

Referate

Oor Bayes-beraming en -hipotesetoetsing*

D.J. de Waal

Departement Wiskundige Statistiek, Universiteit van die Oranje-Vrystaat, Posbus 339, Bloemfontein 9300

UITTREKSEL

Alhoewel Bayes se stelling reeds in 1764 gepubliseer is, is dit maar eers in die jongste tyd dat statistiese analises volgens die Bayes-benadering in die praktyk begin aandag kry. Daar het heelwat ontwikkelinge op dié gebied plaasgevind, veral as ons dink aan die beslissingsteorie en groepsbesluitneming. Twee aspekte, naamlik beraming en hipotesetoetsing – die gebied van statistiese inferensie waar die Wiskundige Statistiek hoofsaaklik figureer, kry in hierdie referaat aandag.

ABSTRACT

Summary on Bayes estimation and hypothesis testing

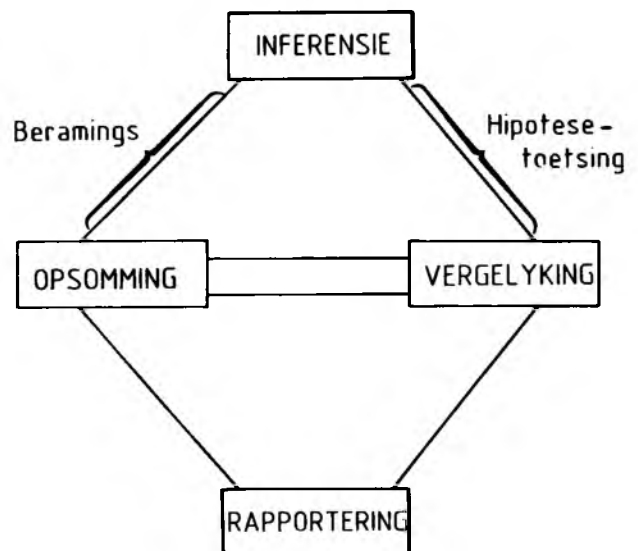
Although Bayes' theorem was published in 1764, it is only recently that Bayesian procedures were used in practice in statistical analyses. Many developments have taken place and are still taking place in the areas of decision theory and group decision making. Two aspects, namely that of estimation and tests of hypotheses, will be looked into. This is the area of statistical inference mainly concerned with Mathematical Statistics.

1. INLEIDING

Efron (1986) vra die vraag: "Waarom is almal nie Bayesiane nie?" Dit is na aanleiding van Lindley (1975) se stelling dat die Statistiek van die een-en-twintigste eeu Bayes-georiënteerd sal wees terwyl die twintigste-eeuse Statistiek hoofsaaklik nie-Bayes is. Morris (1986) sluit sy kommentaar op Efron se vraag af deur te sê dat praktiese statistici in aanraking kom met 'n verskeidenheid probleme waar frekwentistiese, objektiewe Bayes-, subjektiewe Bayes- en empiriese Bayes-metodes oplossings bied. Daar kan nie 'n finale oorwinning van die Bayes-metode oor die klassieke uitgespreek word nie en dit is ons plig om statistici op te lei wat kompromieë kan tref tussen die verskillende benaderings van analises.

Dit is miskien nodig om hier eers 'n duidelike omlýning te gee van wat Statistiek behels en waar die Wiskundige Statistiek hoofsaaklik figureer, en dan net kortliks die drie hoofgroepe van statistici, te wete die frekwentiste, die Fisheriane en die Bayesiane, te omskrywe.

Efron (1982) stel die vier basiese bewerkings van die Statistiek skematies soos volg voor:



FIGUUR 1: Die vier basiese statistiese bewerkings en hoedat hulle met mekaar skakel.

Rapportering is die mees basiese statistiese proses. Dit behels die versameling en tabellering van data.

Opsomming behels die identifisering van die karakteriserende elemente van die data. Interessante waarnemings word uitgewys. Statistieke soos die gemiddelde en variansie word bereken. In hierdie proses het ook heelwat wiskundige ontwikkelinge die afgelope tyd plaasgevind. Dink byvoorbeeld aan die bepaling van 'n waarskynlikheidsmodel waarmee die data opgesom kan word. Hier het Fisher 'n groot rol gespeel en hy het

* Voordrag gelewer tydens 'n simposium van die Afdeling Wis- en Natuurkunde in die Akademiegebou oor die onderwerp: *Statistiek as Ontledingstechniek*.

** Finansiële ondersteun deur die WNNR en UOVs.

veral baie aandag aan die eerste stappe van data-analise gegee. Metodes soos maksimum aanneemlikheid is veral deur Fisher uitgewys as 'n besonder kragtige tegniek vir die opsomming van data. Ander tegnieke soos die "Uitsnit-" en "Skoenlusmetode" het later in hierdie gebied bygekom.

Vergelyking is die ander belangrike element van data-analise wat teenoor opsomming staan. Dit is waar die data ontrafel word om interessante verskille uit te wys, soos byvoorbeeld dat studente beter presteer onder 'n nuwe metode van onderrig as onder die ou metode. Daar vind egter heelwat wisselwerking tussen *opsomming* en *vergeliking* plaas.

In die Wiskundige Statistiek word egter merendeels op die boonste proses naamlik dié van *Inferensie* gefokus. Dit is dan ook hier waar die Bayes-analises meer van toepassing is en wat in die laaste tyd sterk ontwikkel is. Inferensie het te make met die ekstrapolering van die data na wat ons verwag daar in die populasie plaasvind. Om die goudprys vir volgende jaar te voorspel is byvoorbeeld inferensie wat te make het met beraming. Om, daarenteen, die bewering te ondersoek dat byvoorbeeld koringvariëteit A 'n hoër opbrengs as variëteit B lewer, is inferensie ten opsigte van die toets van 'n hipotese. 'n Tipiese analise van data is 'n wisseling tussen *opsomming* en *vergeliking* deur die data te ontleed om ooreenkomste en verskille aan te toon. Inferensie word as 'n toets in die proses gebruik om foute wat as gevolg van besprekinge in die data ontstaan, tot 'n minimum te reduceer. So mag daar vier beramers van die goudprys vir die volgende jaar uit die data gevind word, maar daar moet gepoog word om dié beramer te gee wat die werklike goudprys so goed as moontlik beraam.

2. DIE FREKWENTIS/FISHERIAAN/BAYESIAAN

Veronderstel ons het 'n stogastiese veranderlike X met waarskynlikheidsdigtheidsfunksie $f(x|\theta)$ met θ 'n onbekende parameter. Indien ons belangstel in die beraming van θ uit 'n ewekansige steekproef van n waarnemings $x_1, \dots, x_n$ sal die frekwentis 'n funksie van die waarnemings ('n statistiek), sê $\hat{\theta}(x_1, \dots, x_n)$, of kortweg net as $\hat{\theta}$ geskrywe, soek, sodanig dat $(\hat{\theta} - \theta)^2$ byvoorbeeld 'n minimum is.

Heelwat ander kriteria soos onsydigheid, minimum-variëansie-onsydigheid en andere is ook kriteria wat opgestel kan word en waaraan die beramer moet voldoen. Vergemiddelding is die kos van die frekwentis.

Die Fisheriaan wat in navolging van Fisher se idees kyk na die aanneemlikheidsfunksie (AV)

$$\ell(\theta|x_1, \dots, x_n) = \prod_{i=1}^n f(x_i|\theta) \dots\dots\dots(2.1)$$

soek weer dié θ wat $\ell(\theta|x_1, \dots, x_n)$ 'n maksimum maak. Let op dat AV as 'n funksie van die parameter, gegee die data, beskou word. Fisher se ontwikkelinge het in die 1920's gekom en hy het gesê dat al die informasie uit die data in die AV vervat word.

Die Bayesiaan gebruik ook die AV as die funksie wat die informasie vanaf die data dra, maar voeg

daarby sekere a priori-informasie wat omtrent die parameter θ bestaan. Daaruit, met behulp van Bayes se stelling, word die posterior-waarskynlikheidsdigtheidsfunksie (w.d.f.) van θ , gegee die data, bepaal as

$$\pi(\theta|x_1, \dots, x_n) \propto \pi(\theta) \ell(\theta|x_1, \dots, x_n) \dots\dots\dots(2.2)$$

waar $\pi(\theta)$ die prior w.d.f. van θ is en $\propto$ die proporsionele bewerking aandui. Deur dan dié θ te bepaal wat $\pi(\theta|x_1, \dots, x_n)$ 'n maksimum maak, word 'n Bayes-beramer van θ verkry. Heelwat ander tipe Bayes-beramers i.p.v. die modus kan vanaf $\pi(\theta|x_1, \dots, x_n)$ verkry word. Daarop sal ons nie nou ingaan nie. Lindley (1978) sê 'n Bayesiaan is iemand wat onsekerhede met waarskynlikhede uitdruk. Wat hy daarmee bedoel, is byvoorbeeld dat die gedrag van die onbekende parameter aanvanklik uitgedruk word met $\pi(\theta)$ en dan later met $\pi(\theta|x_1, \dots, x_n)$. Posterior-waarskynlikhede is weer die kos van die Bayesiaan.

3. BAYES-BERAMING

Op die gebied van beraming is daar 'n magdom van literatuur veral sedert 1960 beskikbaar. Met die ontwikkeling van die bekende Stein-tipe beramers (James and Stein (1960)) is heelwat soortgelyke Bayes-beramers ontwikkel. Daar het ook heelwat ontwikkelinge op die gebied van empiriese Bayes-beraming plaasgevind (Efron en Morris (1971, 1972, 1973)) en dan verskyn daar in die boek van Berger (1986) ontwikkelinge op die gebied van robuuste prior-verdelings om beraming te doen. Ons sal aan die hand van 'n voorbeeld 'n paar van hierdie ontwikkelinge illustreer.

Voorbeeld 1 : Beraming van die IK van 'n persoon

Efron (1978) gee die volgende voorbeeld vir die beraming van die ware IK, sê θ , van 'n persoon: 'n Persoon is n keer getoets deur 'n navorser wat 'n nuwe toets ontwikkel het en sy IK (X) is elke keer bereken. Die gemiddelde $\bar{x}$ van die n waarnemings is bereken as $\bar{x} = 160$. Dit word aanvaar dat die stogastiese veranderlike X normaalverdeel is met gemiddelde θ en standaardfout $\sigma / \sqrt{n} = 7,5$. Die onsydige beraming asook die maksimum-aanneemlike beraming van θ is $\hat{\theta} = 160$. 'n Intervalberaming vir θ word ook dikwels gedoen. So, byvoorbeeld, sal 'n 95% vertrouensinterval van θ wees $\bar{x} \pm 1,96 \sigma / \sqrt{n} = (145,3; 174,7)$. Beskou ons egter die Bayes-beramer, dan word die vraag gevra: "Wat weet ons omtrent θ ?" Aangesien die betrokke persoon se ware IK as 'n enkele waarde van die ware IK's van die populasie waaruit die persoon afkomstig is, beskou kan word (en as dit bekend is dat die ware IK's van die populasie normaalverdeel is met gemiddelde $\mu = 100$ en standaardafwyking $\tau = 15$), het ons hier waardevolle informasie omtrent die gedrag van θ . Dit is dus bekend dat die AV gegee word deur

$$\ell(\theta|\bar{x}) = \phi(\theta|160; 7,5^2)$$

en die prior w.d.f. van θ gegee word deur

$$\pi(\theta) = \phi(\theta|100; 15^2).$$

$\phi(\bullet|a;b)$ dui die normaal w.d.f. aan met gemiddelde a en variansie b . Die posterior w.d.f. van θ , gegewe $\bar{x} = 160$, word dus gegee deur

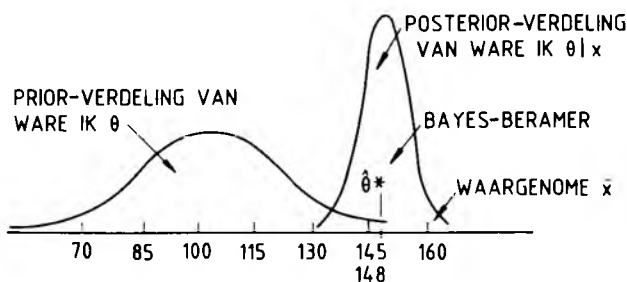
$$\pi(\theta|\bar{x}) = \phi(\theta|\mu', \tau')$$

waar

$$\text{met } \mu' = \mu + n\tau(\bar{x} - \mu)/\sigma^2 = 148$$

$$\text{en } \tau' = (1/\tau + n/\sigma^2)^{-1} = 6,7^2.$$

Deur die modus (of gemiddelde) van die posterior-verdeling van θ as die Bayes-beramer te gebruik, het ons die Bayes-beraming van θ gegee word deur $\hat{\theta}^* = 148$. Die $\hat{\theta}^* = \mu'$ is in die vorm van die bekende Stein-beramer. Figuur 2 illustreer die verdelings.



FIGUUR 2: Prior- en posterior-verdelings van ware IK's (Efron 1978).

Die beraming $\hat{\theta} = 148$ verskil heelwat van die vorige van 160 as gevolg van die ekstra informasie op θ wat nie daar in berekening gebring is nie. Die data word in die geval van die Bayes-beramer nie onvoorwaardelik aanvaar nie. Dit kan wees dat die IK-toets miskien bevraagteken moet word.

In bostaande Bayes-beraming het ons van 'n *egte prior* op θ , nl. $\pi(\theta|100;15^2)$, gebruik gemaak. Indien dit nie bekend was nie, kan nog steeds 'n Bayes-beramer gevind word deur van 'n *vae of onegte prior*, soos byvoorbeeld $\pi(\theta)\alpha 1, -\infty < \theta < \infty$, gebruik te maak. Dit gee 'n posterior-w.d.f. vir θ van

$$\pi(\theta|\bar{x}) = \phi(\theta|160;7,5^2).$$

Die Bayes-beraming is dan $\hat{\theta} = 160$. Berger (1986) wat baie aandag aan die robuustheid van prior-verdelings gegee het, sê dat in 'n sekere sin vae priors nie robuust is nie. Die gebruik van vae priors is egter 'n populêre tipe prior en daar word dikwels daarvan gebruik gemaak. 'n Meer robuuste tipe prior volgens Berger (1986) is egter die *hiërgargiese prior*-verdeling. In hierdie geval word 'n verdere waarskynlikheidsverdeling op die parameter(s) van $\pi(\theta|\mu, \tau)$ geplaas. So byvoorbeeld kan vae priors op μ en τ geplaas word wat weer later in die posterior-verdeling uitgeïntegreer word.

Die parameters van die prior-verdeling, indien onbekend, kan in sekere gevalle – veral as meer as een steekproef geneem is – uit die data beraam word.

Dit is die *empiriese Bayes-metode* en word ook redelik algemeen toegepas (kyk byvoorbeeld Efron en Morris (1972)). Die probleem hiermee is dat die data die prior-verdeling sterk beïnvloed, met die gevolg dat die AV gekonsentreer is in die sentrale gedeelte van die prior-verdeling. Bayes se waarde lê heel dikwels juis in die feit dat informasie buite die AV ingebring kan en wil word.

Berger (1986) het 'n kompromie probeer kry tussen die empiriese Bayes-benadering en die gebruik van 'n egter prior indien daar nie met volle oorgawe van die egte prior soos die $N(100;15^2)$ -verdeling in die IK voorbeeld, gebruik gemaak wil word nie. Hy noem dit die *gekontameneerde prior-verdeling* wat bestaan uit 'n kontaminasie van die egte prior, sê $\pi_0(\theta)$, en 'n verdeling $q(\theta)$ wat die AV se gebied insluit. Hy stel voor dat $q(\theta)$ 'n uniforme verdeling om die modus θ_0 van $\pi_0(\theta)$ is sodanig dat vir $\theta_0 - k < \theta < \theta_0 + k$, k sodanig bepaal word dat

$$m(x_1, \dots, x_n|k) = \frac{1}{2k} \int_{\theta_0 - k}^{\theta_0 + k} \pi(\theta|x_1, \dots, x_n) d\theta$$

'n maksimum is.

Voorbeeld 2 : Gekontameneerde prior

Bepaal ons k vir die data $\bar{x} = 160$ om die modus $\theta_0 = 100$, dan vind ons $k = 72,545$. Indien ons sê ons kontamineer $\pi_0(\theta) = \phi(\theta|100;15^2)$ met 'n 90% waarskynlikheid en $q(\theta) = \frac{1}{2(72,545)}$, $27,5 < \theta$

$< 172,5$, met 'n 1% waarskynlikheid, dan verkry ons die gekontameneerde prior

$$\pi(\theta) = 0,9\phi(\theta|100;15^2) + 0,1u(\theta|27,5;172,5) \dots (3.1)$$

Hier dui $u(\theta|\bullet)$ die uniforme w.d.f. aan. Die posterior-w.d.f. van θ , gegewe $\bar{x} = 160$, word nou (kyk Berger (1986) p. 233).

$$\pi(\theta|\bar{x} = 160) = 0,463\phi(\theta|148;6,7^2) + 0,537\phi(\theta|160;7,5^2)I_{(27,5;172,5)}(\theta) \dots (3.2)$$

$I_{(a,b)}(\theta)$ dui die indikatorfunksie aan sodat laasgenoemde normaalverdeling afgeknot is volgens $I(\theta)$.

Deur die verwagte waarde $E(\theta|\bar{x})$ uit (3.2) te beskou as die Bayes-beramer van θ , is gevind dat die Bayes-beraming nou

$$\hat{\theta}^* = 150$$

is.

Berger (1986) sê dat bogenoemde gekontameneerde klas priors 'n redelike klas is solank die essensiële gedrag van $\pi_0(\theta)$ asook die enkel-modaliteit θ_0 van $\pi_0(\theta)$ behoue bly. Dié tipe prior is robuust in verskeie opsigte.

4. BAYES-HIPOTESETOETSING

Op die gebied van hipotesetoetsing volgens 'n

Bayes-prosedure het heelwat ontwikkelinge reeds plaasgevind. Belangrike bydraes is dié van Lindley (1957), Berger (1986), Shafer (1982), Pratt (1965) en andere. Ons sal hier egter net konsentreer op die basiese stappe van die prosedure en dit illustreer aan die hand van 'n voorbeeld.

Die vraag wat ontstaan nadat so 'n Bayes-toets uitgevoer is, is hoe dit vergelyk met die klassieke toets, of hoe die bekende p-waarde vergelyk met die posterior-waarskynlikheid van die Bayes-toets. Ek is van mening dat die twee waarskynlikhede nie vergelykbaar is nie, maar tog is daar werke soos dié van Dickey (1977) en Casella en Berger (1985) wat vergelykings probeer tref. In die voorbeeld sal ons egter net die verskil tussen die waarskynlikhede uitwys wat tot verskillende gevolgtrekkings kan lei.

Laat, volgens De Groot (1970), Θ die parameter-ruimte van die parameter θ wees en gestel ons stel belang in die hipoteses

$$H_0 : \theta \in \Theta_0$$

$$H_1 : \theta \in \Theta_1$$

waar $\Theta = \{\Theta_0, \Theta_1\}$.

Die prior-w.d.f. van θ word gegee deur

$$\begin{aligned} \pi(\theta) &= \pi_0(\theta), \quad \theta \in \Theta_0 \\ &= \pi_1(\theta), \quad \theta \in \Theta_1 \quad \dots\dots\dots(4.1) \end{aligned}$$

'n Ewekansige steekproef van n waarnemings $x_1, \dots, x_n$ word van X gemaak. Stel die w.d.f. van X is $f(x|\theta)$. Die posterior w.d.f. van θ , gegee $x_1, \dots, x_n$, word dan gegee deur

$$\begin{aligned} \pi(\theta|x_1, \dots, x_n) &= \pi_0(\theta|x_1, \dots, x_n), \quad \theta \in \Theta_0 \\ \pi_1(\theta|x_1, \dots, x_n), \quad \theta \in \Theta_1 \quad \dots(4.2) \end{aligned}$$

waar

$$\pi_i(\theta|x_1, \dots, x_n) = \frac{\pi_i(\theta)\ell_i(\theta|x_1, \dots, x_n)}{\sum_{j=0}^1 \int_{\Theta_j} \pi_j(\theta)\ell_j(\theta|x_1, \dots, x_n)d\theta} \quad i = 0, 1$$

$\ell_j(\theta|x_1, \dots, x_n)$ is die AV onder $\theta \in \Theta_j$.

As *kriterium* vir die aanvaarding of verwerping van H_0 sê ons : Die data ondersteun H_0 indien

$$\pi_0(\theta|x_1, \dots, x_n) > \pi_0(\theta)$$

en Θ_0 'n enkele punt is. Indien Θ egter 'n gebied beslaan, ondersteun die data H_0 as die totale massa van $\pi_0(\theta|x_1, \dots, x_n)$ groter is as dié van $\pi_0(\theta)$.

Ons sal dit illustreer in die geval waar H_0 'n enkele punt is, m.a.w. 'n toets van 'n eenvoudige hipotese teen 'n saamgestelde alternatief.

Voorbeeld 3 : Bayes-toets op IK.

Beskou weer voorbeeld 1. Gestel ons stel belang in die toets

$$H_0 : \theta = 140$$

$$H_1 : \theta \neq 140.$$

Laat die prior-verdeling wees

$$\begin{aligned} \pi(\theta) &= 0,5, \quad \theta = 140 \\ &= 0,5\phi(\theta|100;15^2), \quad \theta \neq 140 \end{aligned}$$

Die prior-waarskynlikheid van $\pi_0(\theta) = 0,5$ is nie van kritiese belang as ons kyk na die kriterium wat uiteindelik gebruik gaan word nie. Die 0,5 impliseer dat ons begin deur H_0 'n 50% kans teenoor H_1 gee om waar te wees.

Die posterior-verdeling van θ , gegee $\bar{x} = 160$ en $\sigma/\sqrt{n} = 7,5$, is volgens (4.2) bereken as

$$\begin{aligned} \pi(\theta|\bar{x} = 160) &= 0,9747, \quad \theta = 140 \\ &= 0,0253\phi(\theta|148;6,7^2), \quad \theta \neq 140 \end{aligned}$$

Aangesien $\pi_0(\theta|\bar{x} = 160) > \pi_0(\theta)$ sê ons dat die data die hipotese H_0 ondersteun. Die posterior-waarskynlikheid 0,9747 vir H_0 om waar te wees, is heelwat groter as die prior-waarskynlikheid van 0,5. Sommige Bayesiane verkies om die Bayes-faktor

$$B = \frac{(1-\pi_0(\theta))\ell_0(\theta|x_1, \dots, x_n)}{\int_{\Theta_2} \pi_1(\theta)\ell_1(\theta|x_1, \dots, x_n)d\theta}$$

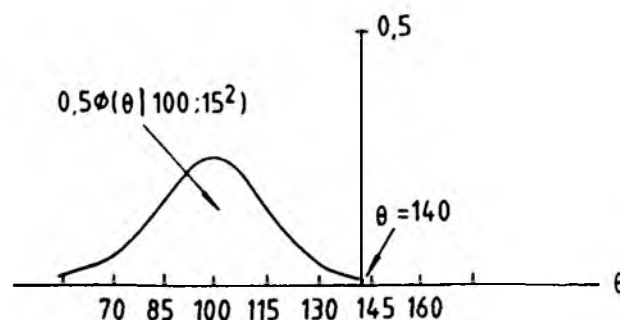
as kriterium te gebruik omdat die prior-waarskynlikheid $\pi_0(\theta) = 0,5$ in die geval van $\Theta_0 = \{140\}$ byvoorbeeld nie gespesifiseer hoef te word nie. $B > 1$ impliseer dat die data H_0 ondersteun.

Volgens die klassieke toets word die P-waarde bereken as

$$\begin{aligned} P &= P(\bar{X} > 160|\theta = 140) + P(\bar{X} < 120|\theta = 140) \\ &= 0,0078. \end{aligned}$$

Hiervolgens word H_0 verwerp.

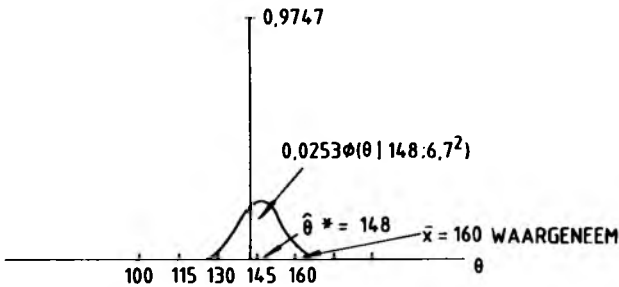
Die uitslag van die twee toetse is teenstrydig, om-



FIGUUR 3: Prior-verdeling van θ

dat die Bayes-toets die prior-informasie dat die ware IK's $N(100;15^2)$ verdeel is, in ag geneem het.

Figure 3 en 4 illustreer respektiewelik die prior- en posterior-verdelings van θ .



FIGUUR 4: Posterior-verdeling van θ gegee $\bar{x} = 160$

VERWYSINGS

- Berger, J.O. (1986). *Statistical Decision Theory*, 2 Ed. Springer-verlag.
- Casella, G. and Berger, J.O. (1985). Reconciling Bayesian and Frequentist evidence in the one-sided testing problem, *Mimeo Series No. 1657*, North Carolina State University.
- De Groot, M.H. (1978). *Optimal Statistical decisions* (McGray-Hill).
- Dickey, J.M. (1977). Is the tail area useful as an approximation, *J. American Statistical As.* 72, 138-142.
- Efron, B. (1978). Controversies in the foundation of Statistics, *American Mathematical Monthly*, 85, 231-246.
- Efron, B. (1982). Maximum likelihood and decision theory, *Annual of Statistics*, 10, 340-356.
- Efron, B. (1986). Why isn't everyone a Bayesian, *The American Statistician*, 40, 1-11.
- Efron B. and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimates – Part I : The Bayes case, *J. Amer. Statistical Assoc.*, 66, 807-815.
- Efron, B. and Morris, C. (1972). Limiting the risk of Bayes and empirical Bayes estimates – Part II : The empirical Bayes case, *J. Amer. Statist. Assoc.*, 67, 130-139.
- Efron, B. and Morris, C. (1973) : Stein's estimation rule and its competitors – an empirical Bayes approach, *J. Amer. Statist. Assoc.*, 68, 177.
- James, W. and Stein, C. (1960). Estimation with quadratic loss, *Proc. Fourth Berk. Symp.*, 1, 361-379.
- Lindley, D.V. (1975). The future of Statistics – a Bayesian 21st century, *Suppl. Adv. Appl. Prob.*, 7, 106-115.
- Lindley, D.V. (1978). Comment on parametric empirical Bayes inference, *J. Amer. Statist. Assoc.*, 78, 61-62.
- Morris, C. (1986). Comment on why everyone is not a Bayesian, *Amer. Statistician*, 40, 7-8.
- Pratt, (1965). Bayesian interpretation of standard statements, *R. Statist. Soc. B.*, 27, 169-203.
- Shafer, G. (1982). Lindley's paradox, *J. Amer. Statistical Assoc.*, 77, 325-351.