**Wat is de stelling van Bayes?**

De stelling van Bayes is een manier om [voorwaardelijke waarschijnlijkheid](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/conditional-probability-definition-examples/) te achterhalen . Voorwaardelijke waarschijnlijkheid is de waarschijnlijkheid dat een gebeurtenis plaatsvindt, aangezien deze een relatie heeft met een of meer andere gebeurtenissen. Uw kans om een ​​parkeerplaats te krijgen, hangt bijvoorbeeld samen met het tijdstip waarop u parkeert, waar u parkeert en welke conventies er op elk moment plaatsvinden. De stelling van Bayes is iets genuanceerder. In een notendop geeft het u de werkelijke [kans](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/probability-and-statistics/probability-main-index/) op een gebeurtenis met informatie over tests.

* 'Gebeurtenissen' zijn anders dan 'tests'. Er is bijvoorbeeld een test voor leverziekte, maar die staat los van het feit dat je daadwerkelijk een leverziekte hebt.
* Tests zijn gebrekkig: alleen omdat u een positieve test heeft, betekent niet dat u de ziekte ook daadwerkelijk heeft. Veel tests hebben een hoog [percentage vals-positieven](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/false-positive-definition-and-examples/) . Zeldzame gebeurtenissen hebben de neiging om hogere percentages vals-positieven te hebben dan meer algemene gebeurtenissen. We hebben het hier niet alleen over medische tests. Spamfilters kunnen bijvoorbeeld hoge percentages valse positieven hebben. De stelling van Bayes neemt de testresultaten en berekent uw *reële waarschijnlijkheid* dat de test de gebeurtenis heeft geïdentificeerd.

**De Formule**

Bekijk de video voor een snel voorbeeld van het werken aan een Bayes 'Theorem-probleem, of lees de onderstaande voorbeelden:

https://youtu.be/WAbndEzC2po

De stelling van Bayes (ook bekend als de regel van Bayes) is een bedrieglijk eenvoudige formule die wordt gebruikt om [voorwaardelijke waarschijnlijkheid](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/conditional-probability-definition-examples/) te berekenen . De stelling is vernoemd naar de Engelse wiskundige Thomas Bayes (1701-1761). De formele definitie van de regel is: In de meeste gevallen kun je niet zomaar getallen in een vergelijking stoppen; U moet eerst uitzoeken wat uw "tests" en "gebeurtenissen" zijn.



 Voor twee gebeurtenissen, A en B, stelt de stelling van Bayes je in staat om p (A | B) (de kans dat gebeurtenis A plaatsvond, gegeven dat test B positief was) uit p (B | A) (de kans dat test B gebeurde, aangezien gebeurtenis A plaatsvond).

Het kan een beetje lastig zijn om je hoofd eromheen te wikkelen, omdat je technisch gezien achteruit werkt; het kan zijn dat u uw tests en gebeurtenissen moet omwisselen, wat verwarrend kan worden.

Een voorbeeld zou moeten verduidelijken wat ik bedoel met 'schakel de tests en gebeurtenissen om'.

**Voorbeeld 1**

U bent misschien geïnteresseerd in het achterhalen van de kans dat een patiënt een leveraandoening heeft als hij alcoholist is. "Alcoholist zijn" is de test (een soort lakmoesproef) voor leverziekte.

* **A** zou de gebeurtenis kunnen betekenen: "Patiënt heeft een leverziekte." Gegevens uit het verleden vertellen u dat 10% van de patiënten die uw kliniek binnenkomen, een leveraandoening heeft. P (A) = 0,10.
* **B** zou de lakmoesproef kunnen betekenen: "Patiënt is alcoholist." Vijf procent van de patiënten in de kliniek is alcoholist. P (B) = 0,05.
* U weet misschien ook dat van de patiënten bij wie een leverziekte is vastgesteld, 7% alcoholist is. Dit is jouw **B | A** : de kans dat een patiënt alcoholist is, gezien het feit dat hij een leverziekte heeft, is 7%.

De stelling van Bayes zegt:
P (A | B) = (0,07 \* 0,1) / 0,05 = 0,14

Met andere woorden, als de patiënt alcoholist is, is de kans op een leverziekte 0,14 (14%).

Dit is een grote stijging ten opzichte van de 10% die wordt gesuggereerd door gegevens uit het verleden. Maar het is nog steeds onwaarschijnlijk dat een bepaalde patiënt een leveraandoening heeft.

**Voorbeeld 2**

Een andere manier om naar de stelling te kijken, is door te zeggen dat de ene gebeurtenis op de andere volgt. Hierboven zei ik "tests" en "gebeurtenissen", maar het is ook legitiem om het te beschouwen als de "eerste gebeurtenis" die leidt tot de "tweede gebeurtenis". Er is niet één juiste manier om dit te doen: gebruik de terminologie die voor u het meest logisch is.

In een bepaalde pijnkliniek krijgt 10% van de patiënten verdovende pijnstillers voorgeschreven. In totaal is vijf procent van de patiënten in de kliniek verslaafd aan verdovende middelen (waaronder pijnstillers en illegale middelen). Van alle mensen die pijnstillers hebben voorgeschreven, is 8% verslaafd. *Als een patiënt verslaafd is, hoe groot is dan de kans dat hij pijnstillers krijgt voorgeschreven?*

**Stap 1: Zoek uit wat uw evenement "A" is uit de vraag.**

Die informatie staat in het cursief gedrukte deel van deze specifieke vraag. De eerste gebeurtenis (A) wordt voorgeschreven pijnstillers. Dat wordt gegeven als 10%.

**Stap 2: Zoek uit wat uw evenement "B" is uit de vraag.**

Die informatie staat ook in het cursief gedrukte deel van deze specifieke vraag. Event B is verslaafd zijn. Dat wordt gegeven als 5%.

**Stap 3: Zoek uit wat de kans is op gebeurtenis B (stap 2) gegeven gebeurtenis A (stap 1).**

Met andere woorden: zoek wat (B | A) is. We willen weten: "Gezien het feit dat mensen pijnstillers krijgen voorgeschreven, wat is de kans dat ze verslaafd zijn?" Dat wordt in de vraag gegeven als 8%, of .8.

**Stap 4: Voeg uw antwoorden uit stap 1, 2 en 3 in de formule in en los het op.**
P (A | B) = P (B | A) \* P (A) / P (B) = (0,08 \* 0,1) / 0,05 = 0,16

De kans dat een verslaafde pijnstillers krijgt voorgeschreven is 0,16 (16%).

**Voorbeeld 3: de medische test**

Een iets ingewikkelder voorbeeld betreft een medische test (in dit geval een genetische test):

Er zijn verschillende vormen van de stelling van Bayes die er zijn, en ze zijn allemaal gelijkwaardig (ze zijn gewoon op een iets andere manier geschreven). In deze volgende vergelijking wordt "X" gebruikt in plaats van "B." Bovendien zie je enkele veranderingen in de noemer. Het bewijs waarom we de vergelijking op deze manier kunnen herschikken, valt buiten het bestek van dit artikel (anders zouden het 5.000 woorden zijn in plaats van 2.000!). Als u echter een vraag tegenkomt met medische tests, gebruikt u waarschijnlijk deze alternatieve formule om het antwoord te vinden:



Bekijk de video voor een snelle oplossing of lees hieronder twee opgeloste Bayes 'Theorem-voorbeelden:

https://youtu.be/Jht31ML2HxI

1% van de mensen heeft een bepaald [genetisch defect](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.genome.gov/10001204) .
90% van de tests voor het gen detecteert het defect (echte positieven).
9,6% van de tests zijn [fout-positieven](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/false-positive-definition-and-examples/) .

Als een persoon een positief testresultaat krijgt, **wat is dan de kans dat hij daadwerkelijk het genetische defect heeft** ?

De eerste stap om de theorema-problemen van Bayes op te lossen, is om letters toe te wijzen aan gebeurtenissen:

* A = kans op het defecte gen. Dat werd in de vraag gegeven als 1%. Dat betekent ook de kans *niet* met het gen (~ A) is 99%.
* X = een positief testresultaat.

Dus:

1. P (A | X) = waarschijnlijkheid dat het gen een positief testresultaat krijgt.
2. P (X | A) = Kans op een positief testresultaat, aangezien de persoon daadwerkelijk het gen heeft. Dat werd in de vraag vermeld als 90%.
3. P (X | ~ A) = Kans op een positieve test als de persoon het gen *niet* heeft. Dat werd in de vraag gegeven als 9,6%

Nu hebben we alle informatie die we nodig hebben om in de vergelijking te stoppen:

P (A | X) = (.9 \* .01) / (.9 \* .01 + .096 \* .99) = 0.0865 (8,65%).

De kans dat het defecte gen in de test voorkomt, is 8,65%.

**Voorbeeld 4: een test op kanker**

Ik schreef over hoe uitdagende artsen [waarschijnlijkheid en statistieken vinden](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/probability-and-statistics/) in mijn post over het [verkeerd lezen van mammogramresultaten.](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/blog/%23physicians#physicians) Het is niet verwonderlijk dat artsen ver weg zijn met hun interpretatie van resultaten, gezien het feit dat er een aantal lastige kansen spelen. Hier is een tweede voorbeeld van hoe de stelling van Bayes werkt. Ik heb vergelijkbare getallen gebruikt, maar de vraag is anders geformuleerd om je nog een kans te geven om je te concentreren op hoe je beslist welke gebeurtenis A en welke gebeurtenis X is.

**V. Gegeven de volgende statistieken, wat is de kans dat een vrouw kanker heeft als ze een positief mammogramresultaat heeft?**

* Een procent van de vrouwen boven de 50 heeft borstkanker.
* Negentig procent van de vrouwen met borstkanker test positief op mammogrammen.
* Acht procent van de vrouwen krijgt valse positieven.

**Stap 1** : Wijs gebeurtenissen toe aan A of X. U wilt weten wat de kans van een vrouw op kanker is, gegeven een positief mammogram. Voor dit probleem is het hebben van kanker A en een positief testresultaat is X.

**Stap 2** : Maak een lijst van de delen van de vergelijking (dit maakt het gemakkelijker om met de eigenlijke vergelijking te werken):
P (A) = 0,01
P (~ A) = 0,99
P (X | A) = 0,9
P (X | ~ A) = 0,08

**Stap 3** : Voeg de onderdelen in de vergelijking in en los ze op. Merk op dat aangezien dit een medische test is, we de vorm van de vergelijking uit voorbeeld 2 gebruiken:
(0,9 \* 0,01) / ((0,9 \* 0,01) + (0,08 \* 0,99) = 0,10.

De kans dat een vrouw kanker krijgt, bij een positief testresultaat, is 10%.

**Weet je nog dat ik (daarboven ^^) zei dat er veel gelijkwaardige manieren zijn om de Bayes-stelling te schrijven?**

Hier is nog een vergelijking die u kunt gebruiken om het bovenstaande probleem te achterhalen. U krijgt exact hetzelfde resultaat:



Het belangrijkste verschil met deze vorm van de vergelijking is dat het de waarschijnlijkheidstermen [*intersectie*](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/probability-and-statistics/probability-main-index/probability-of-a-and-b/) (∩) en [*complement*](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/complementary-events/) ( c ) gebruikt. Beschouw het als een afkorting: het is dezelfde vergelijking, op een andere manier geschreven.

Gebruik de [vermenigvuldigingsregel](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/multiplication-rule-probability/) om de kansen aan de rechterkant van deze vergelijking te vinden :

P (B ∩ EEN) = P (B) \* P (A | B)

De twee kanten van de vergelijking zijn equivalent, en P (B) \* P (A | B) is wat we gebruikten toen we de teller in het bovenstaande probleem oplosten.

P (B) \* P (A | B) = 0,01 \* 0,9 = 0,009

Als noemer hebben we P (B c ∩ A) als onderdeel van de vergelijking. Dit kan (equivalent) worden herschreven als P (B c \* P (A | B c ).

Dit geeft ons:
P (B c \* P (A | B c ) = 0,99 \* 0,08 = 0,0792.

Als we die twee oplossingen in de formule invoegen, krijgen we:
0,009 / (0,009 + 0,0792) = 10%.

**De stellingproblemen van Bayes:**

**een andere manier om ernaar te kijken**.

De theorema-problemen van Bayes kunnen worden achterhaald *zonder* de vergelijking te gebruiken (hoewel het gebruik van de vergelijking waarschijnlijk eenvoudiger is). Maar als je niet kunt begrijpen waarom de vergelijking werkt (of wat hij doet), dan is hier de niet-vergelijkingsoplossing voor hetzelfde probleem in # 1 (het genetische testprobleem) hierboven.

**Stap 1** : Bepaal de waarschijnlijkheid van een echt positief op de test. Dat is gelijk aan mensen die daadwerkelijk het defect hebben (1%) \* echt positieve resultaten (90%) = .009.

**Stap 2** : Zoek de kans op een vals positief resultaat op de test. Dat is gelijk aan mensen die het defect niet hebben (99%) \* fout-positieve resultaten (9,6%) = 0,09504.

**Stap 3** : Bepaal de kans op een positief resultaat op de test. Dat is gelijk aan de kans op een echt positief (stap 1) plus een vals positief (stap 2) = .009 + .09504 = .0.10404.

**Stap 4** : Zoek de kans om het gen daadwerkelijk te hebben, bij een positief resultaat. Deel de kans op een echt, positief resultaat (stap 1) door de kans op een positief resultaat (stap 3) = .009 / .10404 = 0,0865 (8,65%).

**Andere vormen van de stelling van Bayes**

De stelling van Bayes kent verschillende vormen. U zult waarschijnlijk geen van deze andere vormen tegenkomen in een elementaire statistiekenklasse. De verschillende vormen kunnen voor verschillende doeleinden worden gebruikt.

Een versie gebruikt bijvoorbeeld wat Rudolf Carnap de " **waarschijnlijkheidsratio** " noemde . De waarschijnlijkheidsratio-regel stelt dat elke gebeurtenis (zoals een patiënt met een leverziekte) moet worden vermenigvuldigd met deze factor PR (H, E) = P E (H) / P (H).

Dat geeft de waarschijnlijkheid van de gebeurtenis afhankelijk van E. De **oddsratio-regel** lijkt sterk op de waarschijnlijkheidsratio, maar de [likelihood-ratio](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://www.statisticshowto.com/likelihood-ratio/) deelt het werkelijke positieve percentage van een test gedeeld door het percentage vals-positieve waarden.

De formele definitie van de Odds Ratio-regel is OR (H, E) = P H, (E) / P ~ H (E).

**Bayesiaanse spamfiltering**

Hoewel de stelling van Bayes op grote schaal wordt gebruikt in de medische wetenschappen, zijn er andere toepassingen.

Het wordt bijvoorbeeld gebruikt om [spam](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://en.wikipedia.org/wiki/Bayesian_spam_filtering) te [filteren](https://translate.google.com/translate?hl=nl&prev=_t&sl=en&tl=nl&u=https://en.wikipedia.org/wiki/Bayesian_spam_filtering) . Het geval in dit geval is dat het bericht spam is. De test op spam is dat het bericht enkele gemarkeerde woorden bevat (zoals "viagra" of "je hebt gewonnen"). Hier is de vergelijking die is opgesteld (van Wikipedia), gelezen als "De kans dat een bericht spam is, aangezien het bepaalde gemarkeerde woorden bevat":



De feitelijke vergelijkingen die worden gebruikt voor het filteren van spam zijn iets ingewikkelder; ze bevatten meer vlaggen dan alleen inhoud. De timing van het bericht, of hoe vaak het filter eerder dezelfde inhoud heeft gezien, zijn bijvoorbeeld twee andere spamtests.

Bron: https://www.statisticshowto.com/probability-and-statistics/probability-main-index/bayes-theorem-problems/