Statistiek MBW/KW deel 1 week 5

- Normale verdeling: Passingsproblemen
- Normale verdeling: Centrale limietstelling
- Poissonverdeling

Passingsproblemen H5.5

Een jamfabrikant vult lege potten met een vulapparaat

Vulgewicht jam per pot: $\underline{x} \sim N(\mu_1 = 505\ ;\ \sigma_1 = 5)$ in gram.

Volume van pot: $\underline{y} \sim N(\mu_2 = 513\ ;\ \sigma_2 = 7)$ in gram.

Hoe groot is de kans dat een hoeveelheid jam niet in een potje past? Dit is een voorbeeld van een **passingsprobleem**.

De kans dat het **niet** past is $P\left(\underline{x} > \underline{y}\right)$. Hieraan kunnen wij niet rekenen, omdat er **twee** kansvariabelen bij betrokken zijn.

Truc: Gebruik de verschilvariabele $\underline{v} = \underline{x} - \underline{y}$

$$P\left(\underline{x} > \underline{y}\right) = P\left(\underline{v} = \underline{x} - \underline{y} > 0\right)$$

$\underline{v}$ is normaal verdeeld (rekenregels) met

$\mu_v = 505 - 513 = -8$ gram en

$\sigma_v = \sqrt{\sigma_1^2 + \sigma_2^2} = 8{,}6023$ gram, dus

$P\left(\underline{x} > \underline{y}\right) = P\left(\underline{v} > 0\right) = \text{normalcdf}(0\,, 10^{10}\,, -8\,, 8.6023) =$ 0,1762.

Passingsproblemen H5.5

Een jamfabrikant vult lege potten met een vulapparaat
Vulgewicht jam per pot: $\underline{x} \sim N(\mu_1 = 505\ ;\ \sigma_1 = 5)$ in gram.
Volume van pot: $\underline{y} \sim N(\mu_2 = 513\ ;\ \sigma_2 = 7)$ in gram.

De kans dat het niet past is $P\left(\underline{x} > \underline{y}\right) = 0,1762$.

Kan deze kans worden verkleind tot 0,07 door nauwkeurigere potten met een kleinere σ_2 te gebruiken?

Stel $\sigma_2 = s$ en los de vergelijking:

$$\text{normalcdf}\left(0\,,10^{10}\,,-8\,,\sqrt{5^2+s^2}\right) = 0.07$$

op voor s met de Solver op de GR. Dit levert $\sigma_2 = 2{,}0941$.

Kan deze kans vervolgens nog verder worden verkleind tot 0,05 door nog nauwkeurigere potten te gebruiken?

$$\text{normalcdf}\left(0\,,10^{10}\,,-8\,,\sqrt{5^2+s^2}\right) = 0.05.$$

De solver geeft geen oplossing. Dat komt omdat de kleinst mogelijke waarde $s = 0$ is. Dit zou een kans leveren van

$$\text{normalcdf}\left(0\,,10^{10}\,,-8\,,\sqrt{5^2+0^2}\right) = 0{,}0548 > 0{,}05$$

Centrale limietstelling

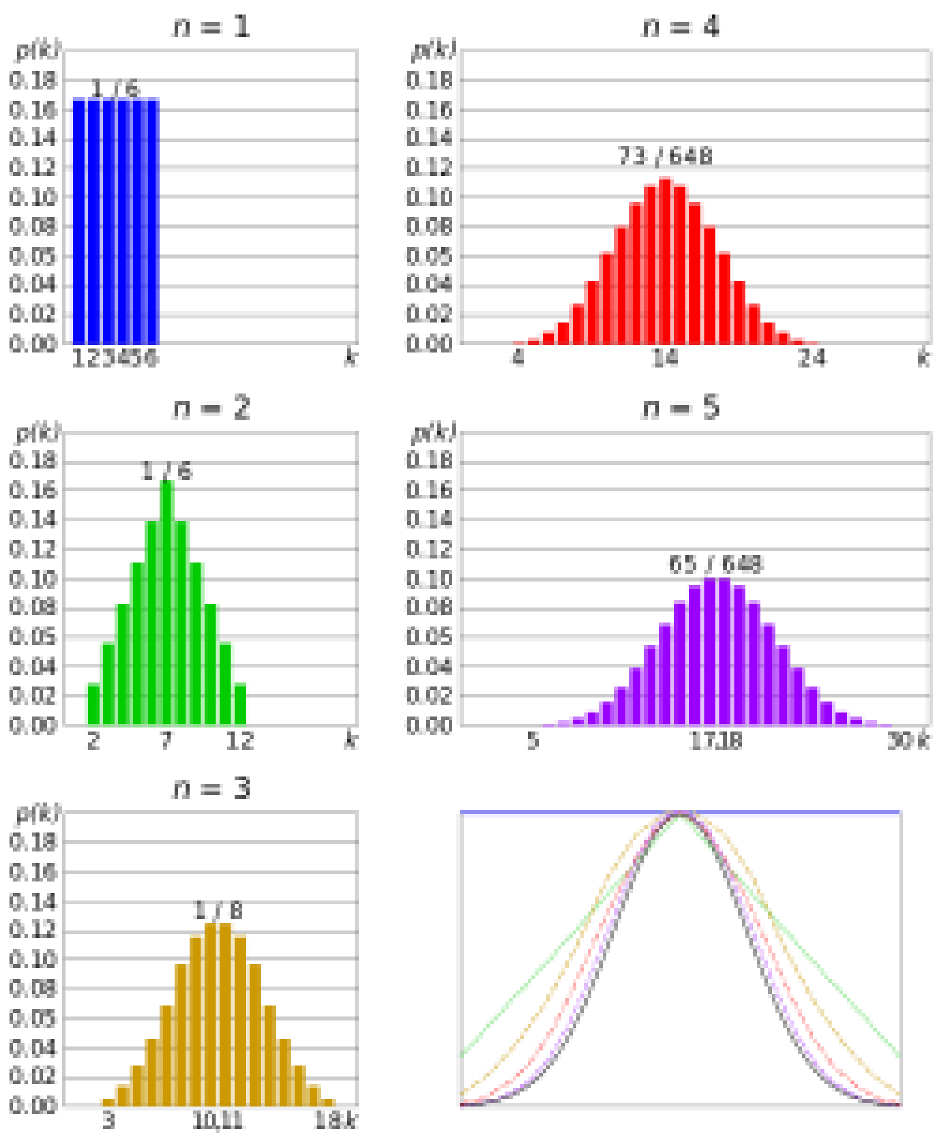
Centrale limietstelling:

1. Het <u>steekproefgemiddelde</u> van n getallen uit een willekeurige kansvariabele $\underline{x}$ met gemiddelde μ en standaarddeviatie σ gaat zich voor grotere n steeds meer gedragen als een normale verdeling met gemiddelde μ en standaarddeviatie $\frac{\sigma}{\sqrt{n}}$.

2. De <u>som</u> van n waarden uit die kansverdeling gedraagt zich voor grotere waarden van n als normaal verdeeld met gemiddelde $n\mu$ en standaarddeviatie $\sqrt{n}\sigma$.

Belangrijk is dat de precieze verdeling van $\underline{x}$ niet bekend hoeft te zijn. Het berekenen van de kansverdeling van een som of gemiddelde van een bepaalde kansverdeling is i.h.a. moeilijk of ondoenlijk. De centrale limietstelling geeft dan de mogelijkheid om toch iets uit te rekenen.

Opmerking: Als $\underline{x}$ zelf normaal verdeeld is, dan geldt de centrale limietstelling zelfs exact voor elke n.



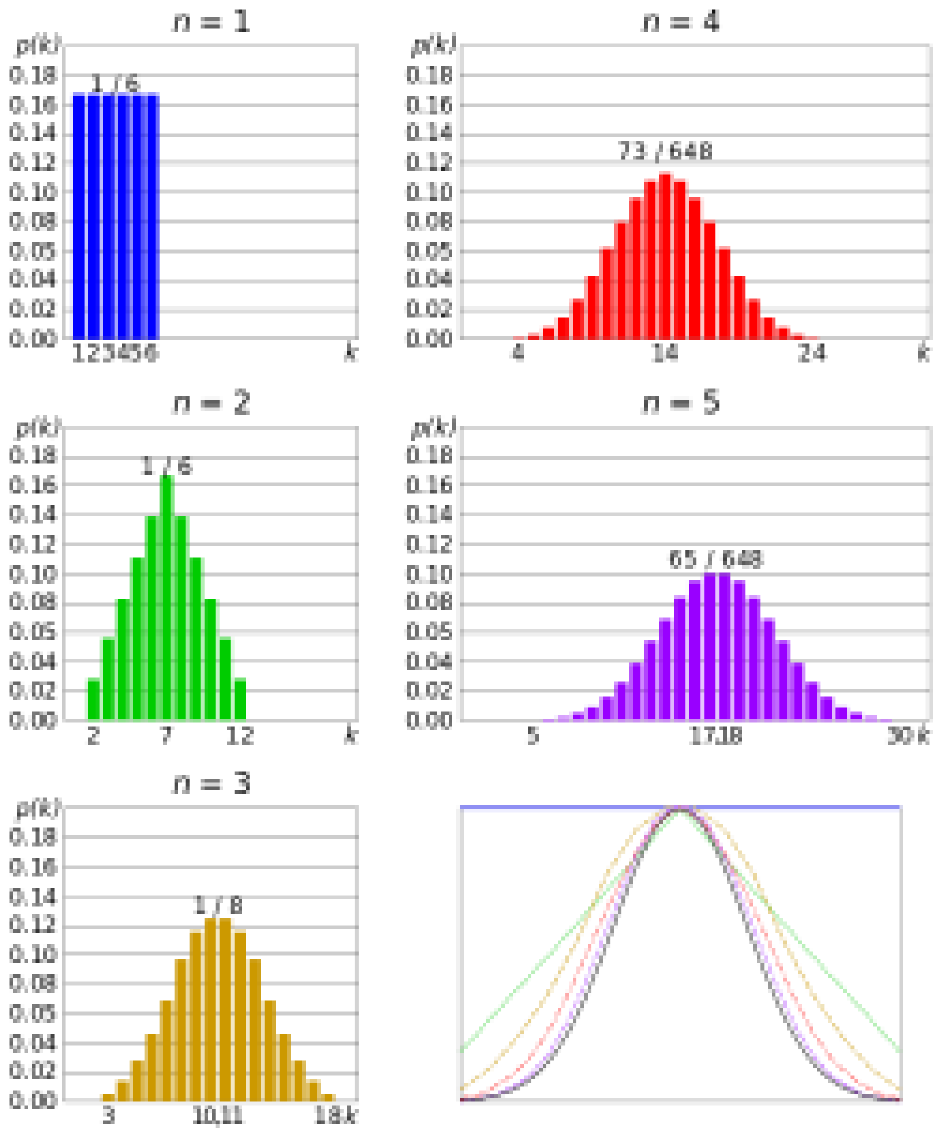
Centrale limietstelling Voorbeeld

De **centrale limietstelling** zegt dat de steekproefsom van een willekeurige kansvariabele $\underline{x}$ met gemiddelde μ en standaarddeviatie σ zich voor grotere aantallen n steeds meer gaat gedragen als een normale verdeling met gemiddelde $n\mu$ en standaarddeviatie $\sqrt{n}\sigma$.

Voorbeeld: Het aantal telefoons dat een winkel van een bepaald model mobiele telefoon verkoopt is uniform (continu) verdeeld tussen 10 en 26 stuks per week. Wat is de kans dat er in een jaar (52 weken) niet meer dan 940 worden verkocht?

Voor de continue uniforme verdeling is de gemiddelde waarde $\mu = \frac{(10+26)}{2} = 18$ en standaarddeviatie $\sigma = \frac{(26-10)}{\sqrt{12}} = 4{,}6188.$
Volgens de centrale limietstelling gedraagt het totale aantal per jaar zich als normaal verdeeld met gemiddelde $52 \cdot 18 = 936$ en standaarddeviatie $\sqrt{52} \cdot 4{,}6188 = 33{,}3066.$
De gevraagde kans is:

$$\text{normalcdf}(-10^{10}, 940, 936, 33.3066) = 0{,}5478.$$



Poissonverdeling Voorbeeld IED

Improvised Explosive Device

= Bermbom

Irrelevante vraag: Is dit een actieopname van een IED die toevallig net afgaat?

https://www.defensie.nl/onderwerpen/taken-in-nederland/explosieven-ruimen/expertisecentra

Poissonverdeling Voorbeeld IED

Terrorist Who Put A Lot Of Work Into Explosive Device Offended By Intelligence Agencies Labeling It As ‘Improvised’

10/08/19 11:55AM • SEE MORE: TERRORISM

https://www.theonion.com/terrorist-who-put-a-lot-of-work-into-explosive-device-o-1838878632

Poissonverdeling Voorbeeld IED

GALGALA MOUNTAINS, SOMALIA—Deeply hurt by the way in which counterterrorism operatives repeatedly discounted the craftsmanship and ingenuity of his anti-personnel bombs, terrorist Ahmad Musa stated Tuesday he resented Western intelligence agencies referring to his explosive devices as "improvised." "I refined the design and construction of this car bomb over a period of weeks, filling two graph paper tablets with notes and staying up all night reading physics textbooks, yet they treat my work like I just threw a bunch of roofing nails in a pressure cooker and called it a day," Musa said of the explosive, which involved extensive work with fertilizer chemistry, careful deciphering and adapting wiring diagrams, and even countless practice sessions with soldering irons. "Devising the right detonator rig was a huge undertaking—did you know flame travels at different speeds through different densities of diesel-fuel vapors? It totally does. Plus, I had to be sure the shrapnel was the right shape to disperse correctly. Yet everyone from the CIA to Interpol continues to use the cute little dismissive acronym 'IED.' Insulting. Everyone knows improvisation is amateurish, and no one in their right mind respects anything improvised. My device killed 67 people—does that sound improvised to you?" Musa promised that his magnificent suicide bombing next month would force authorities to report his demise as "death by elegantly devised explosive device."

Poissonverdeling Voorbeeld IED

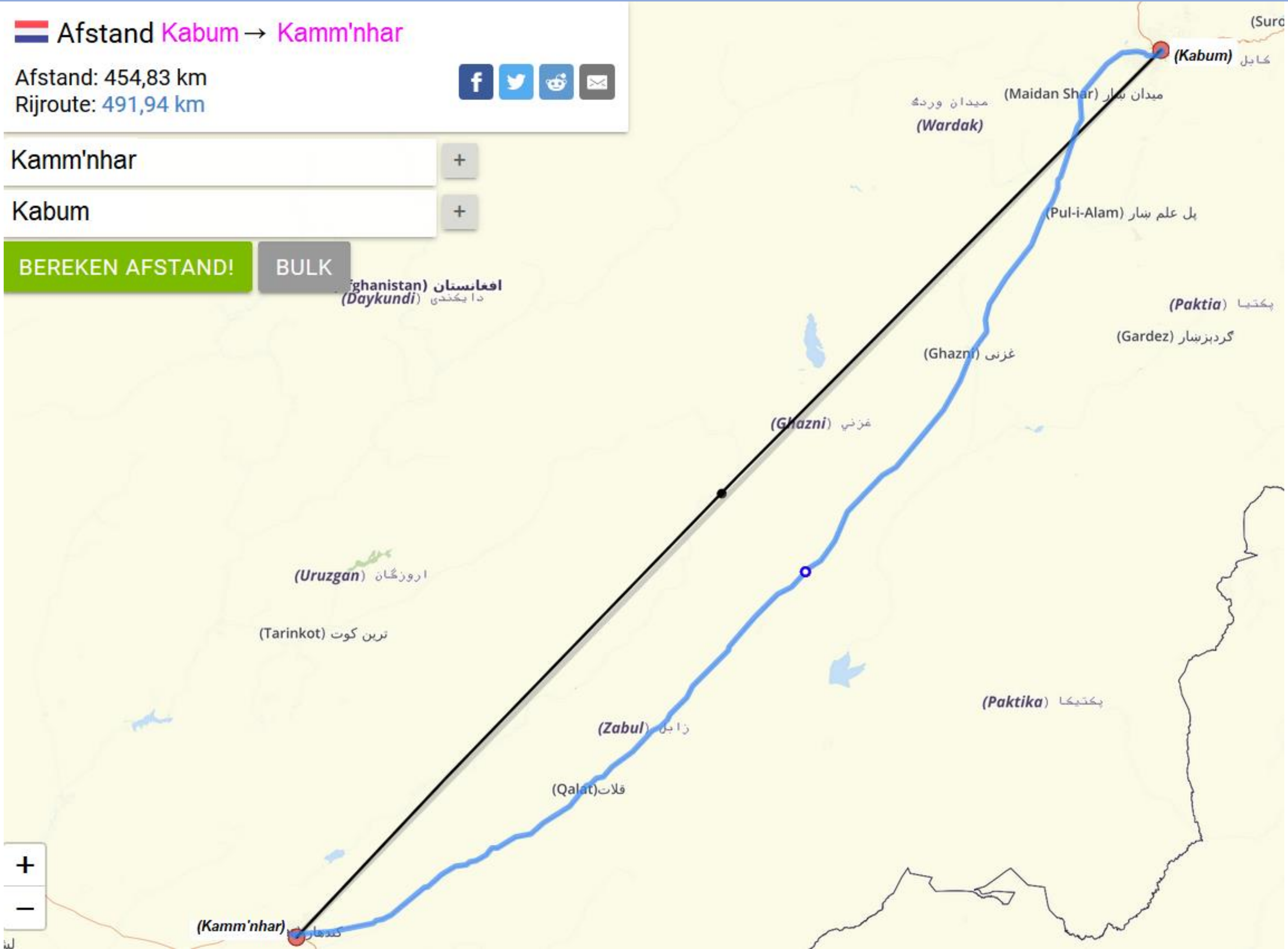
Situatieschets IED's.

- ✓ **Situatie**: Vredesmissie in Kamm'nhar (Talibanon)
- ✓ **Opdracht**: Pelotonscommandant, opdracht tot verplaatsing naar hoofdstad Kabum
- ✓ **Afstand**: 491,94 km
- ✓ **Reistijd**: 8 uur 59 minuten
- ✓ **Intelligence info**: Gemiddeld 2 IED's per 1000 km
- ✓ **Probleem**: Wat is de kans dat je met een IED te maken krijgt?

Antwoord:

a. Meer dan 100%
b. Vrijwel 100%
c. Tussen 50 en 100%
d. Ongeveer 50%
e. Tussen 0 en 50%
f. Vrijwel nihil
g. Bij Defensie krijg je nooit meerkeuzeantwoorden

Poissonverdeling Voorbeeld IED

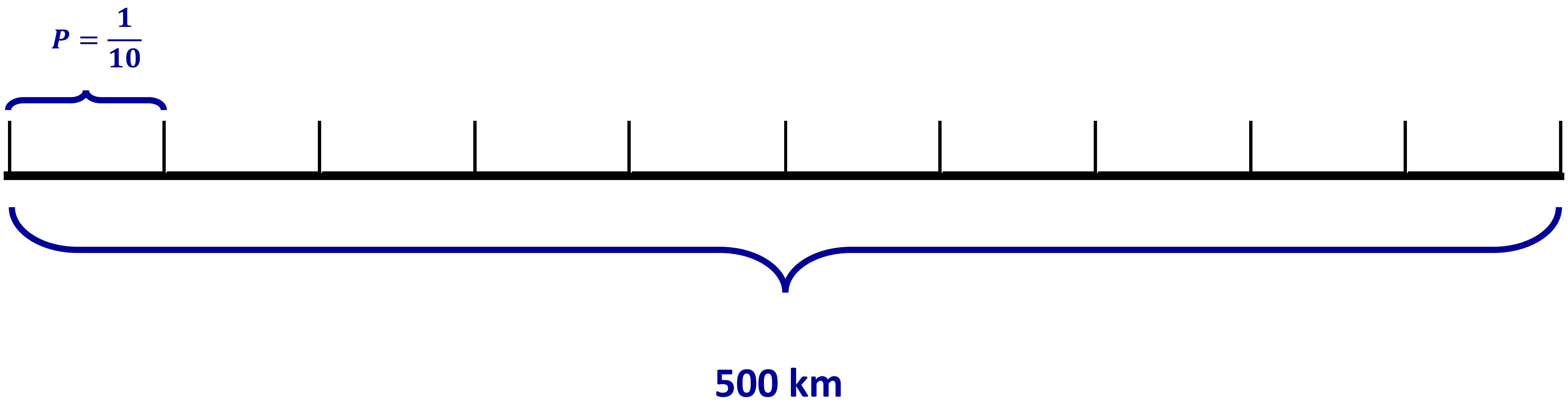


Oplossingsstrategie: Afstand discretiseren - achtergrond geen tentamenstof

Gemiddeld 2 IED's per 1000 km, dus op 500km gemiddeld 1 IED.

Verdeel de 500 km in 10 stukken van 50 km.

Kans op een IED per 50 km is $P = 2 \cdot \frac{50}{1000} = \frac{1}{10}$



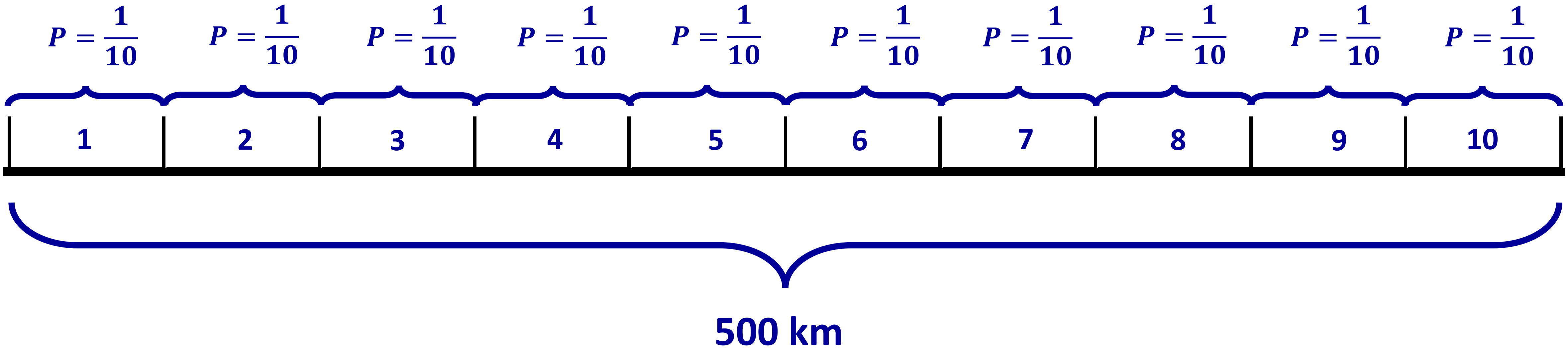
Poissonverdeling Voorbeeld IED **achtergrondinfo, geen tentamenstof**

Oplossingsaanpak: Discretiseren

Verdeel de 500 km in 10 gelijke stukken.
Wat is de kans op precies één IED in deze 500 km?
Deze IED kan in vak nr. 1 liggen, maar dan geen in vak 2, 3, …,10.

Kans daarop is $\frac{1}{10} \cdot \left(1 - \frac{1}{10}\right) \cdots \left(1 - \frac{1}{10}\right) = \frac{1}{10} \cdot \left(1 - \frac{1}{10}\right)^9$

Kans op één IED in precies één van de 10 vakken: $10 \cdot \frac{1}{10} \cdot \left(1 - \frac{1}{10}\right)^9 = \left(1 - \frac{1}{10}\right)^9$



Oplossingsaanpak: Generaliseren

Algemeen: Verdeel de 500 km in n gelijke stukken.
Wat is de kans op precies één IED in deze 500 km?
Dat is n keer de kans op één IED alleen in precies één 500/n km vak :

$$n \cdot \frac{1}{n} \cdot \left(1 - \frac{1}{n}\right) \cdots \left(1 - \frac{1}{n}\right) = \boldsymbol{\left(1 - \frac{1}{n}\right)^{n-1}}$$

Klopt voor $n = 10$, dus klopt voor elke n. (☺)
Wat gebeurt er als n groot wordt?

Je zou kunnen zeggen: $1 - \frac{1}{n} \rightarrow 1$ en $1^{n-1} \rightarrow 1$, dus die kans wordt op den duur 1.
Dus 100% kans dat er een IED ligt op dit traject!

Oplossingsaanpak: Controleren

Maar dat klopt niet:

Reken uit: $\left(1-\frac{1}{n}\right)^{n-1}$ voor $n = 10,\ 100,\ 1000000, \ldots$

$n = 10 \ \rightarrow \ \left(1-\frac{1}{10}\right)^{10-1} = 0{,}9^{9} \qquad = 0{,}387420 \ldots$

$n = 100 \ \rightarrow \ \left(1-\frac{1}{100}\right)^{100-1} = 0{,}99^{99} \qquad = 0{,}369729 \ldots$

$n = 1000000 \ \rightarrow \ \left(1-\frac{1}{1000000}\right)^{1000000-1} = 0{,}999999^{999999} \qquad = 0{,}367879 \ldots$

Bereken $\ln(0{,}367879)$

= -0,9999995… ≈ -1

De kans op precies één IED is niet 1, maar $e^{-1} = 0{,}367879441 \ldots$, dus **36,8%.**

Poissonverdeling Voorbeeld IED **achtergrondinfo, geen tentamenstof**

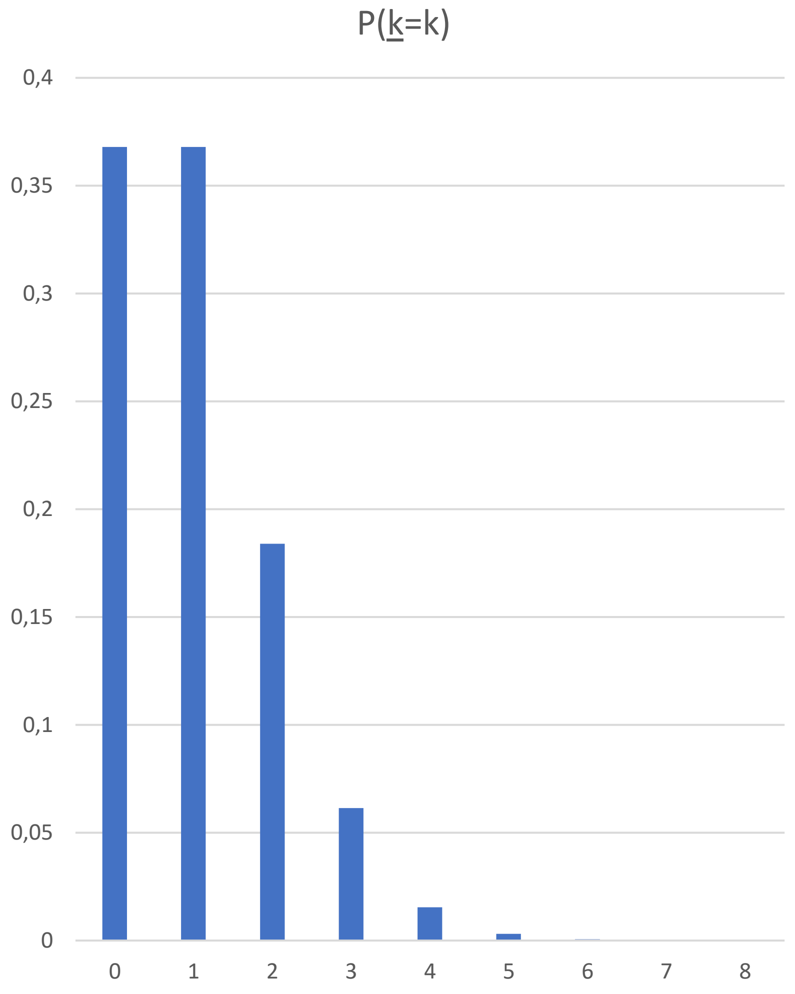
Oplossingsaanpak: Generaliseren

We weten nu wat de kans op precies één bermbom is:

$$P(\underline{k} = 1) = e^{-1}$$

Algemeen blijkt de kans dat er k IED's langs de weg liggen:

$$P(\underline{k} = k) = \frac{1}{k!} e^{-1}$$



Poissonverdeling Voorbeeld IED

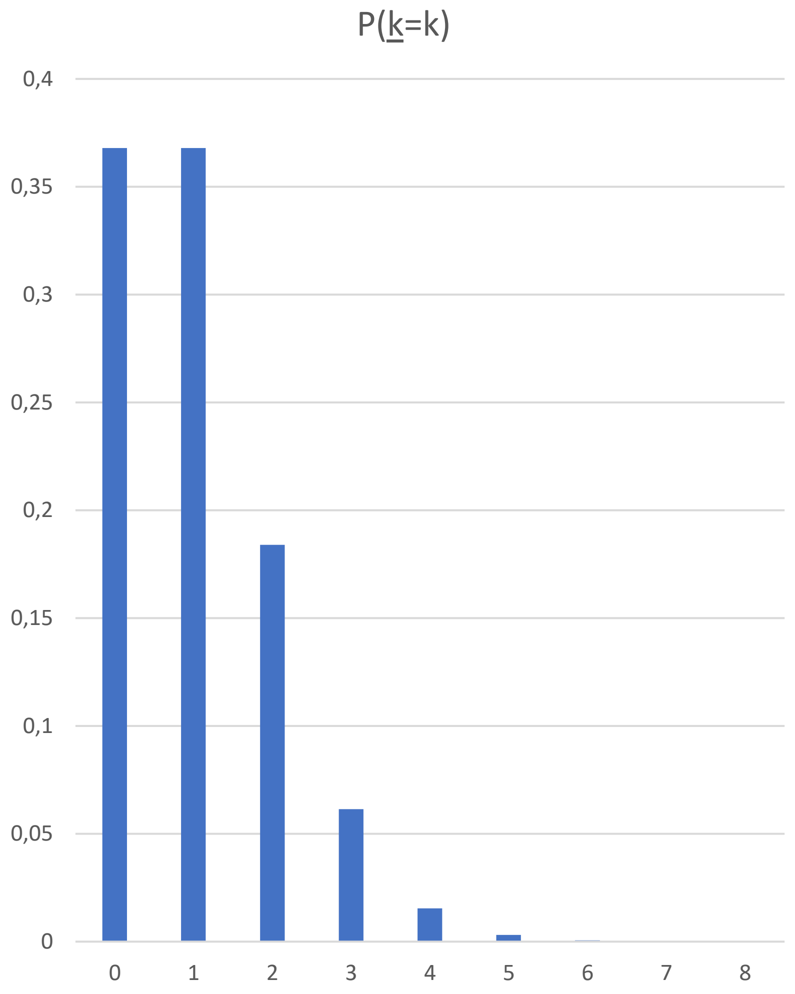
Is dit de goede oplossing?

De kans op (precies) <u>één</u> IED is:

$$P(\underline{k} = 1) = e^{-1} = 0{,}3679$$

Wat is nu de kans op <u>een</u> (spreek uit: úhn) IED?

$$P(\underline{k} \geq 1) = 1 - P(\underline{k} = 0) = 1 - e^{-1} = 0{,}6321$$



Poissonverdeling Voorbeeld IED

Stel, je bent al een IED tegengekomen, wat is dan de kans op nog een IED?

$$P(\underline{k} \geq 2) = 1 - P(\underline{k} < 2) = 1 - e^{-1} - e^{-1} = 0{,}26424?$$

Maar niet heus, conditionele kans gebruiken!

$$P(\underline{k} \geq 2 \mid \underline{k} \geq 1) = \frac{P(\underline{k} \geq 2)}{P(\underline{k} \geq 1)} = \frac{1 - \frac{2}{e}}{1 - \frac{1}{e}} = \frac{e-2}{e-1} = 0{,}41802$$

Poissonverdeling Definitie

Poissonverdeling (1838). Voor een verschijnsel dat gemiddeld μ keer per (tijd- of plaats)eenheid plaatsvindt is de kans op k successen in één eenheid:

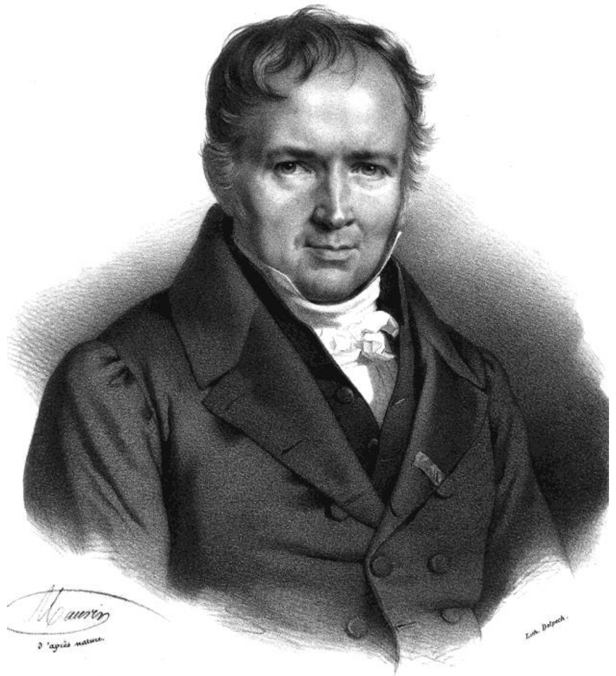
$$P(\underline{k} = k) = \frac{\mu^k}{k!} e^{-\mu}$$

Voorwaarde is dat het verschijnen geheugenloos is, d.w.z. niet afhangt van eerdere optredens.

Voorbeeld 1: Het IED voorbeeld. **Het kiezen van de eenheid is bij de Poissonverdeling belangrijk.** Hier is de rekeneenheid 500 km. Per 1000 km zijn er gemiddeld 2 voorvallen,

dus $\mu = 2 \cdot \frac{500}{1000} = 1$ (IED per 500 km) en de kansfunctie is

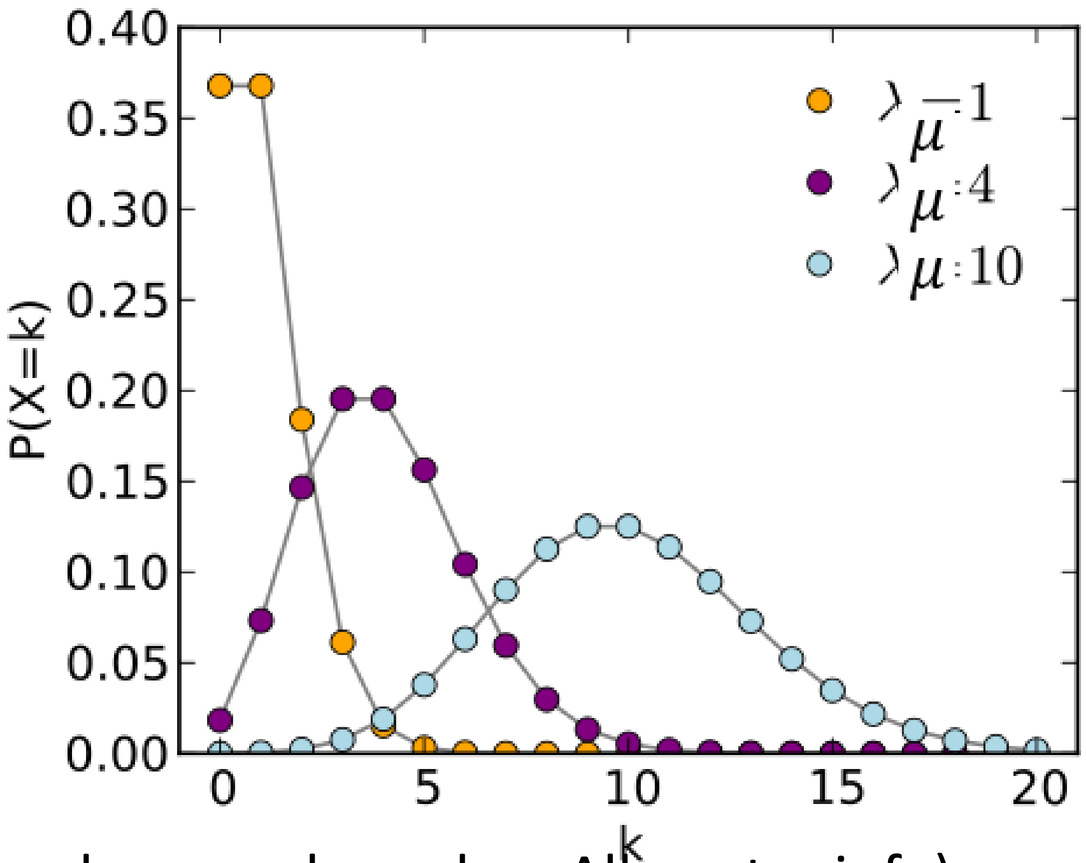
$$P(\underline{k} = k) = \frac{1}{k!} e^{-1}$$



Siméon Poisson (1781-1840)

Poissonverdeling Eigenschappen

De Poissonverdeling is een <u>discrete</u> kansverdeling met één parameter μ.



Kansfunctie:

$$P(\underline{k} = k) = \frac{\mu^k}{k!} e^{-\mu}$$

Gemiddelde waarde: μ

Standaarddeviatie: $\sqrt{\mu}$

Mediaan: $\approx \left\lfloor \mu + \frac{1}{3} - \frac{0{,}02}{\mu} \right\rfloor$ (afronden naar beneden. Alleen ter info)

Poissonverdeling is de limiet van een binomiale verdeling, waarbij n groot is en het verwachte aantal successen $np = \mu$ constant is. Het beschrijft verschijnselen die niet vaak, maar wel geregeld en onafhankelijk van elkaar optreden.

Poissonverdeling Voorbeeld 2: Doodgetrapt door Pruisisch paard

Pruisische Rode Huzaren (schilderij Christian Sell)

Rond 1900 bestond een Pruisisch legerkorps uit 1554 officieren, 43.317 manschappen, 16.934 paarden, 2.933 voertuigen.

Poissonverdeling Voorbeeld 2: Doodgetrapt door Pruisisch paard

24 Zweites Kapitel. § 12.

	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94
G	—	2	2	1	—	—	1	1	—	3	—	2	1	—	—	1	—	1	—	1
I	—	—	—	2	—	3	—	2	—	—	—	1	1	1	—	2	—	3	1	—
II	—	—	—	2	—	2	—	—	1	1	—	—	2	1	1	—	—	2	—	—
III	—	—	—	1	1	1	2	—	2	—	—	—	1	—	1	2	1	—	—	—
IV	—	1	—	1	1	1	1	—	—	—	—	1	—	—	—	—	1	1	—	—
V	—	—	—	—	2	1	—	—	1	—	—	1	—	1	1	1	1	1	1	—
VI	—	—	1	—	2	—	—	1	2	—	1	1	3	1	1	1	—	3	—	—
VII	1	—	1	—	—	—	1	—	1	1	—	—	2	—	—	2	1	—	2	—
VIII	1	—	—	—	1	—	—	1	—	—	—	—	1	—	—	—	1	1	—	1
IX	—	—	—	—	—	2	1	1	1	—	2	1	1	—	1	2	—	1	—	—
X	—	—	1	1	—	1	—	2	—	2	—	—	—	—	2	1	3	—	1	1
XI	—	—	—	—	2	4	—	1	3	—	1	1	1	1	2	1	3	1	3	1
XIV	1	1	2	1	1	3	—	4	—	1	—	3	2	1	—	2	1	1	—	—
XV	—	1	—	—	—	—	—	1	—	1	1	—	—	—	2	2	—	—	—	—

Aantal doden als gevolg van een paardentrap voor 14 cavalerieregimenten van 1875-1894 (*Ladislaus von Bortkiewicz (1898) Das Gesetz der kleinen Zahlen Teubner, Leipzig*)

Aantal doden	Waargenomen frequentie	Berekende frequentie
0	144	139,05
1	91	97,33
2	32	34,07
3	11	7,95
4	2	1,39
5	0	0,19

Totaal: 1x92+2x32+3x11+4x2=196 ongevallen in 20 jaar.

Eenheid: één jaar in één regiment

$$\mu = \frac{196}{20 \cdot 14} = 0{,}7$$

$$P(K = k) = \frac{0{,}7^k}{k!} e^{-0{,}7}$$

Poissonverdeling Voorbeeld 3: WW2 blindgangers

BBC | Sign in | News | Sport | Reel | Worklife | Travel | Future | More

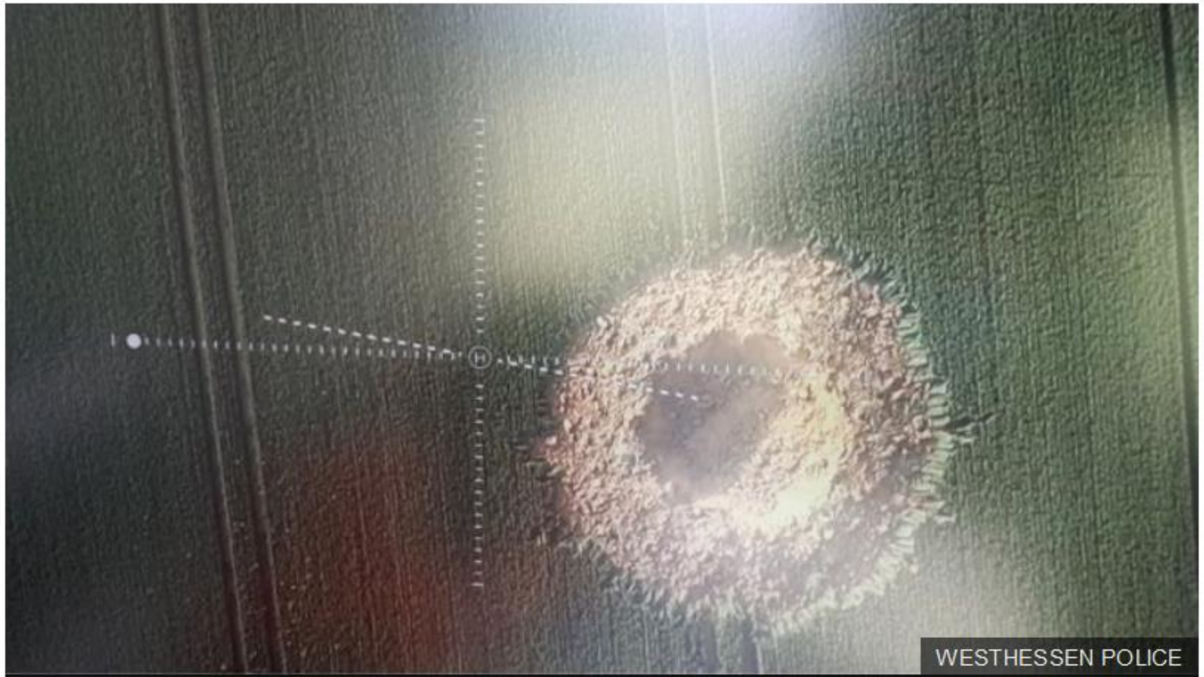
NEWS

German WW2 bomb leaves giant crater in field

24 June 2019

Share

World War Two



A drone showed the impact of the blast

Poissonverdeling Voorbeeld 3: WW2 blindgangers

It was not immediately clear what caused a blast so powerful that it registered as a minor earthquake.

The blast, at 03:52 (01:52 GMT) on Sunday, startled residents near the central German town of Limburg, leaving a crater 10m (33ft) wide and four metres deep in a field.

A photo taken by a drone later revealed the impact of the night-time explosion.

Initially police said "no definitive indication" had been found of a suspected unexploded bomb.

But, on close inspection of the corn field in Ahlbach, bomb disposal experts decided it was "with almost absolute certainty" a World War Two bomb. They believed it was a 250kg (550lb) bomb dropped by a plane.

Experts said it was not surprising that no remains of the bomb had been found

Poissonverdeling Voorbeeld 3: WW2 blindgangers

Voorbeeld 3a: In Duitsland ontploft elk jaar gemiddeld één blindganger uit WO II.
Er is er in 24-06-2019 één afgegaan. Wat is de kans dat er in 2019 nóg een is afgegaan?

Periode is 1 jaar, $\mu = 1$ explosie per jaar, $\underline{k}$ is het aantal explosies in een bepaald jaar.
Bereken de conditionele kans $P\left(\underline{k} \geq 2 \,\middle|\, \underline{k} \geq 1\right)$

$$P\left(\underline{k} \geq 2 \,\middle|\, \underline{k} \geq 1\right) = \frac{P(\underline{k} \geq 2)}{P(\underline{k} \geq 1)} = \frac{1 - P(\underline{k} \leq 1)}{1 - P(\underline{k} = 0)} = \frac{1 - \frac{2}{e}}{1 - \frac{1}{e}} = 0{,}4180,$$

of met de GR, Poissoncdf:

$$\frac{1 - \text{Poissoncdf}(\mu = 1, k = 1)}{1 - \text{Poissoncdf}(\mu = 1, k = 0)} = 0{,}4180$$

Poissonverdeling Voorbeeld 3: WW2 blindgangers

Voorbeeld 3b: In Duitsland ontploft elk jaar gemiddeld één blindganger uit WO II. Er is er in 24-06-2019 één afgegaan. Wat is de kans dat er **later** dat jaar nóg een is afgegaan?

Wat is de kans dat er van 25-06 t/m 31-12-2019 nog (minstens) een is afgegaan?

Er zijn nog 5+31+31+30+31+30+31 = 189 dagen te gaan van de 366 (schrikkeljaar).
Periode is 189/366 = 0,5164 jaar, $\mu = 0{,}5164$ explosies per 189 dagen.

$$P(\underline{k} \geq 1) = 1 - P(\underline{k} = 0) = 1 - \text{Poissonpdf}(\mu = 0{,}5164, k = 0) = 0{,}4033.$$

Kans op precies één extra explosie is

$$P(\underline{k} = 1) = \text{Poissonpdf}(\mu = 0{,}5164, k = 1) = 0{,}3081.$$

Poissonverdeling Voorbeeld 3: WW2 blindgangers

Voorbeeld 3c: In Duitsland ontploft elk jaar gemiddeld één blindganger uit WO II.
Er is er in 24-06-2019 één afgegaan. Wat is de kans dat dit de eerste van dit jaar was?

$$P(\text{Geen explosie t/m } 23-06) = \text{Poissonpdf}\left(\mu = \frac{175}{366}, k = 0\right) = 0{,}6200$$

Voorbeeld 3d: In Duitsland zijn in een jaar twee blindgangers geëxplodeerd. Wat is de kans dat de tweede explosie als eerste plaatsvond?