4.4 Niet-binomiaal

Inleiding

Het binomiale kansmodel is erg overzichtelijk: er zijn maar twee mogelijkheden ‘succes’ of ‘mislukking’ en het gaat om herhaling van steeds dezelfde kanssituatie. Maar natuurlijk bestaan ook discrete stochasten waarbij geen herhaling plaatsvindt. Denk in het vaasmodel aan een trekking zonder teruglegging.

Je leert in dit onderwerp

- werken met een andere belangrijke discrete stochast: de hypergeometrische stochast;
- de bijbehorende kansverdeling opstellen
- regels voor de verwachtingswaarde en de standaardafwijking van deze stochasten.

Voorkennis

- het vaasmodel;
- een kansverdeling opstellen bij een vaasmodel;
- de regels voor de verwachtingswaarde en de standaarddeviatie van de som van meerdere stochasten.

Verkennen

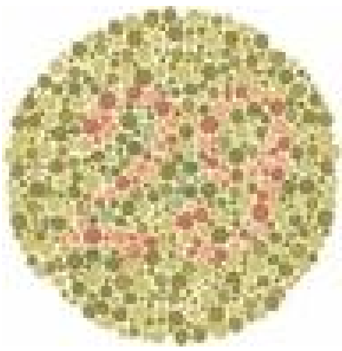
Opgave V1

In een groep van 25 mannen zijn 2 leden van die groep kleurenblind. Je trekt aselect een steekproef van vier mannen uit deze groep.

a Hoe groot is de kans dat daarbij één kleurenblinde man zit?

Van alle westerse mannen is een proportie van 8% kleurenblind. Je trekt aselect een steekproef van vier mannen uit deze groep.

b Hoe groot is de kans dat daarbij één kleurenblinde man zit?



Figuur 1

Uitleg 1

In een groep van 30 personen hebben 10 mensen een bepaalde eigenschap en de rest niet. Uit die groep wordt aselect een steekproef van 5 getrokken. Stochast M is het aantal mensen met deze eigenschap in de steekproef.

Wil je nu een kansverdeling voor M opstellen, bedenk dan dat het hier gaat om trekking zonder teruglegging. Dit betekent dat de kansen afhankelijk van elkaar zijn en dat een binomiaal kansmodel niet mogelijk is. M is nu een hypergeometrische stochast. De kans op bijvoorbeeld $M = 2$ kun je zo berekenen:

$$\text{P}(M = 2) = \frac{10}{30} \cdot \frac{9}{29} \cdot \frac{20}{28} \cdot \frac{19}{27} \cdot \frac{18}{26} \cdot \binom{5}{2} \approx 0{,}3600$$

Deze kans kun je ook uitrekenen door het aantal gunstige mogelijkheden en het totale aantal mogelijkheden te tellen met behulp van combinaties:

$$\text{P}(M = 2) = \frac{\binom{10}{2} \cdot \binom{20}{3}}{\binom{30}{5}}$$

Ga na dat je de volgende kansverdeling krijgt:

m	0	1	2	3	4	5
$\mathrm{P}(M = m)$	0,1088	0,3400	0,3600	0,1600	0,0295	0,0018

Tabel 1

Je kunt met behulp van de tabel en de grafische rekenmachine de verwachtingswaarde en de standaardafwijking berekenen.
Je vindt $\mathrm{E}(M) \approx 1{,}667$ en $\sigma(M) \approx 0{,}979$.

Kennelijk gaat $\mathrm{E}(M) = 5 \cdot \frac{10}{30} = 1\frac{2}{3} \approx 1{,}667$ ook hier op, maar dit geldt niet voor de formule die bij de binomiale verdeling voor de standaardafwijking geldt:

$$\sigma(M) \neq \sqrt{\frac{10}{30}\left(1 - \frac{10}{3}\right)} \approx 0{,}471$$

Dit komt doordat het gaat om zonder terugleggen en de kans op succes niet $\frac{10}{30}$ blijft, nadat je één of meer trekkingen hebt gedaan.

Opgave 1

Bekijk in **Uitleg 1** de kansverdeling van stochast M die het aantal mensen met een bepaalde eigenschap in een steekproef uit een kleine populatie van 30 personen weergeeft.

a Reken de kansen uit de kansverdeling van stochast M na.

b Bereken $\mathrm{E}(M)$ en $\sigma(M)$. Rond af op vier decimalen.

c Waarom is hier geen sprake van een binomiale kansverdeling?

Uitleg 2

In een groep van 30000 personen hebben 10000 personen een bepaalde eigenschap en de rest niet.
Uit die groep wordt aselect een steekproef van 5 getrokken.
Stochast M is het aantal mensen met deze eigenschap in deze steekproef.

Wil je nu een kansverdeling voor M opstellen, bedenk dan opnieuw dat het hier gaat om trekking zonder terugleggen. Dit betekent dat de kansen afhankelijk van elkaar zijn en dat een binomiaal kansmodel niet mogelijk is.
De kans op $M = 2$ is:

$$\mathrm{P}(M = 2) = \frac{10000}{30000} \cdot \frac{9999}{29999} \cdot \frac{20000}{29998} \cdot \frac{19999}{29997} \cdot \frac{19998}{29996} \cdot \binom{5}{2} \approx 0{,}3292$$

Nu verschilt een breuk als $\frac{9999}{29999}$ vrijwel niet van $\frac{10000}{30000} = \frac{1}{3}$.

Daarom kun je als je een kleine steekproef uit een heel grote populatie trekt, toch het binomiale kansmodel gebruiken. Hoewel het eigenlijk niet om onafhankelijke kansen gaat.

$$\mathrm{P}(M = 2) \approx \left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^3 \cdot \binom{5}{2} \approx 0{,}3292$$

Je ziet dat beide kansen bij benadering gelijk zijn aan elkaar. Daarom wordt in de praktijk bij een steekproef uit een veel grotere populatie waarbij het gaat om het wel of niet hebben van een bepaalde eigenschap, gewoon het binomiale kansmodel gebruikt.

Opgave 2

Bekijk in **Uitleg 2** de kansverdeling van stochast M die het aantal mensen met een bepaalde eigenschap in een steekproef uit een grote populatie van 30000 personen weergeeft.

a Bereken $\mathrm{P}(M = 3)$ en $\mathrm{P}(M = 4)$. Benader deze kansen ook met behulp van het binomiale kansmodel. Rond in beide gevallen af op vier decimalen.

b Bereken $\mathrm{E}(M)$ en $\sigma(M)$. Rond af op drie decimalen.

c Waarom kun je de kansverdeling van M heel goed benaderen door een binomiale kansverdeling?

Theorie en voorbeelden

Om te onthouden

Heel vaak is in een bepaalde kanssituatie helemaal geen sprake van een binomiale stochast. Er is dan geen sprake van een herhaling van onafhankelijke Bernoulli-experimenten (succes of mislukking).

Een belangrijk geval is de **hypergeometrische stochast**. Daarbij gaat het om een **populatie** van N elementen waarvan er a een bepaalde eigenschap hebben. Je trekt daaruit zonder teruglegging een **steekproef** van n elementen. De hypergeometrische stochast X is dan het aantal elementen in de steekproef dat deze eigenschap heeft. De kans op $X = x$ is:

$$\text{P}(X = x) = \frac{a}{N} \cdot \frac{a-1}{N-1} \cdot \ldots \cdot \frac{N-a}{N-x} \cdot \frac{N-a-1}{N-x-1} \cdot \ldots \cdot \binom{n}{x}$$

Met behulp van combinaties kun je dit ook uitrekenen:

$$\text{P}(X = x) = \frac{\binom{a}{x} \cdot \binom{N-a}{n-x}}{\binom{N}{n}}$$

Voor de verwachtingswaarde geldt: $\text{E}(X) = n \cdot \frac{a}{N}$. De standaardafwijking van X kun je nu alleen uit de kansverdeling zelf halen. Daarom bepaal je in de praktijk zowel $\text{E}(X)$ als $\sigma(X)$ met behulp van de grafische rekenmachine.

Bij een **kleine steekproef uit een heel grote populatie** kun je toch wel het binomiale kansmodel gebruiken, hoewel het eigenlijk niet om onafhankelijke kansen gaat. Dat komt omdat dan breuken als $\frac{a}{N}$ en $\frac{a-1}{N-1}$ vrijwel gelijk zijn. In de praktijk wordt bij een steekproef uit een heel veel grotere populatie waarbij het gaat om het wel of niet hebben van een bepaalde eigenschap, gewoon het binomiale kansmodel gebruikt.

Voorbeeld 1

In een klas zitten 12 jongens en 9 meisjes. Daaruit wordt een aselecte steekproef van vier personen gekozen. Stochast M is het aantal meisjes in de steekproef.

Stel een kansverdeling op voor M en bepaal de verwachtingswaarde en de standaardafwijking van M.

Antwoord

Bij de steekproef gaat het om trekking zonder terugleggen van vier elementen uit een populatie van 21. M is een hypergeometrische stochast.

De kans op bijvoorbeeld $M = 3$ is:

$$\text{P}(M = 3) = \frac{9}{21} \cdot \frac{8}{20} \cdot \frac{7}{19} \cdot \frac{12}{18} \cdot 4 \approx 0{,}1684$$

De complete kansverdeling wordt:

m	0	1	2	3	4
$\text{P}(M = m)$	0,0827	0,3308	0,3970	0,1684	0,0211

Tabel 2

Met de grafische rekenmachine vind je dan: $\text{E}(M) \approx 1{,}71$ en $\sigma(M) \approx 0{,}91$.

Opgave 3

In **Voorbeeld 1** gaat het om een steekproef van 4 uit een populatie van 21 personen. M is het aantal meisjes in de steekproef.

a Waarom is M geen binomiale stochast?

b Bereken in vier decimalen de kans dat er minstens 3 meisjes in de steekproef voorkomen.

Opgave 4

In een vaas zitten twee witte en drie rode balletjes. Uit deze vaas worden zonder teruglegging balletjes getrokken, totdat er een wit balletje wordt getrokken.

Bereken algebraïsch de verwachting, de variantie en de standaardafwijking van het aantal benodigde trekkingen.

Opgave 5

Een gezelschap bestaat uit zes mannen en zeven vrouwen. Op een buurtfeest moet op aselecte wijze een team van vier personen uit de groep samengesteld worden om aan een spel deel te nemen.

a Welk kansmodel moet je gebruiken om de kans op een bepaald aantal mannen in de groep te berekenen?

b Noem M het aantal mannen in de geselecteerde groep en bereken $\text{P}(M < 2)$. Rond af op vier decimalen.

c Onder de mannen en vrouwen zitten 2 jongens en 4 meisjes jonger dan 16 jaar. Noem J het aantal jongens in je geselecteerde groep. Is J hypergeometrisch verdeeld? Waarom wel of niet?

d Bereken $\text{P}(J < 2)$ exact.

e Bereken $\text{E}(J)$ en $\sigma(J)$ exact.

Voorbeeld 2

Op een scholengemeenschap zitten 1200 jongens en 900 meisjes. Daaruit wordt een aselecte steekproef van vier personen getrokken. Stochast M is het aantal meisjes in de steekproef. Stel een kansverdeling op voor M en bepaal de verwachtingswaarde en de standaardafwijking van M. Laat zien dat je kansen vrijwel hetzelfde zijn als je een binomiaal kansmodel gebruikt.

Antwoord

Bij de steekproef gaat het om trekking zonder terugleggen van vier elementen uit een populatie van 2100. M is een hypergeometrische stochast.

De kans op bijvoorbeeld $M = 3$ is:

$$\text{P}(M = 3) = \frac{900}{2100} \cdot \frac{899}{2099} \cdot \frac{898}{2098} \cdot \frac{1200}{2097} \cdot 4 \approx 0{,}1798$$

Dit is vrijwel gelijk aan $\left(\frac{900}{2100}\right)^3 \cdot \frac{1200}{2100} \cdot 4 \approx 0{,}1799$.

Je kunt de kansen goed benaderen met een binomiaal kansmodel:

m	0	1	2	3	4
$\text{P}(M = m)$	0,1066	0,3199	0,3599	0,1799	0,037

Tabel 3

Nu vind je met behulp van een binomiale verdeling: $\text{E}(M) = 4 \cdot \frac{900}{2100} = \frac{12}{7}$ en

$$\sigma(M) = \sqrt{4 \cdot \frac{900}{2100} \cdot \frac{1200}{2100}} \approx 0{,}9897.$$

Opgave 6

In **Voorbeeld 2** gaat het om een steekproef van 4 uit een populatie van 2100 personen. M is het aantal meisjes in de steekproef.

a Waarom is M nog steeds geen binomiale stochast? Waarom kun je M nu wel goed benaderen met een binomiale stochast?

b Bereken de kans dat er minstens 3 meisjes in de steekproef voorkomen, met M als hypergeometrische stochast en als binomiale benadering. Rond af op vier decimalen.

Voorbeeld 3

Je hebt gelezen dat op dit moment 23% van alle Nederlandse meisjes van 12 tot en met 18 jaar rookt. Je weet dat deze groep meisjes uit ongeveer 450000 personen bestaat. Je vraagt 80 jou onbekende Nederlandse meisjes uit die leeftijdscategorie of ze roken. Hoe groot is de kans dat minstens 20 daarvan roken?

Antwoord

Hier is sprake van een steekproef uit een hele grote populatie. Hoewel in feite sprake is van een hypergeometrische stochast, kun je het aantal rokende meisjes M in de steekproef opvatten als binomiale stochast.

De gevraagde kans is daarom:

$\text{P}(M \geq 20 | n = 80 \text{ en } p = 0{,}23) = 1 - \text{P}(X \leq 19) \approx 0{,}3768$

Opgave 7

In **Voorbeeld 3** gaat het om het berekenen van kansen dat een bepaald aantal meisjes in een steekproef van 80 uit een populatie van 450000 meisjes rookt.

a Hoe moet je $\text{P}(M = 15)$ eigenlijk berekenen?

b Waarom kun je in dit geval heel goed met een binomiaal kansmodel werken?

c Neem nu een steekproefgrootte van 3 en bereken $\text{P}(M = 1)$. Doe dit zowel als hypergeometrische stochast als met een binomiale benadering. Rond af op vier decimalen.

d Beargumenteer in welke algemene situatie(s) het niet meer geoorloofd is om een binomiale benadering te gebruiken voor een hypergeometrische stochast.

Opgave 8

Van alle leerlingen uit het basisonderwijs is bekend dat 90% rechtshandig is. Hoe groot is de kans dat je in een willekeurig gekozen groep van 30 kinderen minder dan 25 rechtshandigen aantreft? Rond af op vier decimalen.

Verwerken

Opgave 9

In een klas zitten 8 jongens en 12 meisjes. Daaruit wordt een aselecte steekproef van 3 personen getrokken. Stochast J is het aantal jongens in de steekproef.

a Stel de kansverdeling voor J op. Rond de kansen af op vier decimalen.

b Bereken de verwachtingswaarde en de standaardafwijking van J. Rond af op één decimaal.

Opgave 10

In een doos zitten 20 uiterlijk allemaal dezelfde bonbons. Vijf bonbons hebben echter een roomvulling en de andere een caramelvulling. Uit de doos worden vier bonbons genomen.

a Hoe groot is de kans dat er precies één bonbon met een roomvulling uit wordt gehaald? Rond af op vier decimalen nauwkeurig.

b Hoe groot is de kans dat er twee of meer bonbons met roomvulling uit de doos worden gehaald? Rond af op vier decimalen nauwkeurig.

c Hoe groot is de kans dat de vier er uitgenomen bonbons op één na allemaal een roomvulling hebben? Rond af op vier decimalen nauwkeurig.

Opgave 11

Een partij van 1000 blikken met groente heeft lange tijd in een magazijn gelegen. Je mag aannemen dat van 10% van de blikken de uiterste verkoopdatum verstreken is. Je kiest aselect 8 blikken uit de partij en controleert de verkoopdatum. Je vraagt je af hoe groot de kans is dat je in die steekproef drie blikken aantreft die te oud zijn.

a Bereken deze kans met behulp van de hypergeometrische kansverdeling. Rond af op vier decimalen.

b Bereken deze kans ook met het binomiale kansmodel. Rond af op vier decimalen. Hoe groot is het verschil tussen beide berekeningen?

c Benader in drie decimalen de kans dat je maximaal 3 blikken gekozen hebt waarvan de uiterste verkoopdatum verstreken is.

Opgave 12

Een grote partij wijnflessen wordt gekeurd door uit de partij een aselecte steekproef van 20 flessen te nemen. Elke fles wordt nauwkeurig onderzocht op gebreken. Wordt er in de steekproef meer dan één fles gevonden met een gebrek, dan wordt de hele partij afgekeurd. Anders wordt de hele partij goedgekeurd.

a 5% van de hele partij flessen heeft gebreken. Hoe groot is de kans dat de partij wordt goedgekeurd? Rond af op vier decimalen.

b Acht afnemers nemen ieder afzonderlijk een steekproef van 20 flessen uit deze grote partij.

Hoeveel afnemers keuren de partij naar verwachtingswaarde af? Rond af op gehelen.

Opgave 13

Bij het spel hartenjagen krijgt elke speler 13 kaarten toegedeeld. Als een speler van één kleur vier kaarten krijgt en van de andere drie kleuren ieder drie kaarten, wordt dat een vlakke verdeling genoemd.
Bereken de kans op een vlakke verdeling. Rond af op vier decimalen.

Opgave 14

In een vaas zitten vijf balletjes genummerd 2, 4, 6, 8 en 10. Er worden zonder teruglegging twee balletjes uit de vaas getrokken. Stochast V is het verschil van de nummers van de twee balletjes.

a Stel de kansverdeling voor V op.

b Bereken zonder grafische rekenmachine de verwachtingswaarde, de variantie en de standaardafwijking van V.

Toepassen

Opgave 15: Politieke partij

Het bestuur van een politieke partij bestaat uit 20 personen, waarvan 40% jonger is dan 28 jaar. Door het lot worden 4 personen aangewezen om deel te nemen aan een buitenlandse reis.

a Hoeveel personen van de groep van 4 zijn naar verwachtingswaarde jonger dan 28 jaar?

b Bepaal de kans dat drie van de vier personen jonger zijn dan 28 jaar. Rond af op vier decimalen.

Tijdens een landelijke bijeenkomst van diezelfde partij zijn n leden aanwezig, met $n > 20$. Van deze leden is 40% jonger dan 35 jaar. Door het lot worden 3 personen aangewezen om deze landelijke groepering te vertegenwoordigen.

c Toon aan dat de kans dat twee van de drie afgevaardigden jonger zijn dan 35 jaar gelijk is aan: $\frac{0{,}72n(0{,}4n-1)}{(n-2)(n-1)}$

d Geef de binomiale benadering van de gevraagde kans in c. Vanaf welke n is het verschil tussen de eigenlijke kans en de binomiale benadering kleiner dan 0,001?

Opgave 16: Steaks

Van een lading van meer dan 20 steaks is bekend dat er twee bedorven zijn. Er worden willekeurig vijf steaks uit de lading gekozen. Twee ervan blijken bedorven. Wat is de minimale grootte van de lading steaks zodanig dat de kans op deze hypergeometrisch verdeelde gebeurtenis hooguit 0,25% verschilt met de binomiale benadering?

Testen

Opgave 17

In een vaas zitten 12 rode en 18 gele balletjes. Joran neemt een greep van 4 balletjes uit de vaas. Stochast G is het aantal gele balletjes in de greep van 4 balletjes.

a Leg uit waarom stochast G niet binomiaal verdeeld is.

b Maak de kansverdeling van stochast G.

c Hoeveel gele balletjes kan Joran in zijn greep van 4 balletjes verwachten?

Opgave 18

Neem aan dat 98% van de Nederlandse bevolking een ziektekostenverzekering hebben.

Hoe groot is de kans dat bij een groep van 20 Nederlanders zich hoogstens twee personen bevinden die geen ziektekostenverzekering hebben?

Opgave 19

Bij de Lotto moet je zes getallen raden en die in de juiste volgorde hebben gezet. In een ronddraaiende trommel zitten 41 balletjes met daarop de getallen 1 tot en met 41. Zesmaal wordt een balletje getrokken om de uitslag te verkrijgen.

a Hoe groot is de kans dat er zes even nummers worden getrokken?

b Als er twee even nummers zijn getrokken, hoe groot is dan nog de kans dat de volgende vier balletjes ook een even nummer hebben?

c Hoe groot is de kans, dat elk van de zes getrokken getallen kleiner is dan 15?

Jos heeft de nummers 5, 10, 15, 20, 25 en 30 op zijn formulier aangekruist.

d Hoe groot is de kans dat hij ze alle zes goed heeft?

Deze paragraaf is een onderdeel van het Math4All wiskundemateriaal.

Math4All stelt het op prijs als onvolkomenheden in het materiaal worden gemeld en ideeën voor verbeteringen in de content of dienstverlening kenbaar worden gemaakt.

Email: f.spijkers@math4all.nl

Met de Math4All maatwerkdienst kunnen complete readers worden samengesteld en toetsen worden gegenereerd. Docenten kunnen bij a.f.otten@xs4all.nl een gratis inlog voor de maatwerkdienst aanvragen.