte en kom die resultaat van die konstante kragte te voorskyn, wat aan al die verskynsels gemeenskaplik is. Hierin kom dan die struktuur van die massa uit, of soos Wagemann dit uitdruk die „samehang van die onsaamhangende" of die „individuele vryheid by 'n kollektiewe dwang."

Ek het gespreek van „toevallig." Op al die filosofiese en ander moeilikhede van die begrip wil ek hier nie ingaan nie maar alleen aanhaal wat ons onder toeval in ons verband moet verstaan. Wat word bedoel met 'n toevallige gebeurtenis? Klaarblyklik 'n gebeurtenis waarby ons deur die gekompliseerdheid van die daarop inwerkende kragte nie uit die aanvangstand kan bereken of voorspel of dit sal voorval of nie. Dus 'n gebeurtenis wat meer of minder vaak optree en waardeur ons die indruk van iets onverwags kry. Ons so ewe genoemde toevallige kragte moet nou aan die volgende eienskappe voldoen: (i) dit moet vir die verskynsel afwykings gee wat afwisselend na die een of na die ander kant gaan, want as die afwyking steeds na een kant was dan sou die krag konstant wees en nie langer toevallig nie; (ii) die resultaat van elke krag moet onafhanklik wees van die van die orige kragte. Dit is nogal 'n belangrike eienskap wanneer ons kom by die wiskundige formulering; (iii) hulle moet talryk wees, want die konstante struktuur verskyn alleen as sulks deurdat dit uit 'n groot aantal gevalle te voorskyn tree; tenslotte (iv) moet die resultaat van elke krag klein wees in verhouding tot die van al die kragte saam, om dieselfde rede as vir die eerste eienskap genoem: as die resultate nie klein bly nie dan word die kragte tipies en konstant en nie langer toevallig nie. Die wetmatigheid in die massa, of die empiriese wet van groot getalle, is dus vervat in die begrip van die massa self. Deur die groot aantal kom die reëlmaat te voorskyn, wat by enkele gevalle verduister word deur die resultaat van die toevallige oorsake.

Die wiskundige formulering van hierdie reëlmaat en die wet van groot getalle vind ons in die waarskynlikheidsrekening. Dis interessant om te bedink dat juis die reëlmaat in die praktyk direkte aanleiding was tot die eerste wiskundige behandeling van die waarskynlikheidsbegrip. Die Chevalier de Méré, 'n bekende dobbelaar uit die middel van die 17e eeu, het uit die feit dat die kanse gelyk is om minstens een ses te kry in ongeveer vier gooie met een dobbelsteen, afgelei dat die kanse om minstens een dubbel ses te kry met twee dobbelstene gelyk is in 24 gooie. Ondervinding het hom egter geleer dat die aantal gooie groter moet wees. Hierdie skynbare teenstrydigheid en 'n paar ander moeilikhede het de Méré aan Pascal voorgelê. Pascal het die moeilikhede opgelos en daaroor gekorrespondeer met Fermat, en hieruit ontstaan die wiskundige waarskynlikheidsrekening. Sy uitdrukkingswyse is tot vandag toe nog altyd in die vorm van geluk-spele: Uit 'n bak, wat seker aantalle balle van verskillende kleur bevat, word balle getrek; uit 'n pak kaarte word een of meer kaarte getrek; met een of meer dobbelstene word gegooi en meer dergelike voorbeelde en kombinasies daarvan.

In die ontwikkeling van die waarskynlikheidsrekening vind ons van die begin af twee strominge wat dikwels vreemd teenoor mekaar staan wat die grondslae en uitgangspunte betref maar wat tog lei tot dieselfde resultate. Dit is die ou teenstelling rasionalisme-empirisme. In die eerste geval word die wiskundige waarskynlikheid of waarskynlikheid a priori gedefinieer deur die verhouding van die aantal gunstige gevalle waarin 'n gebeurtenis voorval tot die aantal gelykmoontlike gevalle. In hierdie een woord „gelykmoontlike" steek daar baie groot moeilikhede wanneer dit kom tot die toepassing van die waarskynlikheidsrekening op die praktyk, en vermy die empiris hierdie voetangels en klemme deur die waarskynlikheid te definieer as die limiet van die verhouding van die aantal gevalle waarin 'n gebeurtenis voorgeval het tot die totaal aantal gevalle waargeneem wanneer die reeks onbepaald voortgeset word.

Dit is interessant om op te let dat die teenstelling oorspronklik ook tussen volke as geheel was. In die Roomse

Frankryk waar 'n verwende adel nog na hartelus kon dobbel, was die belangstelling gekonsentreer op die à priori definisie van waarskynlikheid en die afleidings daaruit. Selfs vandag nog is die Franse skool die vernaamste voorstanders van die rigting. Aan die ander kant het in die Calvinistiese Engeland en Holland, waar dobbel nie toegelaat was nie, die belangstelling gegaan in die praktiese rigting: toepassing op bevolkingsverskynsels soos bv. deur die reeds genoemde John Graunt en Johan de Witt. Hier vind ons die empiriese definisie of die waarskynlikheid à posteriori.

Teen die einde van die 17e eeu het Jacob Bernoulli gesien dat die dubbele oorsprong lei tot 'n dubbele definisie. In die poging om die gelykwaardigheid van die definisies te bewys lei hy die stelling af wat vandag na hom genoem word en wat een van die belangrikste stellings in die waarskynlikheidsrekening is. Ek het reeds met 'n voorbeeld die stelling verduidelik maar wil dit nog 'n maal in woorde as volg weergee: Veronderstel 'n gebeurtenis E het 'n a priori waarskynlikheid p. In n proewe val E m maal voor. Dan is die relatiewe frekwensie van E m/n. Die stelling beweer nou dat met 'n waarskynlikheid wat so naby as gewens aan die sekerheid nader, die relatiewe frekwensie m/n stochasties na die à priori waarde p sal streef, wanneer die aantal herhalings n slegs groot genoeg geneem word.

Of om 'n getalle-voorbeeld te gebruik: Wanneer 'n muntstuk gegooi word dan is die kanse à priori om kop bo te kom ½. Wanneer ons 100 maal gooi dan sal in 16 persent van die gevalle die relatiewe frekwensie wees .49, .50 of .51, dus in waarde met hoogstens 1/100 van die à priori waarde ½ afwyk. Wanneer 1000 maal gegooi word dan is 47 persent binne die grense ½ ± 1/100, en wanneer 10,000 maal gegooi word dan is 95 persent van die gevalle waarin die relatiewe frekwensie met hoogstens 1/100 afwyk van die à priori waarde ½.

Let op: in die stelling word uitgegaan van 'n bekende kans à priori en word die voorspelling van 'n frekwensie bewys vir 'n gegewe aantal herhalings. Maar nou het ons in die praktyk steeds te doen met die omgekeerde geval: hoe kan aan 'n verskynsel waarvan die relatiewe frekwensie in 'n aantal herhalings bekend is, 'n à priori waarskynlikheid toegevoeg word waarop verdere deduksies gebou kan word? Die stelling van Bernoulli bevat alleen teoretiese oorwegings uitgaande van 'n waarskynlikheid à priori. Dit is 'n teoretiese verklaring van die empiriese wet van groot getalle maar nie 'n bewys daarvan nie. Laasgenoemde wet is 'n proefondervindelike feit en kan nie teoreties bewys word nie.

Om die geval nog meer gekompliseerd te maak moet opgemerk word dat in die praktyk so goed as nooit 'n verskynsel gevind word waarvan die waarskynlikheid konstant bly oor verloop van tyd nie. Hierdie moeilikheid word oorkom deur die Poisson-uitbreiding van die Bernoulli-stelling waarin aangeneem word dat die waarskynlikheid p van groep tot groep verander en dan bewys word dat die relatiewe frekwensie stochasties na 'n deursnee-waarskynlikheid streef.

Bernoulli het die probleem om van 'n relatiewe frekwensie te kom tot 'n waarskynlikheid à priori nie opgelos nie en dit is vandag nog 'n strydpunt. Die probleem is van fundamentele belang veral in die steekproef-teorie. Ek wil dit met 'n eenvoudige voorbeeld verduidelik. Veronderstel ons het 'n bak met wit en swart balle van onbekende saamstelling. 100 maal word, met teruglegging na trekking, 'n bal getrek en gevind word dat 63 maal wit te voorskyn gekom het. Hoe groot is nou die kans dat die bak saamgestel was uit 6 wit en 4 swart balle? M.a.w. wat is die kans dat die oorsaak, d.i. die saamstelling 6/10 wit is. In die wiskundige teorie het ons vir die berekening van die kans 'n heeltemal geldige stelling van Bayes waaruit die gevraagde kans afgelei kan word. Maar dit is gebaseer op die veronderstelling dat al die verskillende saamstellings „gelykmoontlik" is. In die praktyk word hierdie veronderstelling egter nie bewaarheid nie en het dit al dikwels tot ongeoorloofde gevolgtrekkings en onmoontlike resultate gelei. 'n Uiterste standpunt in hierdie stryd neem R. A. Fisher in wat die hele stelling van Bayes geskrap wil sien en vervang deur 'n stelling wat die gewraakte veronderstellings nie aanneem nie. Hierin slaag hy deur die invoering van die "confidence intervals" waarin 'n klein maar essensiële verandering in die veronderstellings gemaak word.

Soos reeds gesê verskil die twee rigtings, nl. die wiskundige en die empiriese alleen in grondslag en nie in resultate nie. Albei aksiomatiseer en lei daaruit die resultate af. Die moderne wiskundige waarskynlikheidsrekening is in seker sin die omkering van die klassieke teorie. Dit word gegrond op die wiskundige versamelingsleer waardeur dan die wetenskaplike determinisme versoen word met die objektiewe bestaan van die toeval. Die waarskynlikheid as die verhouding van 'n aantal gunstige tot 'n aantal moontlike gevalle word nou 'n bewysbare stelling, d.i. kan afgelei word uit die aksiomas.

Aan die ander kant aksiomatiseer von Mises, die beste verteenwoordiger van die empiriese rigting, die empiriese wet van groot getalle, na invoering van die begrip „Kollektiv." Vir die opbou van die teorie het hy nog 'n meer kunsmatige aksioma van „Regellosigkeit" nodig, wat reeds veel kritiek uitgelok het.

Vir von Mises is die waarskynlikheidsrekening 'n „wiskundige natuurwetenskap" van dieselfde soort as meetkunde en meganika. Die doel is om vir 'n bepaalde groep verskynsels nl. die massa-verskynsels of die sig herhalende gebeurtenisse, 'n oorsigtelike beskrywing te gee, net soos die meetkunde dit vir die ruimtelike, en die meganika dit vir die verskynsels van beweging doen.

Of dit nou al as aksioma aangeneem word, of as stelling bewys word uit ander aksiomas, vir albei metodes lê die kernpunt hierin dat vir 'n onbepaald voortgesette reeks proewe of herhalings die proefondervindelike relatiewe frekwensie vereenselwig kan word met 'n wiskundige waarskynlikheid. Dit is die brug tussen teorie en praktyk en 'n antwoord op die vraag: hoe kan die waarskynlikheidsrekening toegepas word?

Waarom word die waarskynlikheidsrekening toegepas? In alle teorie soek ons 'n hulpmiddel vir die beskrywing en verklaring van die voorstelling van die fisiese werklikheid. Ons probeer ons kennis daarvan uit te brei deur 'n verklaring van die verskynsels te soek. Deur middel van statistiek en waarskynlikheidsrekening probeer ons ook dieselfde te doen vir die massa-verskynsels. Gepoog word om die verskynsels terug te voer en so noukeurig moontlik te beskryf deur bak-skemas, d.i. deur 'n waarskynlikheid aan te gee waardeur die optree van die verskynsel beskryf kan word. Die eenvoudigste geval is die waarin by herhaling getrek word uit 'n bak met 'n konstante saamstelling, met teruglegging na elke trekking. Hieruit kom die binomiale frekwensieverdeling te voorskyn wat in die limiet oorgaan in die Gauss-normaalkromme. Hierdie normaalkromme kan opgevat word as 'n tweede en algemene wet van groot getalle en ontstaan, soos uit die afleiding blyk, wanneer die afwykings alleen resultaat is van toevallige onafhanklike kragte. Die kromme kom voor in die foute-teorie, erflikheidsleer en biometrie waarin louter toevallige faktore as verklaring van die afwykingsverskynsels aangeneem kan word. Op grond hiervan is dit ook een van die kernpunte van die steekproef-teorie.

In alle takke van wetenskap vind ons egter meer gevalle van skewe en nie-normale frekwensieverdelings as van normale verdelings. Sommige van die nie-normale verdelings kan ons nog terugvoer tot 'n bak-skema, waarin die waarskynlikhede nie langer onafhanklik bly nie maar mekaar beinvloed. Twee gevalle wil ek u noem: In die geval van „aansteking" word, wanneer 'n wit bal getrek word, 'n paar verdere wit balle bygevoeg sodat by die volgende trekking die kans om wit te trek verhoog word. Analoog vir die ander kleure in die bak. In die ander geval van „erflikheid" word twee bakke gebruik van verskillende saamstelling en word beurtelings uit die eerste of uit die tweede bak getrek al na die kleur by die vorige trekking wit of swart was.

In albei hierdie voorbeelde kry ons die sg. waarskynlikheidskettings waarin elke kans beïnvloed word deur die onmiddellik voorafgaande kans.

Die twee skemas kan gebruik word vir die beskrywing van verdelings soos by aansteeklike siektes voorkom of by brandversekerings waarin die voorval van een gebeurtenis die kans op die voorval van die volgende of naburige gebeurtenis verhoog.

Genoemde bakskemas lei nog tot 'n oorsigtelike wiskundige formulering en kan gebruik word om tipes I en II van die sg. Pearson-stelsel van frekwensiekrommes voor te stel. Maar alleen in die eenvoudigste gevalle geluk dit om 'n verklaring van die afwykingsverskynsels deur middel van teoretiese bak-skemas te gee. Gewoonlik is ons nie in staat om met enige sekerheid aan te gee met watter soort waarskynlikheid ons in die verskynsel te doen het nie en lei algemener of saamgestelder bak-skemas tot sulke wiskundige moeilikhede dat van 'n verklaring nie langer sprake is nie. Ons kry dan alleen nog 'n oorsigtelike beskrywende vorm deur middel van hoër wiskundige funksies soos die geval is met die Pearson-krommes.

Die verklaring dat die afwykings deur louter toeval ontstaan, is hier ook nie meer geldig nie en ons moet aanneem dat ander faktore, soos bv. die menslike faktor, in die spel kom, en sy ekstra-afwykende invloede laat geld. Dit is veral die geval in 'n wetenskap soos ekonomie.

As 'n eenvoudige en tipiese voorbeeld kan ons die verdeling van 'n aantal mans volgens hoogte en gewig neem. Daar ons so goed as geen invloed op ons hoogte kan uitoefen nie is die frekwensie-verdeling daarvan simmetries en normaal, maar, waar ons wel invloed op ons gewig kan uitoefen is die verdeling volgens gewig skeef en kan die normaalkromme daar nie so goed op toegepas word nie.

Met die verklaring in die eenvoudigste gevalle van die ontstaan van die afwykings, het ons egter nog nie 'n verklaring vir die stabiliteit en reëlmaat as gevolg van die konstante kragte nie. In die bewys van die stabiliteit stel ons 'n feit van beskrywende aard vas wat ons egter niks kan vertel van die aard van die agterliggende konstante kragte en samehange nie. Ons stel vas dat die geslagsverhouding by geboorte nagenoeg konstant bly op ongeveer 105:100 as voldoende groot getalle geneem word, maar 'n bevredigende verklaring van die feit is nog nie gegee nie. Hierdie stabiliteit is van ontsaglik groot praktiese belang maar van minder wetenskaplike belang waar dit deduksie en die opsporing van kousaliteitsrelasies geld.

Tenslotte wil ek nog 'n oomblik stilstaan by die statistiese deduksie en die rol wat die statistiek speel in die aflei van kousaliteitsrelasies.

Vir hierdie afleiding is nie die stabiliteit nie, maar juis die veranderlikheid van die statistiese getal of gemiddelde van meer belang. Aansienlike veranderings in die getalle is gewoonlik simptomaties van die veranderings in die agterliggende kragte wat dit veroorsaak. Statisties kan ons op verskillende maniere vasstel dat daar sulke kousaliteitsrelasies en meer direkte oorsaaklike verbande bestaan. Die belangrikste werktuig hiervoor is wel die korrelasierekening.

Wat die oorsake is en hoe die relasies ontstaan kan ons nie uit die statistiese metodes verklaar nie, maar die verklaring moet gesoek word in die bepaalde tak van wetenskap waarvan die getalle onder bespreking is. Die vasstelling van 'n statistiese reëlmaat of kousaal-verband beteken nie dat ons 'n volledige wetenskaplike verklaring van die verskynsels kan gee nie. Die kousaal- of funksioneel-verband wat ons kan bewys, beteken nog geensins dat ons 'n volledige kousaalketting het nie.

Wanneer bv. 'n Sweedse entoesias 'n hoë korrelasie kan vasstel tussen die aantal ooievaars, wat nes gemaak het en die aantal geboortes, dan beteken dit nog nie dat dit 'n werklike kousale verklaring van die verskynsel is nie. Net so min as wanneer in New York gevind word dat daar 'n betreklik hoë korrelasie bestaan tussen melkpryse en aandelepryse. Op hierdie manier is al baie ontledings gemaak wat van alle waarde ontbloot is, omdat die agtergrond uit die oog verloor is, sodat dit geen wonder is dat 'n skeptikus daarop die bekende Engelse woord kan toepas nie: "There are three types of liars: a liar, a damn liar, and a statistician."

In alle toepassings van statistiese metodes moet ons altyd die rol wat hierdie metodes speel in gedagte hou. Statistiek is die onmisbare werktuig by wetenskaplike ondersoek maar moet altyd 'n werktuig bly. Die agtergrond en die beperkings wat daaruit ontstaan mag nie by die ontledings uit die oog verloor word nie. Anders sou dit 'n spel met abstrakte getalle word wat waardeloos word vir die bepaalde tak van wetenskap.

Statistiese metodes lewer alleen empiriese wette op: reëlmatighede en kousaal- of funksioneel-verbande, wat eers binne die bepaalde tak van wetenskap tot sy volle reg kom en 'n verklaring vind, en gebruik kan word tot verdere opbou van die wetenskap.

Maar elke werktuig kan ook terwille van homself bestudeer word in sover dit kan lei tot 'n verbetering in die toepassing van die werktuig.

Hierin, Mnr. die Rektor, Dames en Here, lê ook die taak van die nuwe departement van Statistiek en Handelswiskunde. Net soos die wiskunde bestudeer word terwille van homself en om te dien as werktuig om 'n eksakter formulering in ander wetenskappe te gee om daardeur kennis uit te brei moet die statistiek en waarskynlikheidsrekening wees 'n werktuig vir die ander wetenskappe. Maar ook moet die werktuig self bestudeer en verbeter word, om daardeur weer beter toegepas te kan word in die verskillende vertakkings van die wetenskap.

Mag ek eindig met 'n persoonlike woord?

Aan u Mnr. die Rektor, lede van die Raad en lede van die Senaat, my hartlike dank vir die vertroue in my gestel om my die opdrag te gee in die pas gestigte departement. Ek sal probeer om die vertroue te regverdig en my deel, al is dit gering, by te dra tot uitbreiding van die wetenskap en van die Universiteit sodat dit daardeur kan word tot die Voortrekker-Universiteit van ons volk.

Aan u, prof. Duminy, my hartlike dank vir die goeie saamwerking in die verlede. In die ou wiskundedepartement onder prof. du Toit het daar, ondanks verskille wat soms opgekom het, altyd 'n aangename gees geheers, en waar dit op ons rus om die twee departemente wat nou naas mekaar staan verder op te bou en uit te brei, spreek ek die hoop uit dat in die toekoms die saamwerking nog beter sal word.

PUBLIKASIES VAN DIE UNIVERSITEIT VAN PRETORIA.

Reeks III: Geesteswetenskappe.

No. 1. KRITZINGER, M. S. B.—
Afrikaanse en Nederlandse Letterkunde as Studievak aan die Universiteit van Pretoria.

N. 2. DUMINY, J. P.—
Die Wiskunde en Filosofie.

No. 3. SKAWRAN, P. R.—
Die Sentetiese Studie van die Persoonlikheid.

No. 4. GEY VAN PITTIUS, E. F. W.—
Volksregering met Besondere Verwysing na S.A.

No. 5. COERTZE, L. Ign.—
Watter Regsisteem Beheers die Verhouding tussen Owerheid en Onderdaan in die Unie, Romeins-Hollandse Reg of Engelse Reg?

No. 6. BOKHORST, M.—
Kultuur van 'n Waterland.

No. 7. VILJOEN, S. P.—
Die Voortbestaan van die Blanke Ras.

No. 8. ENGELBRECHT, J. A.—
Die Wordings- en Verwordingsgeskiedenis van die Koranna.

No. 9. PELLISSIER, G. M.—
Die Godsdiens as illusie volgens Freud.

No. 10. CRONJE G.—
Die Deterministiese Standpunt in die Sosiologie.

No. 11. WILLEMSE, W. A.—
Die Psigologie en Maatskaplike Afwykinge.

No. 12. WOLMARANS, H. P.—
Die Betekenis van die Openbaringsbegrip vir onse Tyd.

No. 13. GROENEWALD, E. P.—
Die Eksegese van die Nuwe Testament.

No. 14. LOMBAARD, B.—
Geologiese Ondersoek in Suid-Afrika.

No. 15. LOOR, B. DE.—
Die Statistiek en Waarskynlikheidsrekening en hul Toepassing.

Gedruk deur
TRANSVAALSE PERS BEPERK,
Kerkstraat 142, - Pretoria.