

**Adaptive significance levels
in linear regression models**

Alejandra Estefanía Patiño Hoyos

THESIS PRESENTED
TO THE
INSTITUTE OF MATHEMATICS AND STATISTICS
OF
UNIVERSITY OF SÃO PAULO
TO
OBTAIN THE TITLE
OF
DOCTOR IN SCIENCE

Program: Statistics

Advisor: Victor Fossaluza

Co-Advisor: Carlos Alberto de Bragança Pereira

During the development of this work the author received
financial support from CAPES/CNPq

São Paulo, February of 2020

Níveis de significância adaptativos em modelos de regressão linear

Alejandra Estefanía Patiño Hoyos

TESE APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
DOUTOR EM CIÊNCIAS

Programa: Estatística

Orientador: Victor Fossaluzza

Coorientador: Carlos Alberto de Bragança Pereira

Durante o desenvolvimento deste trabalho o autor recebeu auxílio
financeiro da CAPES/CNPq

São Paulo, Fevereiro de 2020

Níveis de significância adaptativos em modelos de regressão linear

Esta versão da tese contém as correções e alterações sugeridas pela Comissão Julgadora durante a defesa da versão original do trabalho, realizada em 05/12/2019. Uma cópia da versão original está disponível no Instituto de Matemática e Estatística da Universidade de São Paulo.

Comissão Julgadora:

- Prof. Dr. Victor Fossaluza (orientador) - IME-USP
- Prof. Dr. Luís Gustavo Esteves - IME-USP
- Prof. Dr. Adriano Polpo de Campos - UWA
- Prof. Dr. Rafael Bassi Stern - UFSCar
- Prof. Dr. Eduardo Yoshio Nakano - UnB

Agradecimentos

Expresso minha eterna gratidão aos meus professores Carlos Alberto de Bragança Pereira e Luís Gustavo Esteves por todos os seus valiosos ensinamentos e conselhos. Agradeço muito especialmente ao meu orientador e amigo Victor Fossaluzza por me guiar com sabedoria e paciência, pela sua ajuda incondicional que foi o suporte que me levou a concluir com sucesso o meu trabalho.

Agradecimentos infinitos aos meus pais, que foram o motor que dirigiu e manteve viva a esperança de alcançar este objetivo. Sem eles e sem a mão que meus grandes amigos estenderam para mim, não teria sido possível consegui-lo.

Gostaria de fazer um reconhecimento especial ao IME-USP, prestigioso instituto que me abriu suas portas e me forneceu todos os elementos necessários para que meu sonho de ser doutora em estatística se tornasse realidade. É minha responsabilidade continuar com o trabalho científico para honrar o legado deixado por esta grande instituição. Agradeço também às agências CAPES e CNPq pelo apoio financeiro.

Abstract

PATIÑO HOYOS, A.E. **Adaptive significance levels in linear regression models.** 2019. Doctoral thesis - Institute of Mathematics and Statistics, São Paulo, 2019.

The Full Bayesian Significance Test (FBST) for precise hypotheses is presented by [Pereira and Stern \(1999\)](#) as a Bayesian alternative to the traditional significance tests based on p -values. With the FBST the authors introduce the e -value as an evidence index in favor of the null hypothesis (\mathbf{H}). An important practical issue for the implementation of the FBST is to establish how small the evidence against \mathbf{H} must be in order to decide for its rejection. In this work we present a method to find a cutoff value for the evidence in the FBST by minimizing the linear combination of the averaged type-I and type-II error probabilities for a given sample size and also for a given dimensionality of the parameter space. Furthermore, we compare our methodology with the results obtained from the test proposed by [Pereira *et al.* \(2017\)](#) and [Gannon *et al.* \(2019\)](#) which presents the P -value as a decision-making evidence measure and includes an adaptive significance level. For that purpose, the scenario of linear regression models under the Bayesian approach is considered.

Keywords: Adaptive significance levels, Bayesian test, e -value, P -value, Linear regression, Predictive distribution, Bayes Factor, Significance test.

Resumo

PATIÑO HOYOS, A.E. **Níveis de significância adaptativos em modelos de regressão linear**. 2019. Tese de Doutorado - Instituto de Matemática e Estatística, São Paulo, 2019.

O Teste de Significância Totalmente Bayesiano (FBST) para hipóteses precisas foi apresentado por [Pereira and Stern \(1999\)](#) como uma alternativa Bayesiana aos testes de significância tradicionais baseados no p -value. Com o FBST, os autores introduzem o e -value como um índice de evidência em favor da hipótese nula (\mathbf{H}). Uma questão prática importante para a implementação do FBST é estabelecer quão pequena deve ser a evidência contra \mathbf{H} para decidir pela sua rejeição. Neste trabalho apresentamos um método para encontrar um valor de corte para a evidência no FBST minimizando a combinação linear das probabilidades ponderadas de erro tipo I e tipo II para um determinado tamanho de amostra e também para uma certa dimensão do espaço paramétrico. Além disso, comparamos nossa metodologia com os resultados obtidos usando o teste proposto por [Pereira et al. \(2017\)](#) e [Gannon et al. \(2019\)](#) que apresenta o P -value como medida de evidência e inclui um nível de significância adaptativo. Para isso, é considerado o cenário de modelos de regressão linear sob a abordagem Bayesiana.

Palavras-chave: Níveis de significância adaptativos, Teste Bayesiano, e -value, P -value, Regressão linear, Distribuição preditiva, Factor de Bayes, Teste de Significância.

Contents

Abbreviation	ix
List of Symbols	xi
List of Figures	xiii
List of Tables	xvii
1 Introduction	1
1.1 About this work	1
1.2 Organization of the thesis	3
2 Basic Concepts	5
2.1 Hypothesis testing	5
2.1.1 Frequentist hypothesis testing	6
2.1.2 Bayesian hypothesis testing	8
2.2 FBST	10
2.2.1 FBST cutoff value under Decision Theory approach	11
2.3 Significance index: P -value	12
2.4 Adaptive significance levels	14
2.5 Adaptive cutoff values for evidence in the FBST	17
3 Simple Illustrative Examples	21
3.1 Adaptive significance levels in proportion hypothesis testing	21
3.1.1 Evidence index: e -value	22
3.1.2 Significance index: P -value	32
3.1.3 Relationship between e -value and P -value	36
3.2 Adaptive significance levels in normal mean hypothesis testing	40
3.2.1 Evidence index: e -value	41
3.2.2 Significance index: P -value	48
3.2.3 Relationship between e -value and P -value	52

4	Adaptive Significance Levels in Linear Regression Models	59
4.1	Unknown-variance model	61
4.1.1	Conditional distributions	63
4.1.2	Marginal distributions	64
4.1.3	Prior predictive densities in regression-coefficient hypothesis testing	65
4.2	Known-variance model	68
4.2.1	Prior predictive densities in regression coefficients hypothesis testing	69
4.3	Adaptive significance levels in regression coefficients hypothesis testing	71
4.3.1	Evidence index: e -value	71
4.3.2	Significance index: P -value	76
4.4	Simulation study	77
4.5	Applications	94
4.5.1	Budget shares of British households dataset	94
4.5.2	Boston housing dataset	96
5	Conclusions	99
5.1	Final considerations	99
5.2	Suggestions for future research	101
A	Support Material	103
A.1	Marginal distribution of a normal sample mean	103
	Bibliography	105

Abbreviation

FBST	Full Bayesian Significance Test
CR	Critical Region
UMP	Uniformly Most Powerful Test
HPD	Highest Posterior Density
BF	Bayes Factor
<i>c.i.i.d</i>	conditionally independent and identically distributed

List of Symbols

\mathbb{R}	Set of real numbers
\mathbb{R}_+	Set of positive real numbers
\mathbb{R}^m	m -dimensional Euclidean space
\mathbf{H}	Null hypothesis
\mathbf{A}	Alternative hypothesis
$\dot{\cup}$	Disjoint union
$\sup_{\mathbf{H}}$	Supremum over the subspace defined by \mathbf{H}
\propto	Proportional to
$\mathbb{1}$	Indicator function
$\oint_{\mathbf{H}}$	Surface integral over the subspace defined by \mathbf{H}
E_{θ}	Expected value on θ
Γ	Gamma function
$\mathbf{1}_n$	n -dimensional vector of ones
$\mathbf{0}$	Vector of zeros
\mathbf{I}_n	Identity matrix of dimension $n \times n$
N_n	n -variate Normal distribution
IG	Inverse Gamma distribution
t_p	p -variate t distribution
χ_p^2	chi-square distribution with p degrees of freedom
$\arg \sup_{\mathbf{H}}$	Argument at which the sup over the subspace defined by \mathbf{H} occurs

List of Figures

3.1	Tangential set to $\mathbf{H} : \theta = 0.5$, with $\theta x^* \sim \text{Beta}(1, 3)$	24
3.2	Tangential set to $\mathbf{H} : \theta = 0.5$, with $\theta x^* \sim \text{Beta}(2, 2)$	24
3.3	Tangential set to $\mathbf{H} : \theta = 0.5$, with $\theta x^* \sim \text{Beta}(3, 1)$	25
3.4	(a) Averaged error probabilities (α_{φ_e} , β_{φ_e} and $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as functions of k , with $n = 2$ and $\theta \sim \text{Beta}(1, 1)$. (b) Power function for $k = 0.1$, $k = 0.5$ and $k = 0.8$, with $n = 2$ e $\theta \sim \text{Beta}(1, 1)$	27
3.5	Linear combination of averaged error probabilities ($\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k , with $n = 2$ and $\theta \sim \text{Beta}(1, 1)$	27
3.6	Prior distributions $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$	28
3.7	(a), (c) and (e) Averaged error probabilities (α_{φ_e} , β_{φ_e} e $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k ; (b), (d) and (f) Power function when $k = 0.1$, $k = 0.5$ and $k = 0.8$. Sample size $n = 40$	29
3.8	Linear combination of averaged error probabilities ($\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k . Sample size $n = 40$	30
3.9	Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim \text{Beta}(1, 1)$	30
3.10	Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim \text{Beta}(1, 3)$	31
3.11	Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim \text{Beta}(10, 10)$	31
3.12	Minimum k^* as function of n , for $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$	32
3.13	Optimal averaged error probabilities ($\alpha_{\varphi_e}^*$, $\beta_{\varphi_e}^*$ and $\alpha_{\varphi_e}^* + \beta_{\varphi_e}^*$) as function of n	33
3.14	Optimal averaged type-I error probability ($\alpha_{\varphi_P}^*$) as function of n , with $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$	35
3.15	Optimal averaged error probabilities ($\alpha_{\varphi_P}^*$, $\beta_{\varphi_P}^*$ and $\alpha_{\varphi_P}^* + \beta_{\varphi_P}^*$) as function of n , with $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$	35
3.16	Relationship between e -value and P -value as function of n in proportion hypothesis testing, with $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$	36
3.17	e -value and P -value densities under \mathbf{H} in proportion hypothesis testing, with $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$. Sample size $n = 60$	37
3.18	Posterior distributions with $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$, $\theta \sim \text{Beta}(10, 10)$, $\bar{x}_0^* = 0.6$ and $n = 10$	39

3.19	Tangential set to $\mathbf{H} : \theta = \theta_0$	42
3.20	Prior distributions $\theta \sim Normal(0, 0.1)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 100)$	45
3.21	(a), (c) and (e) Averaged error probabilities (α_{φ_e} , β_{φ_e} and $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k ; (b), (d) and (f) Power function when $k = 0.1$, $k = 0.5$ and $k = 0.8$. Sample size $n = 50$	46
3.22	(a), (c) and (e) Cutoff values k^* for $ev(0; \bar{x}_0)$ as function of n ; (b), (d) and (f) Optimal averaged error probabilities ($\alpha_{\varphi_e}^*$, $\beta_{\varphi_e}^*$ and $\alpha_{\varphi_e}^* + \beta_{\varphi_e}^*$) as function of n	47
3.23	Cutoff values k^* for $ev(0; \bar{x}_0)$ as function of n , with $\theta \sim Normal(0, 100)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 0.1)$	48
3.24	Optimal averaged type-I error probability (α_{φ_P}) as function of n , with $\theta \sim Normal(0, 100)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 0.1)$	51
3.25	Optimal averaged error probabilities ($\alpha_{\varphi_P}^*$, $\beta_{\varphi_P}^*$ and $\alpha_{\varphi_P}^* + \beta_{\varphi_P}^*$) as function of n , with $\theta \sim Normal(0, 0.1)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 100)$	52
3.26	Relationship between e -value and P -value as function of n in normal mean hypothesis testing, with $\theta \sim Normal(0, 100)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 0.1)$	53
3.27	e -value and P -value densities under \mathbf{H} in normal mean hypothesis testing, with $\theta \sim Normal(0, 100)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 0.1)$. Sample size $n = 100$	54
3.28	Posterior distributions $\theta \sim Normal(0, 0.1)$, $\theta \sim Normal(0, 1)$, $\theta \sim Normal(0, 100)$, $\bar{x}_0 = 0.3$ and $n = 10$	56
4.1	Unknown-variance model averaged error probabilities (α_{φ_e} , β_{φ_e} and $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k . Sample size $n = 100$	81
4.2	Known-variance model averaged error probabilities (α_{φ_e} , β_{φ_e} and $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k . Sample size $n = 100$	81
4.3	Unknown-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of n , with $d = 2$ and $d = 3$	82
4.4	Known-variance cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of n , with $d = 1$ and $d = 2$	82
4.5	Unknown-variance model optimal averaged type I error probability ($\alpha_{\varphi_P}^*$) as a function of n , with $d = 2$ and $d = 3$	83
4.6	Known-variance model optimal averaged type I error probability ($\alpha_{\varphi_P}^*$) as a function of n , with $d = 1$ and $d = 2$	83
4.7	Unknown-variance model optimal averaged error probabilities ($\alpha_{\varphi_e}^*$, $\beta_{\varphi_e}^*$ and $\alpha_{\varphi_e}^* + \beta_{\varphi_e}^*$) as functions of n	84
4.8	Known-variance model optimal averaged error probabilities ($\alpha_{\varphi_e}^*$, $\beta_{\varphi_e}^*$ and $\alpha_{\varphi_e}^* + \beta_{\varphi_e}^*$) as functions of n	84
4.9	Unknown-variance model optimal averaged error probabilities ($\alpha_{\varphi_P}^*$, $\beta_{\varphi_P}^*$ and $\alpha_{\varphi_P}^* + \beta_{\varphi_P}^*$) as functions of n	85
4.10	Known-variance model optimal averaged error probabilities ($\alpha_{\varphi_P}^*$, $\beta_{\varphi_P}^*$ and $\alpha_{\varphi_P}^* + \beta_{\varphi_P}^*$) as functions of n	85

4.11	Unknown-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of d , with $n = 60$ and $n = 120$	87
4.12	Known-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of d , with $n = 60$ and $n = 120$	87
4.13	Unknown-variance model optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as a function of d , with $n = 60$ and $n = 120$	88
4.14	Known-variance model optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as a function of d , with $n = 60$ and $n = 120$	88
4.15	Unknown-variance model relationship between e -value and P -value. Sample size $n = 60$	89
4.16	Known-variance model relationship between e -value and P -value. Sample size $n = 60$	90
4.17	Unknown-variance model e -value and P -value densities under \mathbf{H} . Sample size $n = 60$	91
4.18	Known-variance model e -value and P -value densities under \mathbf{H} . Sample size $n = 60$	92
4.19	Budget shares of British households dataset matrix plot.	95
4.20	Boston housing dataset matrix plot.	97

List of Tables

3.1	Cutoff values k^* for $ev(0.5; x_0^*)$, with $n = 2$ and $\theta \sim Beta(1, 1)$	27
3.2	Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim Beta(1, 1)$	30
3.3	Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim Beta(1, 3)$	31
3.4	Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim Beta(10, 10)$	31
3.5	Minimum k^* as function of n , for $\theta \sim Beta(1, 1)$, $\theta \sim Beta(1, 3)$ and $\theta \sim Beta(10, 10)$	32
3.6	Optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as function of n , with $\theta \sim Beta(1, 1)$, $\theta \sim Beta(1, 3)$ and $\theta \sim Beta(10, 10)$	35
3.7	Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim Beta(1, 1)$	38
3.8	Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim Beta(1, 3)$	38
3.9	Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim Beta(10, 10)$	39
3.10	Cutoff values k_{inf}^* , $ev(0.5; \bar{x}_0^*)$, $\alpha_{\varphi_P^*}$ and P -value (\bar{x}_0^*) as functions of n , with $\theta \sim$ $Beta(1, 1)$, $\theta \sim Beta(1, 3)$ and $\theta \sim Beta(10, 10)$	40
3.11	Cutoff values k_{sup}^* , $ev(0.5; \bar{x}_0^*)$, $\alpha_{\varphi_P^*}$ and P -value (\bar{x}_0^*) as functions of n , with $\theta \sim$ $Beta(1, 1)$, $\theta \sim Beta(1, 3)$ and $\theta \sim Beta(10, 10)$	40
3.12	Cutoff values k^* for $ev(0; \bar{x}_0)$ as function of n , with $\theta \sim Normal(0, 100)$, $\theta \sim$ $Normal(0, 1)$ and $\theta \sim Normal(0, 0.1)$	48
3.13	Optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as function of n , with $\theta \sim Normal(0, 100)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 0.1)$	51
3.14	Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim Normal(0, 100)$	55
3.15	Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim Normal(0, 1)$	55
3.16	Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim Normal(0, 0.1)$	56
3.17	Cutoff values k^* , $ev(0; \bar{x}_0)$, $\alpha_{\varphi_P^*}$ and P -value (\bar{x}_0) as functions of n , with $\theta \sim$ $Normal(0, 100)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 0.1)$	57

4.1	Unknown-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of n , with $d = 2$ and $d = 3$	82
4.2	Known-variance cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of n , with $d = 1$ and $d = 2$	82
4.3	Unknown-variance model optimal averaged type I error probability ($\alpha_{\varphi_P^*}$) as a function of n , with $d = 2$ and $d = 3$	83
4.4	Known-variance model optimal averaged type I error probability ($\alpha_{\varphi_P^*}$) as a function of n , with $d = 1$ and $d = 2$	83
4.5	Unknown-variance model cutoff values k^* , $ev(\mathbf{H}; \mathbf{y}_0)$ and P -value (\mathbf{y}_0) as function of n , with $d = 2$ and $d = 3$	86
4.6	Known-variance model cutoff values k^* , $ev(\mathbf{H}; \mathbf{y}_0)$ and P -value (\mathbf{y}_0) as function of n , with $d = 1$ and $d = 2$	86
4.7	Unknown-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of d , with $n = 60$ and $n = 120$	87
4.8	Known-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of d , with $n = 60$ and $n = 120$	87
4.9	Unknown-variance model optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as a function of d , with $n = 60$ and $n = 120$	88
4.10	Known-variance model optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as a function of d , with $n = 60$ and $n = 120$	88
4.11	Unknown-variance model cutoff values k^* , $ev(\mathbf{H}; \mathbf{y}_0)$ and P -value (\mathbf{y}_0) as functions of d , with $n = 60$ and $n = 120$	93
4.12	Known-variance model cutoff values k^* , $ev(\mathbf{H}; \mathbf{y}_0)$ and P -value (\mathbf{y}_0) as functions of d , with $n = 60$ and $n = 120$	93
4.13	Budget shares of British households dataset hypothesis-testing summary.	95
4.14	Budget shares of British households dataset optimal averaged error probabilities.	95
4.15	Boston housing dataset hypothesis-testing summary.	97
4.16	Boston housing dataset optimal averaged error probabilities.	98

Introduction

1.1 About this work

The Full Bayesian Significance Test (FBST) for precise hypotheses is presented by [Pereira and Stern \(1999\)](#) as a Bayesian alternative to the traditional significance tests based on p -values. With the FBST the authors introduce the e -value as an evidence index in favor of the null hypothesis (\mathbf{H}). An important practical issue for the implementation of the FBST is to establish how small the evidence must be to decide to reject \mathbf{H} . In that sense, [Madruga *et al.* \(2001\)](#) present loss functions such that the minimization of their posterior expected value gives “Bayesianity” to the FBST, having a characterization within the Decision Theory approach. This procedure provides a cutoff point for the evidence that depends on the severity of the error for deciding whether to reject or accept \mathbf{H} .

In the frequentist significance-test context, it is known that the p -value decreases as sample size increases, so by setting a single significance level, it usually leads to rejection of the null hypothesis. In the FBST procedure, the e -value exhibits similar behavior to the p -value when the sample size increases, which suggests that the cutoff point to define the rejection of \mathbf{H} should be a function of sample size. However, in the proposal of [Madruga *et al.* \(2001\)](#), no loss functions that explicitly take into account the sample size are studied.

In order to solve the problem of testing hypotheses in the usual way, in which changing the sample size influences the probability of rejecting or accepting the null hypothesis, [Oliveira \(2014\)](#) motivated by [Pereira \(1985\)](#), suggests that the level of significance in hypothesis testing should be a function of sample size. Instead of setting a single level of

significance, [Oliveira \(2014\)](#) proposes fixing the ratio of severity between type I and type II error probabilities based on the incurred losses in each case, and thus, given a sample size, defining the level of significance that minimizes the linear combination of the decision error probabilities. [Oliveira \(2014\)](#) shows that, by increasing the sample size, the probabilities of both kind of errors and their linear combination decrease, when in most cases, setting a single level of significance independent of sample size, only type-II error probability decreases. The tests proposed by [Oliveira \(2014\)](#) takes the same conceptual grounds of the usual tests for simple hypotheses based on the Neyman-Pearson Lemma as presented in [DeGroot \(1986\)](#). [Oliveira \(2014\)](#) extends the idea to composite and sharp hypotheses, according to the initial work of [Pereira \(1985\)](#).

Following the same line of work, [Pereira et al. \(2017\)](#) and [Gannon et al. \(2019\)](#) present a new hypothesis-testing procedure formulated from the ideas developed in previous works ([Irony and Pereira, 1995](#); [Montoya-Delgado et al., 2001](#); [Pereira, 1985](#); [Pereira and Wechsler, 1993](#)) and using a mixture of frequentist and Bayesian tools. This procedure introduces the P -value as a decision-making evidence measure and also includes an adaptive significance level, i.e., a significance level that is function of sample size. Such an adaptive significance level is obtained from the generalized form of the Neyman–Pearson Lemma where the linear combination of the type-I and type-II error probabilities is minimized. According to [Gannon et al. \(2019\)](#), the resulting hypothesis tests do not violate the Likelihood Principle and do not require any constraints on the dimensionalities of the sample space and parameter space.

Linear models are probably the most used statistical models to establish the influence of a set of covariates on a response variable. In that sense, the proper identification of the relevant variables in the model is an important issue in any scientific investigation, being a more challenging task in the context of Big-Data problems. In addition to high dimensionality, in recent statistical learning problems, it is common to find large datasets with thousands of observations. This fact may cause the hypothesis of nullity of the regression coefficients to be rejected, most of the time, due to the large sample size when the significance level is fixed.

The main goal of our work is to determine how small the Bayesian evidence in the FBST

should be in order to reject the null hypothesis. Therefore, taking into account the concepts in DeGroot (1986) and Pereira (1985) associated with optimal hypothesis tests, as well as the conclusions of Oliveira (2014) about the relationship between the significance levels and the sample size, and finally, considering the ideas developed recently by Pereira *et al.* (2017) and Gannon *et al.* (2019) related to adaptive significance levels, we present a method to find a cutoff point for the e -value by minimizing a linear combination of the averaged type-I and type-II error probabilities for a given sample size and also for a given dimensionality of the parameter space. For that purpose, the scenario of linear regression models under the Bayesian approach is considered. So, by providing an adaptive level for decision making and controlling the probabilities of both kind of errors, we intend to avoid the problems associated with the rejection of the hypotheses on the regression coefficients when the sample size is very large. In addition to the e -value, we calculate the P -value as well as its corresponding adaptive significance levels in order to compare the decision that can be made by performing the tests with each of these measures.

1.2 Organization of the thesis

In chapter 2, we present the basic definitions of hypothesis testing under the frequentist and Bayesian approaches, the concepts related to the adaptive significance levels as well as the description of the hypothesis-testing procedure given by Pereira *et al.* (2017) and Gannon *et al.* (2019). Lastly in this chapter, the proposed methodology to find a cutoff value for the evidence in the FBST is presented. Two simple examples illustrating the suggested methodology and comparisons with the results derived from the P -value are shown in chapter 3. In chapter 4 we extend both procedures to the context of linear regression models under the Bayesian point of view. For that purpose, we develop a simulation study with different scenarios and present two applications with real data. Finally, in chapter 5 the conclusions obtained in this work are discussed.

Basic Concepts

2.1 Hypothesis testing

A problem of statistical inference consists of making statements about unknown quantities that may be unobservable in a given context. An unknown quantity is usually called a parameter, denoted by θ . We suppose that θ lies in a certain known parameter space Θ . In order to make affirmations about θ , an experiment that allows the observation of a realization x of a set Ω ($x \in \Omega$) associated with a random vector $X = (X_1, \dots, X_n)$ is performed, where Ω is the set in which this random vector takes values, called the sample space, and n is the sample size. In general, the points of the sample space will be represented by x and a particular observation of the experiment by x_0 . Given knowledge of θ , it is assumed that the random vector X has known distribution $P_\theta \in \mathcal{P}$, where $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ is the set of distributions indexed by θ . For each $x_0 \in \Omega$, the function $L_{x_0} : \Theta \rightarrow \mathbb{R}_+$ that associates $L_{x_0}(\theta) = P(X = x_0 | \theta)$ (or $f(x_0 | \theta)$, in the case where X is absolutely continuous) with each $\theta \in \Theta$, is called the *likelihood function* of θ , resulting from the observation x_0 , and is the function that describes the relation between the parameter θ (unknown) and the experimental result x_0 (known) (Fossaluza, 2008).

A *statistical hypothesis* is a statement about a parameter θ . In the most usual scenario, we have two hypotheses, $\mathbf{H} : \theta \in \Theta_{\mathbf{H}}$, called the *null hypothesis*, and $\mathbf{A} : \theta \in \Theta_{\mathbf{A}}$, called the *alternative hypothesis*. Thus, the parameter space is divided into complementary subsets, $\Theta_{\mathbf{H}}$ and $\Theta_{\mathbf{A}} = \Theta_{\mathbf{H}}^c$, with $\Theta = \Theta_{\mathbf{H}} \dot{\cup} \Theta_{\mathbf{A}}$. A decision problem in which we must choose one of two (or more) hypotheses is called a problem of *hypothesis testing*. A procedure to decide

whether to accept the hypothesis \mathbf{H} or to accept the hypothesis \mathbf{A} is called a *test procedure*.

2.1.1 Frequentist hypothesis testing

Definition 1 A *hypothesis test* $\varphi : \Omega \rightarrow \{0, 1\}$ is a decision rule that specifies for which points $x \in \Omega$, the hypothesis $\mathbf{H} : \theta \in \Theta_{\mathbf{H}}$ must be rejected ($\varphi(x) = 1$) or accepted ($\varphi(x) = 0$). The subset of the sample space which contains the points leading to rejection of \mathbf{H} , that is, $\varphi^{-1}(1) = \{x \in \Omega : \varphi(x) = 1\}$, is called the **critical region (CR)** or **rejection region** of the test.

Once the hypothesis-test procedure φ is determined, its characteristics can be described by specifying, for each value of $\theta \in \Theta$, the probability $\pi_{\varphi}(\theta)$ that the procedure will lead to the rejection of \mathbf{H} or the probability $1 - \pi_{\varphi}(\theta)$ that it will lead to the acceptance of \mathbf{H} . The function $\pi_{\varphi}(\theta)$ is called the *power function* of the test φ and is defined as follows:

Definition 2 The *power function* of a hypothesis test φ is the function

$$\begin{aligned} \pi_{\varphi} : \Theta &\longrightarrow [0, 1] \\ \theta &\longmapsto \pi_{\varphi}(\theta) = P(\{x \in \Omega : \varphi(x) = 1\} | \theta) \\ &= P(\varphi^{-1}(\{1\}) | \theta) \\ &= P(X \in CR | \theta), \quad \theta \in \Theta. \end{aligned}$$

The ideal power function would be one for which $\pi_{\varphi}(\theta) = 0$ for every value $\theta \in \Theta_{\mathbf{H}}$ and $\pi_{\varphi}(\theta) = 1$ for every value of $\theta \in \Theta_{\mathbf{A}}$. In a practical problem, however, these rarely exists any test procedure having this ideal power function (DeGroot, 1986).

Whenever a hypothesis test is performed, there is a possibility of error in the decision, more precisely, two types of errors that can be committed must be considered

$$\text{Error of Type I} = \{\text{Rejecting } \mathbf{H} | \mathbf{H} \text{ is true}\} = \{\varphi(x) = 1 | \theta \in \Theta_{\mathbf{H}}\},$$

$$\text{Error of Type II} = \{\text{Accepting } \mathbf{H} | \mathbf{A} \text{ is true}\} = \{\varphi(x) = 0 | \theta \in \Theta_{\mathbf{A}}\}.$$

In terms of the power function, for each $\theta \in \Theta_{\mathbf{H}}$, $\pi_{\varphi}(\theta)$ is the probability of committing a type-I error, and for each $\theta \in \Theta_{\mathbf{A}}$, $1 - \pi_{\varphi}(\theta)$ would be the probability of committing

a type-II error. The error probabilities will be denoted as $\alpha_\varphi = P(\text{Error of Type I})$ and $\beta_\varphi = P(\text{Error of Type II})$.

Note that when $\Theta_{\mathbf{H}} = \{\theta_{\mathbf{H}}\}$ and $\Theta_{\mathbf{A}} = \{\theta_{\mathbf{A}}\}$ (simple hypotheses), $\alpha_\varphi = P(\{\text{Rejecting } \mathbf{H}\} | \Theta_{\mathbf{H}}) = \pi_\varphi(\theta_{\mathbf{H}})$ and $\beta_\varphi = P(\{\text{Accepting } \mathbf{H}\} | \Theta_{\mathbf{A}}) = 1 - \pi_\varphi(\theta_{\mathbf{A}})$ are the probabilities of committing a type-I error and a type-II error respectively.

In the case of composite hypotheses, an alternative definition of α_φ and β_φ is needed. In the frequentist approach, α_φ is considered as the size of a given test, that is,

$$\sup_{\mathbf{H}} \pi_\varphi(\theta) = \alpha_\varphi.$$

It is desirable to find a test procedure for which the probabilities α_φ and β_φ of the two types of error are small. However, for a given sample size, it is typically not possible to find a test procedure for which both α_φ and β_φ are arbitrarily small (DeGroot, 1986).

According to the approach proposed by Neyman and Pearson, one way to find a reasonable test is to fix the maximum probability of a type-I error that the decision maker is willing to commit and try to find a test with the lowest probability of a type-II error. In this sense, the tests of interest are those in which the probability of a type-I error is less than or equal to a fixed value α_φ , $0 \leq \alpha_\varphi \leq 1$, for all $\theta \in \Theta_{\mathbf{H}}$, denoted as the tests with *level of significance* α_φ . Once the level of significance is fixed, one can determine the critical region of the test, finding the critical value k such that

$$\alpha_\varphi = \sup_{\mathbf{H}} \pi_\varphi(\theta) = \sup_{\mathbf{H}} P(\{X \in \Omega : \varphi(X) = 1\} | \theta) = \sup_{\mathbf{H}} P(W(X) \geq k | \theta) = \alpha_\varphi(k),$$

where $W(X)$ is called *test statistic*. It is generally a function of a sufficient statistics and is used to sort the sample space such that large values of W indicate points more unfavorable to \mathbf{H} .

From the fact that a good test should have a power function close to 0 when $\theta \in \Theta_{\mathbf{H}}$ and close to 1 when $\theta \in \Theta_{\mathbf{A}}$, it is possible to seek in the class of tests of smaller or equal size α_φ the one that presents the greatest power function and, consequently, lower probability of a type-II error for each $\theta \in \Theta_{\mathbf{A}}$. These tests are called *uniformly most powerful tests* (UMP) and they are considered as optimal procedures from the frequentist point of view.

In the case where both hypotheses are simple, it can be obtained by a well-known result called the Neyman-Pearson Lemma (DeGroot, 1986). To find UMP tests is not always a trivial task, and therefore, a more general procedure known as a *generalized likelihood ratio test* was constructed providing reasonable results in most cases and having good asymptotic properties. For more details on these tests see Casella and Berger (2002).

Another way to report the result of a hypothesis test is to present the value of a statistic suggested by Fisher called a “*p*-value”.

Definition 3 A ***p*-value** $p(X)$ is a test statistic satisfying $0 \leq p(x) \leq 1$ for every sampling point $x \in \Omega$. Let $W(X)$ be a test statistic such that large values of W provide evidence against the hypothesis $\mathbf{H} : \theta \in \Theta_{\mathbf{H}}$. For each $x_0 \in \Omega$, we define

$$\mathbf{p}\text{-value} = p(x_0) = \sup_{\mathbf{H}} P(W(X) \geq W(x_0) | \theta).$$

We say that a ***p*-value** is ***valid*** if, for all $\theta \in \Theta_{\mathbf{H}}$ and all $\alpha_{\varphi} \in [0, 1]$,

$$P(p(X) \leq \alpha_{\varphi} | \theta) \leq \alpha_{\varphi}.$$

“Small” values of $p(X)$ provide evidence against the hypothesis \mathbf{H} . The *p*-value $p(x_0)$ can be interpreted as the lowest level of significance for which the test would lead to the rejection of \mathbf{H} based on observation x_0 (Schervish, 1995). An alternative definition of the *p*-value for an observation x_0 is described in Pereira and Wechsler (1993) as follows:

Definition 4 The *p*-value is the probability, under \mathbf{H} , of the event composed by all sample points that are at least as ***extreme*** as x_0 is.

The tests based on *p*-value are commonly called *significance tests* and propose the use of a measure of evidence in favor the null hypothesis (Schervish, 1995).

2.1.2 Bayesian hypothesis testing

In the Bayesian context, Bayes’ Theorem is used to join the information brought by observation x_0 , expressed in the likelihood function $L_{x_0}(\theta)$, with the *prior* distribution $g(\theta)$,

in order to update the known information about θ to generate the *posterior* distribution,

$$f(\theta|x_0) = \frac{g(\theta) L_{x_0}(\theta)}{\int_{\Theta} g(\theta) L_{x_0}(\theta) d\theta} \propto g(\theta) L_{x_0}(\theta), \quad (2.1)$$

for $\theta \in \Theta$ and $x_0 \in \Omega$.

The posterior distribution incorporates all the available information about θ , and therefore all Bayesian inference procedures will be based exclusively on $f(\theta|x)$. Under this approach, θ is considered a random variable because it is something unknown, and, within the subjectivist perspective, should receive probability (DeGroot, 1986).

From the Bayesian point of view, estimation and hypothesis-testing procedures can be constructed based on *Decision Theory*. In the context of hypothesis testing, it is basically tested through the posterior probability of the hypotheses, choosing under some criterion the one with the highest probability.

Among the alternatives for testing non-precise hypotheses we can find the *Bayes Test*. In this procedure, the expected loss (or risk) is minimized, and its advantage is the simplicity both in the formulation and in the interpretation of the results. Another proposal is the procedures based on the *Bayes Factor* (BF), which is a measure of the evidence favoring \mathbf{H} over \mathbf{A} , formulated as the ratio of the prior expectation of the likelihood under the two sets of hypotheses. In cases where $\Theta_{\mathbf{H}}$ has zero Lebesgue measure (precise hypothesis), the tests mentioned are not suitable (DeGroot, 1970). The search for a solution to this type of testing has been a controversial and longstanding problem in statistical inference (Pereira *et al.*, 2008). One of the proposed solutions is through the Jeffreys test, in which a positive probability is given over the set $\Theta_{\mathbf{H}}$ which is considered in the construction of the prior for θ . However, the use of a point mass prior on the null hypothesis, as in this case, can be considered as an *ad hoc* assumption (Pereira and Stern, 2001), and also generates the effect of Lindley's Paradox (Lindley, 1957). The *FBST* (Pereira and Stern, 1999) appears as a completely Bayesian alternative to test precise hypotheses, attempting to solve some of the inconsistencies arising from standard tests, both Frequentist and Bayesian. This is a test of interest in our work, so we will present it in detail in next section.

2.2 FBST

The *Full Bayesian Significance Test* (FBST) was proposed by [Pereira and Stern \(1999\)](#) for precise or “sharp” hypotheses (subsets of Θ with fewer dimensions than the dimension of the whole parameter space, and therefore, with null Lebesgue measure) based on the evidence in favor of \mathbf{H} , calculated as the complement of the posterior probability of the HPD (Highest Posterior Density) region, which is tangent to the set that defines the null hypothesis. The following is the formal definition of FBST ([Madruga and Pereira, 2005](#)):

Definition 5 Consider a standard parametric statistical model, i.e., for an integer m , $\theta \in \Theta \subset \mathbb{R}^m$ is the parameter, $g(\theta)$ a prior probability density over Θ , x_0 is the observed data, and $L_{x_0}(\theta)$ is the likelihood generated by x_0 . After data x_0 have been observed, the posterior probability density for θ given x_0 is denoted by $f(\theta|x_0) \propto L_{x_0}(\theta)g(\theta)$.

Consider a null hypothesis $\mathbf{H}: \theta \in \Theta_{\mathbf{H}}$ (sharp hypothesis, i.e., $\dim(\Theta_{\mathbf{H}}) < \dim(\Theta)$). The tangential set to \mathbf{H} is given by

$$T_{x_0} = \left\{ \theta \in \Theta : f(\theta|x_0) > \sup_{\mathbf{H}} f(\theta|x_0) \right\}.$$

The measure of evidence (*e-value*) in favor of \mathbf{H} is defined as the posterior probability of the complement of T_{x_0} , that is,

$$ev(\mathbf{H}; x_0) = 1 - P(\theta \in T_{x_0}|x_0) = 1 - \int_{T_{x_0}} f(\theta|x_0) d\theta.$$

The **FBST** is the procedure that rejects \mathbf{H} whenever $ev(\mathbf{H}; x_0)$ is small ([Pereira et al., 2008](#)).

The evidence index, *e-value*, in favor of a precise hypothesis, considers all points of the parameter space which are less “probable” than some point in $\Theta_{\mathbf{H}}$. A large value of $ev(\mathbf{H}; x_0)$ means that the subset $\Theta_{\mathbf{H}}$ lies in a high-probability region of Θ , and therefore, the data support the null hypothesis; on the other hand, a small value of $ev(\mathbf{H}; x_0)$ means that $\Theta_{\mathbf{H}}$ is in a low-probability region of Θ and the data would make us discredit the null hypothesis ([Madruga and Pereira, 2005](#)).

According to [Pereira and Stern \(1999\)](#), the FBST has the following properties:

- i) Gives an intuitive and simple measure of significance for the null hypothesis, a probability in the parameter space, besides being able to be characterized geometrically.
- ii) Can be easily implemented using numerical integration and optimization techniques.
- iii) Is fully Bayesian, only requiring knowledge of the parameter space represented in the posterior distribution.
- iv) Considers only the observed sample, allowing no ad hoc artifice, like a positive prior probability distribution on the precise hypothesis as in the common Bayesian tests using Bayes Factor.
- v) Considers the alternative hypothesis in equal standing with the null hypothesis.

2.2.1 FBST cutoff value under Decision Theory approach

A major practical issue for the implementation of the FBST is the determination of how small the Bayesian evidence in favor of \mathbf{H} must be in order to decide for its rejection. The formal identification of the FBST as a Bayes hypothesis test yields critical values derived from the loss functions allowing such identification. In this sense, [Madruga *et al.* \(2001\)](#) present loss functions such that minimization of their posterior expected values confers “Bayesianity” to the FBST, having in this way a characterization within the Decision Theory approach ([Pereira *et al.*, 2008](#)). The authors describe the problem as follows:

The FBST procedure is given by the criterion according to which

- \mathbf{H} is to be rejected if $ev(\mathbf{H}; x_0) \leq k$.
- \mathbf{H} is to be accepted if $ev(\mathbf{H}; x_0) > k$.

Formally, consider $D = \{Acceptance\ of\ \mathbf{H}(d_0),\ Rejection\ of\ H(d_1)\}$ the decision space of a statistical hypothesis test problem, and let $L : D \times \Theta \rightarrow \mathbb{R}_+$ be the loss function defined by

$$\begin{aligned} L(Rejection\ of\ \mathbf{H}, \theta) &= a[1 - \mathbb{1}(\theta \in T_{x_0})] \text{ and} \\ L(Acceptance\ of\ \mathbf{H}, \theta) &= b + c\mathbb{1}(\theta \in T_{x_0}), \end{aligned}$$

where a , b and d are positive real numbers. [Madruga et al. \(2001\)](#) prove that minimization of a posterior expected- L loss function is a FBST procedure. In addition, [Madruga et al. \(2001\)](#) prove that, for this loss function, the FBST cutoff value is $k = (b + c)/(a + c)$.

Note that L is a loss function which depends on the action (acceptance or rejection), on the parameter θ and on the tangential set T_{x_0} , therefore, it is a loss function dependent on the observed sample point x_0 and on the prior density for θ . In practice, it is not simple to make a choice for the a , b , and c values because that involves the researcher's opinion about how damaging an error in the decision would be, making them particular values for each problem. There are other loss functions whose minimization is equivalent to the FBST procedure, but they are only minor variations of the L function. However, loss functions that explicitly take into account the sample size have not yet been studied.

2.3 Significance index: P -value

A significance index is a real function over the sample space that is used as an evidence measure for decision-making with respect to accepting or rejecting the null hypothesis \mathbf{H} ([Pereira et al., 2017](#)). The most well-known significance index is the p -value, stated in a previous section. In order to define a p -value which regards the alternative hypothesis \mathbf{A} , contrasting the idea given by definitions like [Definition 4](#), which could be problematic in the sense of its calculation or even in its interpretation, [Pereira and Wechsler \(1993\)](#) present an alternative point of view for the p -value and define a quantity called “ P -value”, written with capital P with the aim of making a distinction between both quantities.

Definition 6 *The P -value is the probability, under \mathbf{H} , of the event composed by all sample points that favor \mathbf{A} (against \mathbf{H}) at least as much as x_0 does.*

This new index, the P -value, can be calculated for arbitrarily complex hypotheses that lead to complex rejection regions, like those where the null hypotheses will be rejected if the experimental observation were to occur there, as explained by [Gannon et al. \(2019\)](#), which is an advantage over the small- p p -value. However, to do so, it is necessary to establish an ordering of all the points in the sample space according to how much each possible observation favors one of the hypotheses over the other. In that way, [Pereira et al. \(2017\)](#) and

Gannon *et al.* (2019) formalize a Bayesian definition for the P -value adopting the approach from Montoya-Delgado *et al.* (2001), which suggests the use of the Bayes factor values of all sample points to induce the necessary order. Then the Bayes factor takes the place of the likelihood ratio and the average value of the likelihood function replaces its maximum value. The mean of the likelihood function under the null hypotheses will be the density used in the calculation of the new index.

Definition 7 *With the prior density $g(\theta)$ defined over the two pieces of the parameter space $\Theta_{\mathbf{H}}$ and $\Theta_{\mathbf{A}}$, let $f_{\mathbf{H}}(x)$ and $f_{\mathbf{A}}(x)$ be the Bayesian prior predictive densities under the respective hypotheses. Both are probability density functions over the sample space Ω , and they are calculated as the following conditional expectations:*

$$\begin{aligned} f_{\mathbf{H}}(x) &= f(x|\mathbf{H}) = E_{\theta} [f(x|\theta)|\mathbf{H}] \\ &= \int_{\Theta} f(x|\theta) dP_{\mathbf{H}}(\theta) \end{aligned} \tag{2.2}$$

and

$$\begin{aligned} f_{\mathbf{A}}(x) &= f(x|\mathbf{A}) = E_{\theta} [f(x|\theta)|\mathbf{A}] \\ &= \int_{\Theta} f(x|\theta) dP_{\mathbf{A}}(\theta), \end{aligned} \tag{2.3}$$

where $P_{\mathbf{H}}$ and $P_{\mathbf{A}}$ are the prior probability measure of θ restricted to the sets \mathbf{H} and \mathbf{A} respectively.

The ratio between the two functions is known as the Bayes factor,

$$BF(x) = \frac{f_{\mathbf{H}}(x)}{f_{\mathbf{A}}(x)}. \tag{2.4}$$

If an experiment yields a result x_0 , the P -value is calculated as the probability, given probability (density) function $f_{\mathbf{H}}(x)$, of a point in the sample space favoring \mathbf{A} as much as or more than x_0 does. That is, it is the sum or integral of the probability (density) function $f_{\mathbf{H}}(x)$ over the part of the sample space where the Bayes factor $BF(x)$ is less than or equal to the Bayes factor calculated at x_0 , $BF(x_0)$. Calling that part of the sample space Ψ_P , it is defined as

$$\Psi_P = \{x \in \Omega : BF(x) \leq BF(x_0)\} \quad (2.5)$$

and the P -value is

$$P\text{-value}(x_0) = P(\{X \in \Omega : BF(X) \leq BF(x_0)\} | \mathbf{H}) \quad (2.6)$$

i.e.,

$$P\text{-value}(x_0) = \begin{cases} \frac{\sum_{\Psi_P} f_{\mathbf{H}}(x)}{\Psi_P} & \text{if } \Omega \text{ is discrete} \\ \int_{\Psi_P} f_{\mathbf{H}}(x) dx & \text{if } \Omega \text{ is continuous.} \end{cases} \quad (2.7)$$

In the absolutely continuous case for which the parametric subspaces defined by the hypotheses have different dimensionalities, the definition of a prior density under the subset of smaller dimension, say \mathbf{H} , is obtained from the following expression

$$g_{\mathbf{H}}(\theta) = \begin{cases} 0 & \text{if } \theta \in \Theta_{\mathbf{A}} \\ \frac{g(\theta) \mathbb{1}(\theta \in \Theta_{\mathbf{H}})}{\oint_{\Theta_{\mathbf{H}}} g(t) dt} & \text{if } \theta \in \Theta_{\mathbf{H}}. \end{cases} \quad (2.8)$$

The denominator is the surface integral over the subspace $\Theta_{\mathbf{H}}$ (subject to the condition that can be defined). When $\Theta_{\mathbf{H}}$ consists of a single point, it is not necessary to calculate the integral (Pereira *et al.*, 2017).

2.4 Adaptive significance levels

In the search for optimal tests, DeGroot (1986) showed that in the simple hypotheses cases, it is possible to construct a procedure for which the value of a specific linear combination of α_{φ} and β_{φ} (weighted sum), say $a\alpha_{\varphi} + b\beta_{\varphi}$, with $a, b > 0$, is minimized. DeGroot (1986) proved that the test $\varphi^*(x)$ which minimizes a linear combination of the type I and

type II error probabilities, $a\alpha_\varphi + b\beta_\varphi$, is given by:

$$\varphi^*(x) = \begin{cases} 0, & \text{if } af_{\mathbf{H}}(x) \geq bf_{\mathbf{A}}(x) \\ 1, & \text{if } af_{\mathbf{H}}(x) < bf_{\mathbf{A}}(x) \end{cases} \Rightarrow \varphi^*(x) = \begin{cases} 0, & \text{if } \frac{f_{\mathbf{H}}(x)}{f_{\mathbf{A}}(x)} \geq \frac{b}{a} \\ 1, & \text{if } \frac{f_{\mathbf{H}}(x)}{f_{\mathbf{A}}(x)} < \frac{b}{a}. \end{cases}$$

Then, for any other test φ ,

$$a\alpha_\varphi + b\beta_\varphi \geq a\alpha_{\varphi^*} + b\beta_{\varphi^*}. \quad (2.9)$$

This is the **generalized Neyman-Pearson Lemma** in its Bayesian form. According to [Pereira \(1985\)](#), this result involves interesting aspects of the sample space, since both β_φ and α_φ values could vary with the sample size. So, under this concept, instead of α_φ being fixed and β_φ tending to decrease with increasing sample size, both error probabilities depend on the sample size. The a and b values represent the relative seriousness of errors of the two types or, equivalently, relative prior preferences for the competing hypotheses. For example if $b/a = 1$ it is said that β_φ and α_φ are equally severe, whereas if $b/a < 1$, then α_φ undergoes a more intense minimization than β_φ , which means that type-I error is considered more serious than type II error and also indicates a prior preference for \mathbf{H} . Note that minimizing $a\alpha_\varphi + b\beta_\varphi$ is equivalent to minimizing $\alpha_\varphi + w\beta_\varphi$ when $w = b/a$, which makes the Neyman-Pearson Lemma a Corollary of this result. Even so, both have a very different practical interpretation. [Pereira \(1985\)](#) proves that the $\varphi^*(x)$ test is also a Bayesian solution to the problem of testing simple versus simple hypotheses, and furthermore, proposes a method under the same approach to test hypotheses involving spaces of different dimensionalities that can receive a frequentist and Bayesian interpretation and has no restriction of applicability. [Oliveira \(2014\)](#) motivated by [Pereira \(1985\)](#), proposes a test where the ratio of severity between type-I and type-II error probabilities is fixed based on the incurred losses in each case, and thus, given a sample size, an ideal level of significance that minimizes the linear combination of the decision errors is defined. This test is extended to composite and sharp hypotheses. On the other hand, [Pericchi and Pereira \(2016\)](#) generalize the error-weighting approach by defining the expected type-I and type-II error probabilities, and find that the

optimal region is given by the ratio of evidence, i.e., the ratio of averaged likelihoods (in relation to a prior measure) and a fixed threshold.

Following the same line of work, [Pereira *et al.* \(2017\)](#) and [Gannon *et al.* \(2019\)](#) present a new hypothesis-testing procedure using a mixture of frequentist and Bayesian tools. This procedure introduces the P -value as a decision-making evidence measure and also includes an adaptive significance level which is function of the sample size. To describe the procedure, consider the test

$$\varphi_P^*(x) = \begin{cases} 0 & \text{if } \text{BF}(x) > \frac{b}{a} \\ 1 & \text{if } \text{BF}(x) \leq \frac{b}{a}. \end{cases} \quad (2.10)$$

$\varphi_P^*(x)$ minimizes the linear combination $a\alpha_\varphi + b\beta_\varphi$. Here, the Bayes factor is compared to the ratio of the coefficients b/a . The power function of φ_P^* is given by

$$\pi_{\varphi_P^*}(\theta) = P(\{X \in \Omega : \varphi_P^*(X) = 1\} | \theta) = P\left(\left\{X \in \Omega : \text{BF}(X) \leq \frac{b}{a}\right\} \mid \theta\right). \quad (2.11)$$

Therefore, the optimal averaged error probabilities from the generalized Neyman-Pearson Lemma, that will depend on the sample size, are

$$\alpha_{\varphi_P^*} = E_\theta [\pi_{\varphi_P^*}(\theta) | \mathbf{H}], \quad (2.12)$$

$$\beta_{\varphi_P^*} = E_\theta [1 - \pi_{\varphi_P^*}(\theta) | \mathbf{A}], \quad (2.13)$$

which can also be calculated as

$$\alpha_{\varphi_P^*} = P(\{X \in \Omega : \varphi_P^*(X) = 1\} | \mathbf{H}) = P\left(\left\{X \in \Omega : \text{BF}(X) \leq \frac{b}{a}\right\} \mid \mathbf{H}\right), \quad (2.14)$$

$$\beta_{\varphi_P^*} = P(\{X \in \Omega : \varphi_P^*(X) = 0\} | \mathbf{A}) = P\left(\left\{X \in \Omega : \text{BF}(X) > \frac{b}{a}\right\} \mid \mathbf{A}\right). \quad (2.15)$$

Consequently,

$$\alpha_{\varphi_P^*} = \begin{cases} \sum_{x \in \Psi_P^*} f_{\mathbf{H}}(x) & \text{if } \Omega \text{ is discrete} \\ \int_{x \in \Psi_P^*} f_{\mathbf{H}}(x) dx & \text{if } \Omega \text{ is continuous.} \end{cases} \quad (2.16)$$

$$\beta_{\varphi_P^*} = \begin{cases} \sum_{x \notin \Psi_P^*} f_{\mathbf{A}}(x) & \text{if } \Omega \text{ discrete} \\ \int_{x \notin \Psi_P^*} f_{\mathbf{A}}(x) dx & \text{if } \Omega \text{ continuous,} \end{cases} \quad (2.17)$$

where Ψ_P^* is the set defined as

$$\Psi_P^* = \left\{ x \in \Omega : \text{BF}(x) \leq \frac{b}{a} \right\}. \quad (2.18)$$

In order to make a decision, the P -value, computed as shown in Section 2.3, is compared to the optimal adaptive significance level $\alpha_{\varphi_P^*}$. Then, the hypothesis \mathbf{H} will be rejected when x_0 is observed if $P\text{-value}(x_0) < \alpha_{\varphi_P^*}$.

According to Gannon *et al.* (2019), once a significance level or cutoff for the P -value is determined, a corresponding cutoff for Bayes factors at the same significance level can be determined, and this is because of the one-to-one correspondence between them.

2.5 Adaptive cutoff values for evidence in the FBST

Considering the concepts in Pereira (1985), in Oliveira (2014) and the recent work of Pereira *et al.* (2017) and Gannon *et al.* (2019) related to adaptive significance levels, we propose to establish a cutoff value k^* for the $ev(\mathbf{H}; x_0)$ in the FBST as a function of the sample size n and the dimensionality of the parameter space d , i.e., $k^* = k^*(n, d)$ with $k^* \in [0, 1]$, such that k^* minimizes the linear combination of the averaged type-I and type-II error probabilities, $a\alpha_{\varphi_e} + b\beta_{\varphi_e}$. To describe the procedure, consider the test

$$\varphi_e(x) = \begin{cases} 0 & \text{if } ev(\mathbf{H}; x) > k \\ 1 & \text{if } ev(\mathbf{H}; x) \leq k. \end{cases} \quad (2.19)$$

Thus, the power function and the averaged error probabilities are respectively

$$\pi_{\varphi_e}(\theta) = P(\{X \in \Omega : \varphi_e(X) = 1\} | \theta) = P(\{X \in \Omega : ev(\mathbf{H}; X) \leq k\} | \theta), \quad (2.20)$$

$$\alpha_{\varphi_e} = E_{\theta}[\pi_{\varphi_e}(\theta) | \mathbf{H}], \quad (2.21)$$

$$\beta_{\varphi_e} = E_{\theta}[1 - \pi_{\varphi_e}(\theta) | \mathbf{A}]. \quad (2.22)$$

Alternatively, it is possible to calculate (2.21) and (2.22) as

$$\alpha_{\varphi_e} = P(\{X \in \Omega : \varphi_e(X) = 1\} | \mathbf{H}) = P(\{X \in \Omega : ev(\mathbf{H}; X) \leq k\} | \mathbf{H}), \quad (2.23)$$

$$\beta_{\varphi_e} = P(\{X \in \Omega : \varphi_e(X) = 0\} | \mathbf{A}) = P(\{X \in \Omega : ev(\mathbf{H}; X) > k\} | \mathbf{A}). \quad (2.24)$$

The averaged error probabilities can also be expressed in terms of the Bayesian prior predictive densities under the respective hypotheses defined in (2.2) and (2.3), as follows

$$\alpha_{\varphi_e} = \begin{cases} \sum_{x \in \Psi_e} f_{\mathbf{H}}(x) & \text{if } \Omega \text{ is discrete} \\ \int_{x \in \Psi_e} f_{\mathbf{H}}(x) dx & \text{if } \Omega \text{ is continuous.} \end{cases} \quad (2.25)$$

$$\beta_{\varphi_e} = \begin{cases} \sum_{x \notin \Psi_e} f_{\mathbf{A}}(x) & \text{if } \Omega \text{ is discrete} \\ \int_{x \notin \Psi_e} f_{\mathbf{A}}(x) dx & \text{if } \Omega \text{ is continuous,} \end{cases} \quad (2.26)$$

where Ψ_e is the set

$$\Psi_e = \{x \in \Omega : ev(\mathbf{H}; x) \leq k\}. \quad (2.27)$$

So the adaptive cutoff value k^* for $ev(\mathbf{H}; x)$ will be the value of k that minimizes $a\alpha_{\varphi_e} + b\beta_{\varphi_e}$, for fixed $a, b > 0$.

Finally, consider the test

$$\varphi_e^*(x) = \begin{cases} 0 & \text{if } ev(\mathbf{H}; x) > k^* \\ 1 & \text{if } ev(\mathbf{H}; x) \leq k^*. \end{cases} \quad (2.28)$$

Then the optimal averaged error probabilities that depend on the sample size will be

$$\alpha_{\varphi_e^*}^* = P(\{X \in \Omega : \varphi_e^*(X) = 1\} | \mathbf{H}) = P(\{X \in \Omega : ev(\mathbf{H}; X) \leq k^*\} | \mathbf{H}) \quad (2.29)$$

$$\beta_{\varphi_e^*}^* = P(\{X \in \Omega : \varphi_e^*(X) = 0\} | \mathbf{A}) = P(\{X \in \Omega : ev(\mathbf{H}; X) > k^*\} | \mathbf{A}). \quad (2.30)$$

Simple Illustrative Examples

In this chapter, two examples of bilateral hypothesis testing for θ ($\mathbf{H} : \theta = \theta_0$ vs. $\mathbf{A} : \theta \neq \theta_0$) are presented, where a precise or “sharp” hypothesis against a composite alternative is tested and also a one-dimensional parameter space is defined. The objective is to illustrate the proposed methodology in the previous chapter to determine the adaptive cutoff value k^* for $ev(\mathbf{H}; x_0)$ that minimizes the linear combination of the averaged error probabilities $a\alpha_{\varphi_e} + b\beta_{\varphi_e}$. On the other hand, the adaptive significance levels for decisions using P -values are presented in order to compare and also to establish a possible relationship with the results from FBST. In the first example, a discrete distribution for X is considered, $X|\theta \sim \text{Bernoulli}(\theta)$, so the parameter of interest in the test is a proportion. In the second one, X obeys a continuous distribution, $X|\theta \sim \text{Normal}(\theta, \sigma^2)$, so the test is related to the mean. Furthermore, conjugate priors are taken into account to make calculations easier (see [Paulino *et al.*, 2003](#)). In both situations the values of a and b , which represent the relative seriousness of errors, are $a = b = 1$, that is, the two types of errors are equally severe.

3.1 Adaptive significance levels in proportion hypothesis testing

Consider conditionally independent and identically distributed random variables (*c.i.i.d*) $X_1, \dots, X_n|\theta$ from a Bernoulli distribution with unknown parameter $\theta \in [0, 1]$, and define $X = (X_1, \dots, X_n)$. Consider also that $\theta \sim \text{Beta}(\gamma, \delta)$ ($\gamma > 0$ and $\delta > 0$) is a prior for θ . Then

the posterior distribution of θ given that $X = x$ can be calculated as

$$\begin{aligned} f(\theta|x) &\propto P(X = x|\theta)g(\theta) \\ &\propto \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \theta^{\gamma-1} (1 - \theta)^{\delta-1} \\ &\propto \theta^{\gamma + \sum_{i=1}^n x_i - 1} (1 - \theta)^{\delta + n - \sum_{i=1}^n x_i - 1}, \quad \theta \in (0, 1), x_i = \{0, 1\}. \end{aligned}$$

Therefore, the posterior distribution is

$$\theta|x \sim \text{Beta} \left(\gamma + \sum_{i=1}^n x_i, \delta + n - \sum_{i=1}^n x_i \right).$$

The hypotheses to be tested are given by

$$\mathbf{H} : \theta = \theta_0$$

$$\mathbf{A} : \theta \neq \theta_0,$$

where $\theta_0 \in (0, 1)$.

Let $X^* = T(X) = \sum_{i=1}^n X_i$ be the sufficient statistic and also let $\Omega^* = \{0, 1, \dots, n\}$ be the space induced by X^* . Then, $X^*|\theta \sim \text{Binomial}(n, \theta)$ and $f(\theta|x) = f(\theta|x^*)$, where $x^* = T(x)$.

3.1.1 Evidence index: e -value

With $\Theta_{\mathbf{H}} = \{\theta_0\}$, the tangential set to the null hypothesis and the evidence in favor of \mathbf{H} for $x_0^* \in \Omega^*$ are, respectively,

$$T_{x_0^*} = \left\{ \theta \in \Theta : f(\theta|x_0^*) > \sup_{\mathbf{H}} f(\theta|x_0^*) \right\} = \{ \theta \in \Theta : f(\theta|x_0^*) > f(\theta_0|x_0^*) \},$$

$$ev(\mathbf{H}; x_0^*) = 1 - P(\theta \in T_{x_0^*} | x_0^*).$$

Let $\varphi_e(x^*)$ be the test given by

$$\varphi_e(x^*) = \begin{cases} 0 & \text{if } ev(\mathbf{H}; x^*) > k \\ 1 & \text{if } ev(\mathbf{H}; x^*) \leq k. \end{cases}$$

So the power function and the averaged error probabilities are, respectively,

$$\begin{aligned} \pi_{\varphi_e}(\theta) &= P(\{X^* \in \Omega^* : \varphi_e^*(X^*) = 1\}|\theta) \\ &= P(\{X^* \in \Omega^* : ev(\mathbf{H}; X^*) \leq k\}|\theta) \\ &= \sum_{x^*=0}^n P(X^* = x^*|\theta) \mathbb{1}(ev(\mathbf{H}; x^*) \leq k) \\ &= \sum_{x^*=0}^n \binom{n}{x^*} \theta^{x^*} (1 - \theta)^{n-x^*} \mathbb{1}(ev(\mathbf{H}; x^*) \leq k). \end{aligned}$$

$$\begin{aligned} \alpha_{\varphi_e} &= P(\{X^* \in \Omega^* : \varphi_e^*(X^*) = 1\}|\mathbf{H}) \\ &= P(\{X^* \in \Omega^* : ev(\mathbf{H}; X^*) \leq k\}|\mathbf{H}) \\ &= \sum_{x^*=0}^n P(X^* = x^*|\theta = \theta_0) \mathbb{1}(ev(\mathbf{H}; x^*) \leq k) \\ &= \sum_{x^*=0}^n \binom{n}{x^*} \theta_0^{x^*} (1 - \theta_0)^{n-x^*} \mathbb{1}(ev(\mathbf{H}; x^*) \leq k). \end{aligned}$$

$$\begin{aligned} \beta_{\varphi_e} &= E_{\theta} [1 - \pi_{\varphi_e}(\theta)|\mathbf{A}] \\ &= \int_{\mathbf{A}} [1 - \pi_{\varphi_e}(\theta)] g(\theta) d\theta. \end{aligned}$$

For example, taking the particular case in which $n = 2$ and the prior $\theta \sim Beta(1, 1)$ ($Uniform(0, 1)$), then $X^*|\theta \sim Binomial(2, \theta)$ and $\theta|x^* \sim Beta(1 + x^*, 3 - x^*)$. Consider the hypotheses

$$\mathbf{H} : \theta = 0.5$$

$$\mathbf{A} : \theta \neq 0.5.$$

Below, according to the value that the random variable x^* takes, the procedure to calculate the evidence is presented.

- When $x^* = 0$,

$$\begin{aligned} \theta|x \sim \text{Beta}(1, 3) \quad \Rightarrow \quad f(\theta|x^* = 0) &= \frac{\Gamma(1+3)}{\Gamma(1)\Gamma(3)} (1-\theta)^2 \\ &= 3(1-\theta)^2, \quad \theta \in (0, 1), \end{aligned}$$

where Γ is the Gamma function.

The evidence in favor of \mathbf{H} is calculated as

$$\begin{aligned} ev(0.5; 0) &= 1 - P(\theta \in T_{x_0^*} | x^* = x_0^*) \\ &= 1 - P(\theta < 0.5 | x^* = 0) \\ &= 1 - 3 \int_0^{0.5} (1-\theta)^2 d\theta \\ &= 0.125. \end{aligned}$$

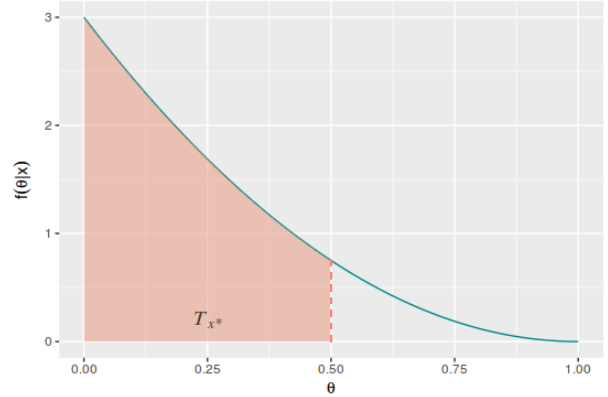


Figure 3.1: Tangential set to $\mathbf{H} : \theta = 0.5$, with $\theta|x^* \sim \text{Beta}(1, 3)$.

- When $x^* = 1$,

$$\begin{aligned} \theta|x \sim \text{Beta}(2, 2) \quad \Rightarrow \quad f(\theta|x^* = 1) &= \frac{\Gamma(2+2)}{\Gamma(2)\Gamma(2)} \theta(1-\theta) \\ &= 6\theta(1-\theta), \quad \theta \in (0, 1). \end{aligned}$$

The evidence in favor of \mathbf{H} is calculated as

$$\begin{aligned} ev(0.5; 1) &= 1 - P(\theta \in T_{x_0^*} | x^* = x_0^*) \\ &= 1 - 0 \\ &= 1. \end{aligned}$$

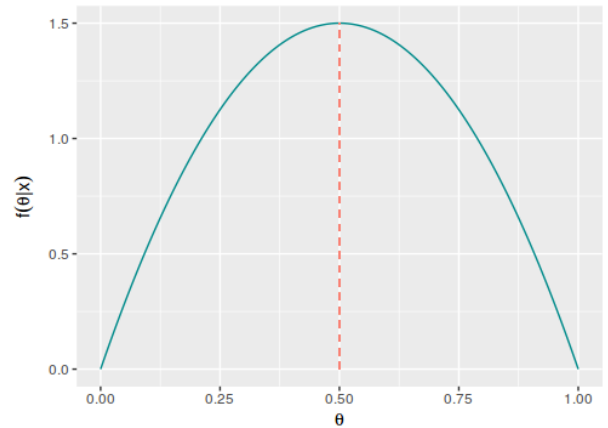


Figure 3.2: Tangential set to $\mathbf{H} : \theta = 0.5$, with $\theta|x^* \sim \text{Beta}(2, 2)$.

- When $x^* = 2$,

$$\begin{aligned} \theta|x \sim \text{Beta}(3, 1) \quad \Rightarrow \quad f(\theta|x^* = 2) &= \frac{\Gamma(1+3)}{\Gamma(1)\Gamma(3)} \theta^2 \\ &= 3\theta^2, \quad \theta \in (0, 1). \end{aligned}$$

The evidence in favor of \mathbf{H} is calculated as

$$\begin{aligned} ev(0.5; 2) &= 1 - P(\theta \in T_{x_0^*} | x^* = x_0^*) \\ &= 1 - P(\theta > 0.5 | x^* = 2) \\ &= 1 - 3 \int_{0.5}^1 \theta^2 d\theta \\ &= 0.125. \end{aligned}$$

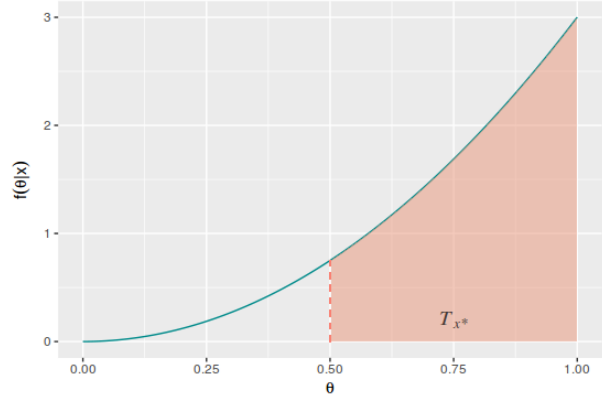


Figure 3.3: Tangential set to \mathbf{H} : $\theta = 0.5$, with $\theta|x^* \sim \text{Beta}(3, 1)$.

In this way, the power function and the averaged error probabilities can be calculated as follows

$$\begin{aligned} \pi_{\varphi_e}(\theta) &= P(\{X^* \in \Omega^* : ev(0.5; X^*) \leq k\} | \theta) \\ &= \sum_{x^*=0}^n P(X^* = x^* | \theta) \mathbb{1}(ev(0.5; x^*) \leq k) \\ &= \sum_{x^*=0}^2 \binom{2}{x^*} \theta^{x^*} (1-\theta)^{2-x^*} \mathbb{1}(ev(0.5; x^*) \leq k) \\ &= (1-\theta)^2 \mathbb{1}(ev(0.5; 0) \leq k) + 2\theta(1-\theta) \mathbb{1}(ev(0.5; 1) \leq k) + \theta^2 \mathbb{1}(ev(0.5; 2) \leq k) \\ &= (1-\theta)^2 \mathbb{1}(k \geq 0.125) + 2\theta(1-\theta) \mathbb{1}(k \geq 1) + \theta^2 \mathbb{1}(k \geq 0.125). \end{aligned}$$

$$\begin{aligned} \alpha_{\varphi_e} &= P(\{X^* \in \Omega^* : ev(0.5; X^*) \leq k\} | \mathbf{H}) \\ &= \sum_{x^*=0}^n P(X^* = x^* | \theta_0 = 0.5) \mathbb{1}(ev(0.5; x^*) \leq k) \\ &= \sum_{x^*=0}^2 \binom{2}{x^*} \theta_0^{x^*} (1-\theta_0)^{2-x^*} \mathbb{1}(ev(0.5; x^*) \leq k) \end{aligned}$$

$$\begin{aligned}
&= (1 - \theta_0)^2 \mathbb{1}(ev(0.5; 0) \leq k) + 2\theta_0(1 - \theta_0) \mathbb{1}(ev(0.5; 1) \leq k) + \theta_0^2 \mathbb{1}(ev(0.5; 2) \leq k) \\
&= (0.5)^2 \mathbb{1}(k \geq 0.125) + 2(0.5)^2 \mathbb{1}(k \geq 1) + (0.5)^2 \mathbb{1}(k \geq 0.125).
\end{aligned}$$

$$\begin{aligned}
\beta_{\varphi_e} &= E_{\theta} [1 - \pi_{\varphi_e}(\theta) | \mathbf{A}] \\
&= \int_{\mathbf{A}} [1 - \pi_{\varphi_e}(\theta)] g(\theta) d\theta \\
&= \int_{\mathbf{A}} [1 - \pi_{\varphi_e}(\theta)] \mathbb{1} d\theta \\
&= 1 - \int_0^1 (1 - \theta)^2 \mathbb{1}(k \geq 0.125) d\theta + \int_0^1 2\theta(1 - \theta) \mathbb{1}(k \geq 1) d\theta + \int_0^1 \theta^2 \mathbb{1}(k \geq 0.125) d\theta \\
&= 1 - \left[\frac{\Gamma(1)\Gamma(3)}{\Gamma(1+3)} \mathbb{1}(k \geq 0.125) + 2 \frac{\Gamma(2)\Gamma(2)}{\Gamma(2+2)} \mathbb{1}(k \geq 1) + \frac{\Gamma(1)\Gamma(3)}{\Gamma(1+3)} \mathbb{1}(k \geq 0.125) \right] \\
&= 1 - \left[\frac{1}{3} \mathbb{1}(k \geq 0.125) + \frac{2}{6} \mathbb{1}(k \geq 1) + \frac{1}{3} \mathbb{1}(k \geq 0.125) \right].
\end{aligned}$$

Therefore, the averaged type-I and type-II error probabilities are given by

$$\alpha_{\varphi_e} = \begin{cases} 0, & 0 \leq k < 0.125 \\ 0.5, & 0.125 \leq k < 1 \\ 1, & k = 1. \end{cases} \quad \beta_{\varphi_e} = \begin{cases} 1, & 0 \leq k < 0.125 \\ 0.333, & 0.125 \leq k < 1 \\ 0, & k = 1. \end{cases}$$

In Figures 3.4. (a) and 3.4. (b) respectively, the averaged error probabilities as a function of k and the power function for a fixed k are plotted.

Because of the discrete nature of X^* , the probabilities of both kind of errors and their linear combination as function of k are simple functions, that is, can be written as $g(k) = \sum_{i=1}^r a_i \mathbb{1}(k \in A_i)$, where $a_i \in \mathbb{R}_+$, $\forall i$, and A_1, \dots, A_r form a partition of $[0, 1]$. Therefore, $\alpha_{\varphi_e} + \beta_{\varphi_e}$ is minimized in a range of values of k , i.e., is minimized by any value of k^* such that $k^* \in [k_{\text{inf}}^*, k_{\text{sup}}^*]$.

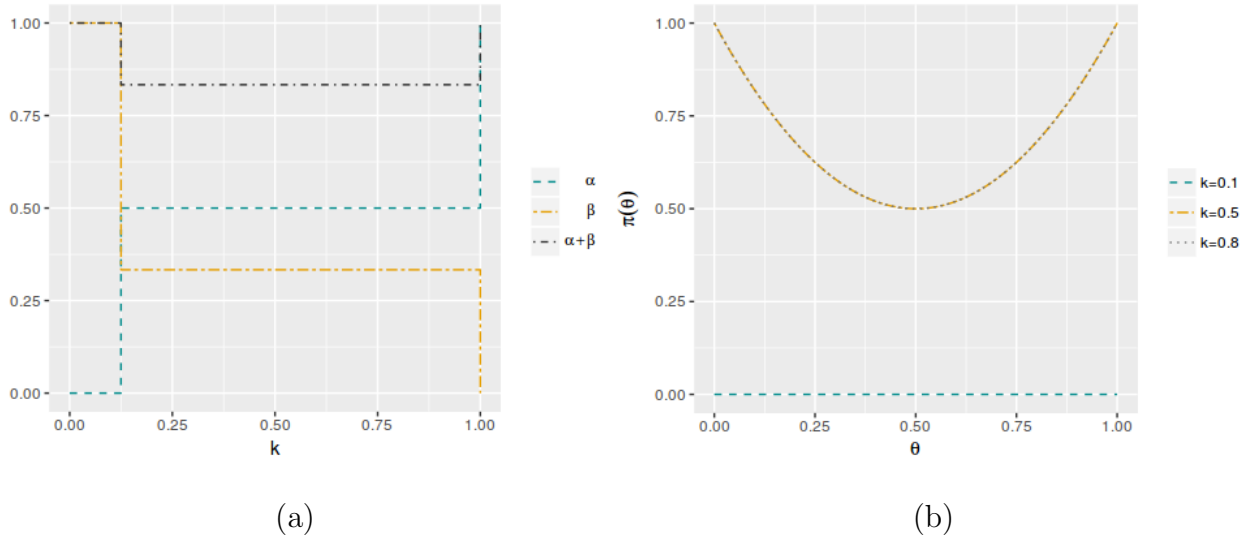


Figure 3.4: (a) Averaged error probabilities (α_{φ_e} , β_{φ_e} and $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as functions of k , with $n = 2$ and $\theta \sim \text{Beta}(1, 1)$. (b) Power function for $k = 0.1$, $k = 0.5$ and $k = 0.8$, with $n = 2$ e $\theta \sim \text{Beta}(1, 1)$.

Table 3.1 and Figure 3.5 show the minimum k^* , middle k^* and maximum k^* values that minimize the sum of the error probabilities.

minimum k^* (k_{inf}^*)	0.1251
middle k^* (\bar{k}^*)	0.5625
maximum k^* (k_{sup}^*)	0.9999

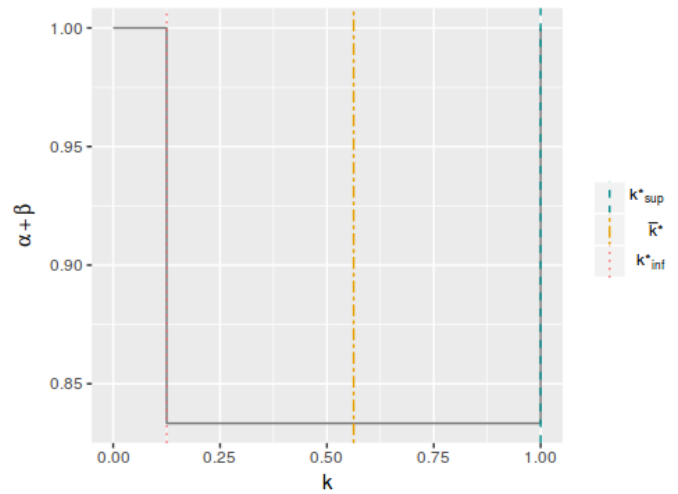


Table 3.1: Cutoff values k^* for $ev(0.5; x_0^*)$, with $n = 2$ and $\theta \sim \text{Beta}(1, 1)$. **Figure 3.5:** Linear combination of averaged error probabilities ($\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k , with $n = 2$ and $\theta \sim \text{Beta}(1, 1)$.

In order to evaluate how much the choice of the prior affects the cutoff value that we are looking for, consider beyond the Uniform prior, two informative distributions for θ , $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$ (Figure 3.6).

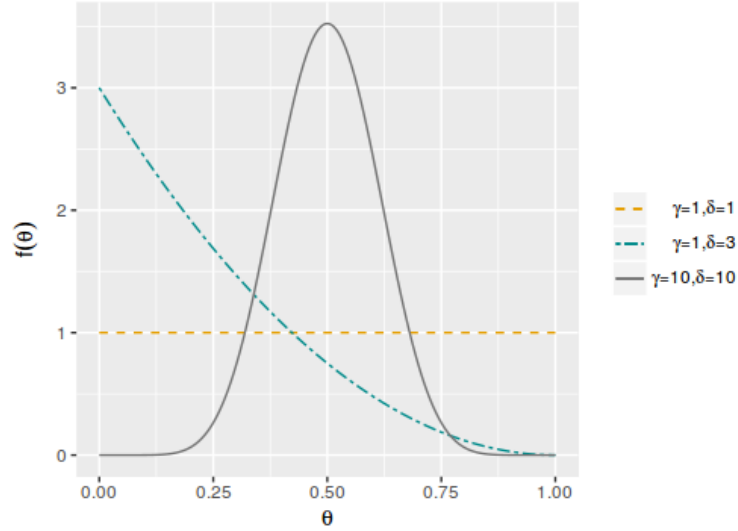


Figure 3.6: Prior distributions $\theta \sim \text{Beta}(1,1)$, $\theta \sim \text{Beta}(1,3)$ and $\theta \sim \text{Beta}(10,10)$.

In Figure 3.7 and 3.8 are plotted for the three priors and a sample size $n = 40$, the averaged error probabilities as a function of k (Figures 3.7.(a), 3.7.(c) and 3.7.(e)), the power function for a fixed k (Figures 3.7.(b), 3.7.(d) and 3.7.(f)) and the minimum k^* , middle k^* and maximum k^* values that minimize the sum of the error probabilities (Figure 3.8).

The procedure presented to find the $k^* \in [k_{\text{inf}}^*, k_{\text{sup}}^*]$ values for a fixed n is replicated for different sample sizes and taking into account the prior distributions for θ . In Tables 3.2, 3.3, 3.4 and Figures 3.9, 3.10, 3.11 the results are displayed. Note that by increasing n , the k^* values show a decreasing trend in the three cases ($\text{Beta}(1,1)$, $\text{Beta}(1,3)$ and $\text{Beta}(10,10)$), which allows us to interpret that the influence of sample size on the determination of the cutoff for $ev(0.5; x_0^*)$ is very relevant.

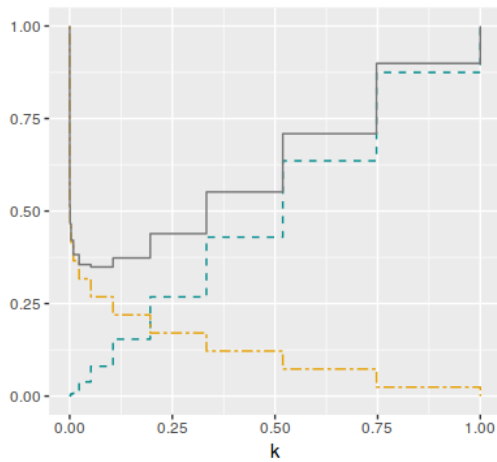
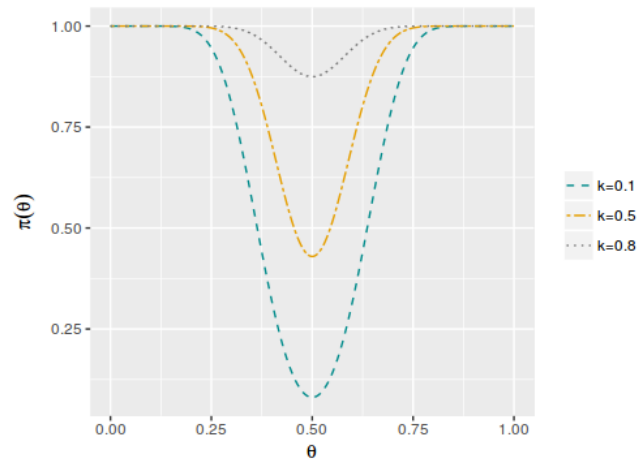
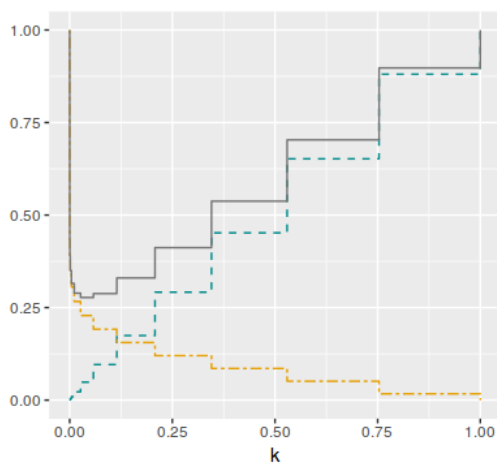
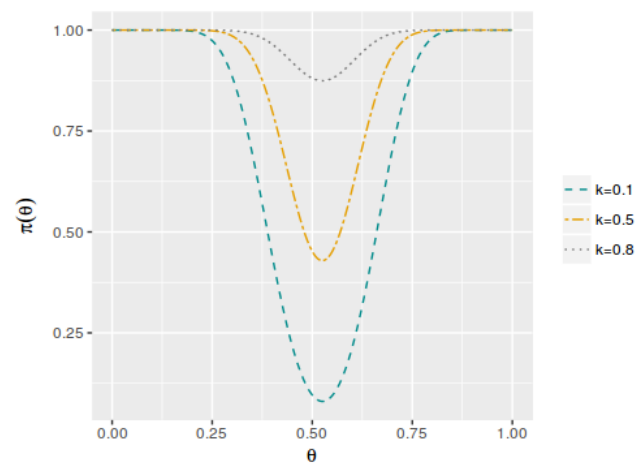
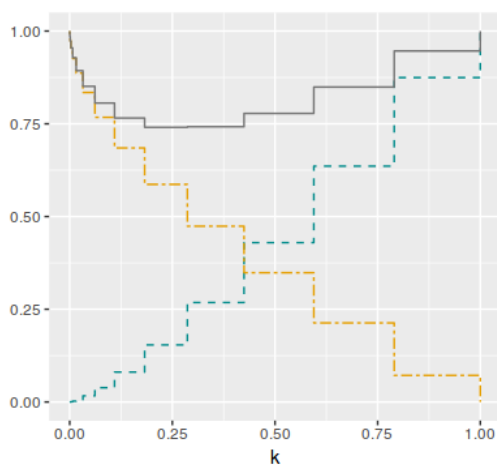
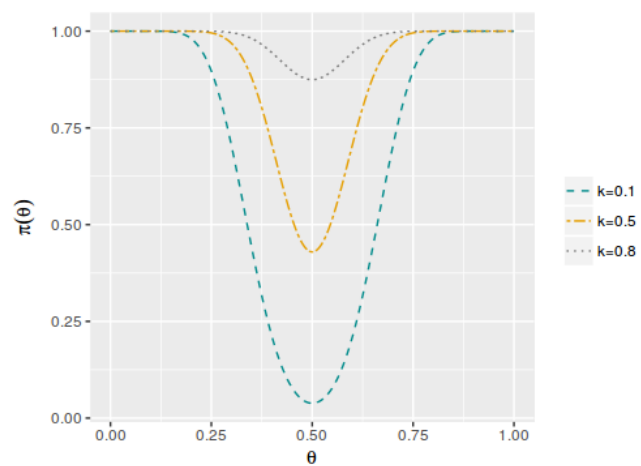
(a) $Beta(1,1)$ (b) $Beta(1,1)$ (c) $Beta(1,3)$ (d) $Beta(1,3)$ (e) $Beta(10,10)$ (f) $Beta(10,10)$

Figure 3.7: (a), (c) and (e) Averaged error probabilities (α_{φ_e} , β_{φ_e} e $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k ; (b), (d) and (f) Power function when $k = 0.1$, $k = 0.5$ and $k = 0.8$. Sample size $n = 40$.

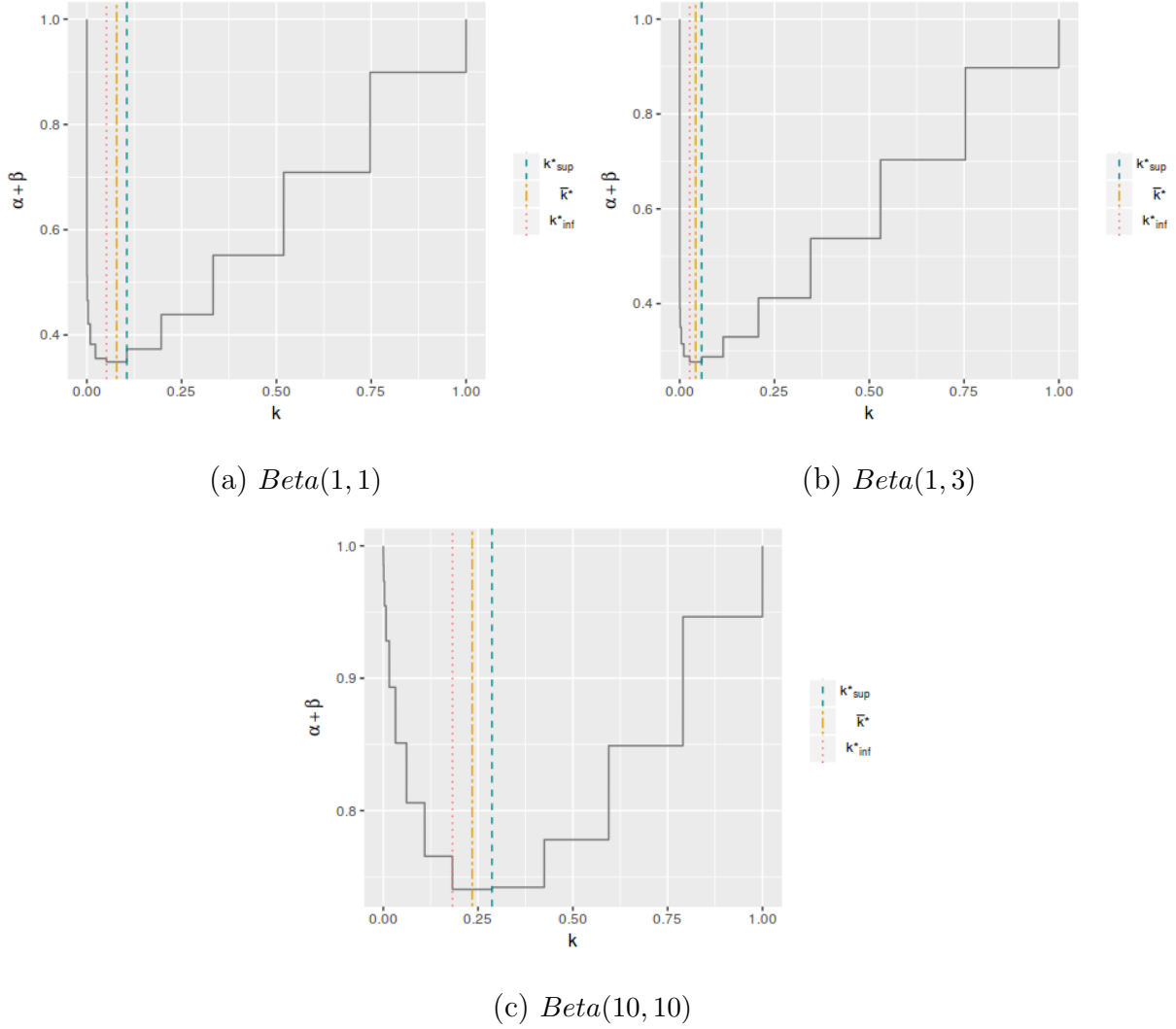


Figure 3.8: Linear combination of averaged error probabilities ($\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k . Sample size $n = 40$.

n	k_{inf}^*	\bar{k}^*	k_{sup}^*
10	0.03910	0.10530	0.17150
20	0.06080	0.11115	0.16150
30	0.05930	0.09585	0.13240
40	0.05170	0.07860	0.10550
50	0.04320	0.06340	0.08360
60	0.03550	0.05075	0.06600
70	0.02890	0.04055	0.05220
80	0.04140	0.05570	0.07000
90	0.03290	0.04390	0.05490
100	0.02610	0.03465	0.04320
150	0.02130	0.02685	0.03240
200	0.02290	0.02790	0.03290
250	0.02220	0.02645	0.03070
300	0.02050	0.02405	0.02760

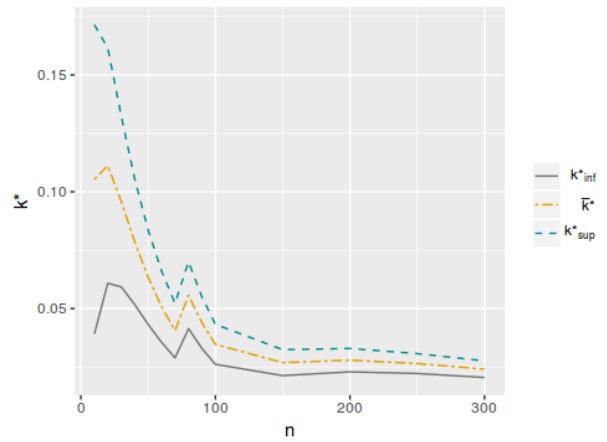


Figure 3.9: Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim Beta(1,1)$.

Table 3.2: Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim Beta(1,1)$.

n	k_{inf}^*	\bar{k}^*	k_{sup}^*
10	0.06270	0.14015	0.21760
20	0.02540	0.05025	0.07510
30	0.02830	0.04835	0.06840
40	0.02660	0.04220	0.05780
50	0.02340	0.03550	0.04760
60	0.01990	0.02925	0.03860
70	0.03130	0.04350	0.05570
80	0.02520	0.03460	0.04400
90	0.02020	0.02750	0.03480
100	0.01620	0.02190	0.02760
150	0.01420	0.01815	0.02210
200	0.01620	0.01985	0.02350
250	0.01620	0.01945	0.02270
300	0.01530	0.01805	0.02080

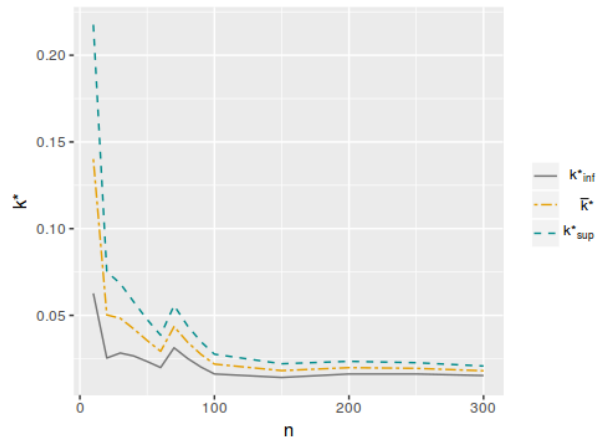


Figure 3.10: Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim \text{Beta}(1, 3)$.

Table 3.3: Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim \text{Beta}(1, 3)$.

n	k_{inf}^*	\bar{k}^*	k_{sup}^*
10	0.43720	0.56755	0.69790
20	0.32010	0.41395	0.50780
30	0.24000	0.30925	0.37850
40	0.18260	0.23460	0.28660
50	0.21950	0.27285	0.32620
60	0.16940	0.21085	0.25230
70	0.19640	0.23910	0.28180
80	0.15350	0.18740	0.22130
90	0.17440	0.20940	0.24440
100	0.13770	0.16590	0.19410
150	0.12070	0.14165	0.16260
200	0.10250	0.11850	0.13450
250	0.08610	0.09855	0.11100
300	0.09150	0.10335	0.11520

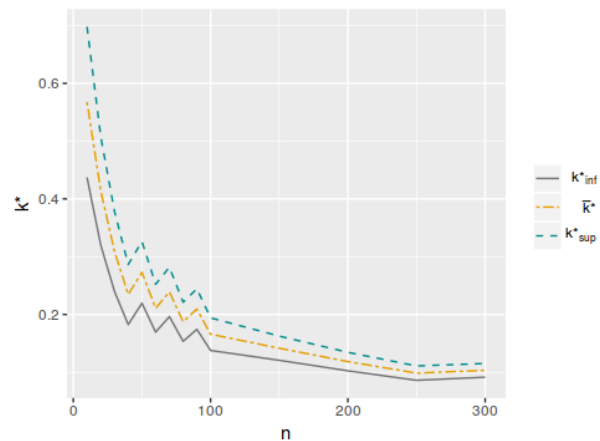


Figure 3.11: Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim \text{Beta}(10, 10)$.

Table 3.4: Cutoff values k^* for $ev(0.5; x_0^*)$ as function of n , with $\theta \sim \text{Beta}(10, 10)$.

Since for any $k^* \in [k_{\text{inf}}^*, k_{\text{sup}}^*]$ the sum of the averaged error probabilities, $\alpha_{\varphi_e} + \beta_{\varphi_e}$, is minimal, we propose to choose the minimum k^* as a cutoff value for $ev(0.5; x_0^*)$, because it could be more conservative, that is, it could reject the null hypothesis less. Thus, in Table 3.5 and Figure 3.12 the comparisons of the minimum k^* as function of sample size for each of the predetermined priors is exposed. The differences in the results between priors are evident, and therefore, their influence in the moment of choosing the cutoff value should be taken into account. In the $\theta \sim \text{Beta}(10, 10)$ case, for example, the k^* value is higher than in the other two priors.

n	k_{inf}^*		
	$\gamma = 1,$ $\delta = 1$	$\gamma = 1,$ $\delta = 3$	$\gamma = 10,$ $\delta = 10$
10	0.03910	0.06270	0.43720
20	0.06080	0.02540	0.32010
30	0.05930	0.02830	0.24000
40	0.05170	0.02660	0.18260
50	0.04320	0.02340	0.21950
60	0.03550	0.01990	0.16940
70	0.02890	0.03130	0.19640
80	0.04140	0.02520	0.15350
90	0.03290	0.02020	0.17440
100	0.02610	0.01620	0.13770
150	0.02130	0.01420	0.12070
200	0.02290	0.01620	0.10250
250	0.02220	0.01620	0.08610
300	0.02050	0.01530	0.09150

Table 3.5: Minimum k^* as function of n , for $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$.

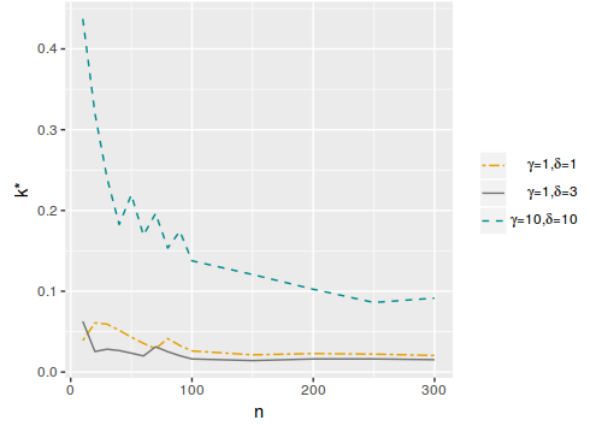


Figure 3.12: Minimum k^* as function of n , for $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$.

Finally, Figure 3.13 presents for each prior the optimal averaged error probabilities $\alpha_{\varphi_e}^*$ and $\beta_{\varphi_e}^*$, that is, the type-I and type-II error probabilities for which $\alpha_{\varphi_e} + \beta_{\varphi_e}$ is minimal as a function of n . Note that, the probabilities of both kind of errors and their linear combination generally decrease with increasing sample size.

3.1.2 Significance index: P -value

Define the same hypotheses to be tested

$$\mathbf{H} : \theta = 0.5$$

$$\mathbf{A} : \theta \neq 0.5.$$

The prior predictive probability functions under the two hypotheses are

$$\begin{aligned} f_{\mathbf{H}}(x^*) &= \int_{\mathbf{H}} f(x^* | \theta) g_{\mathbf{H}}(\theta) d\theta \\ &= f(x^* | \theta = 0.5) \\ &= \binom{n}{x^*} (0.5)^n, \quad x_0^* \in \Omega^*. \end{aligned}$$

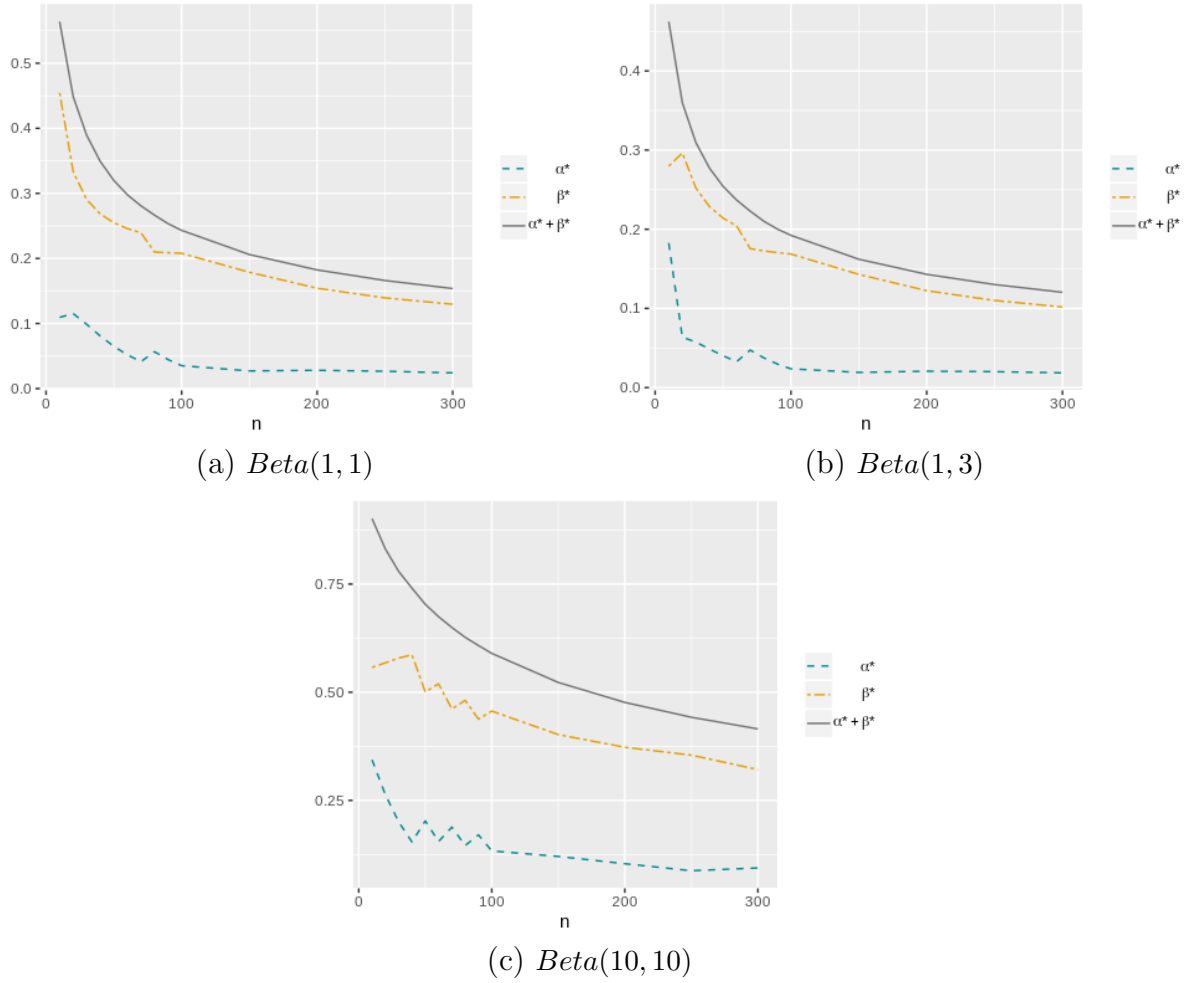


Figure 3.13: Optimal averaged error probabilities ($\alpha_{\varphi_e^*}^*$, $\beta_{\varphi_e^*}^*$ and $\alpha_{\varphi_e^*}^* + \beta_{\varphi_e^*}^*$) as function of n .

and

$$\begin{aligned}
 f_{\mathbf{A}}(x^*) &= \int_{\mathbf{A}} f(x^* | \theta) g_{\mathbf{A}}(\theta) d\theta \\
 &= \int_{\mathbf{A}} f(x^* | \theta) g(\theta) d\theta \\
 &= \binom{n}{x^*} \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma)\Gamma(\delta)} \int_0^1 \theta^{\gamma+x^*-1} (1-\theta)^{\delta+n-x^*-1} d\theta \\
 &= \binom{n}{x^*} \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma)\Gamma(\delta)} \frac{\Gamma(\gamma + x^*)\Gamma(\delta + n - x^*)}{\Gamma(\gamma + \delta + n)}.
 \end{aligned}$$

Let $\varphi_P^*(x^*)$ be the test given by

$$\varphi_P^*(x^*) = \begin{cases} 0 & \text{if } \text{BF}(x^*) > 1 \\ 1 & \text{if } \text{BF}(x^*) \leq 1. \end{cases}$$

To obtain the P -value, define the set Ψ_P of sample points x^* for which the Bayes factors are smaller than or equal to the Bayes factor of the observed sample point x_0^* , that is

$$\Psi_P = \{x^* \in \Omega^* : \text{BF}(x^*) \leq \text{BF}(x_0^*)\},$$

so the P -value is the sum of the predictive probabilities under \mathbf{H} in Ψ_P

$$P\text{-value}(x_0^*) = \sum_{\Psi_P} f_{\mathbf{H}}(x^*) = \sum_{\Psi_P} \binom{n}{x^*} (0.5)^n.$$

Now, defining the set Ψ_P^* of sample points x^* with Bayes factors smaller than or equal to $b/a = 1$, i.e.,

$$\Psi_P^* = \{x^* \in \Omega^* : \text{BF}(x^*) \leq 1\},$$

the optimal averaged error probabilities from the generalized Neyman-Pearson Lemma are

$$\alpha_{\varphi_P^*} = \sum_{x^* \in \Psi_P^*} f_{\mathbf{H}}(x^*) = \sum_{x^* \in \Psi_P^*} \binom{n}{x^*} (0.5)^n$$

and

$$\beta_{\varphi_P^*} = \sum_{x^* \notin \Psi_P^*} f_{\mathbf{A}}(x^*) = \sum_{x^* \notin \Psi_P^*} \binom{n}{x^*} \frac{\Gamma(\gamma + \delta)}{\Gamma(\gamma)\Gamma(\delta)} \frac{\Gamma(\gamma + x^*)\Gamma(\delta + n - x^*)}{\Gamma(\gamma + \delta + n)}.$$

In Table 3.6 and Figure 3.14 the comparisons of the optimal adaptive significance level for each of the predetermined priors are shown. When the prior is more informative, the $\theta \sim \text{Beta}(10, 10)$ case, the $\alpha_{\varphi_P^*}$ value is higher than in the other two priors.

Figure 3.15 presents the optimal adaptive significance level $\alpha_{\varphi_P^*}$, the optimal adaptive type-II averaged error probability and the minimized linear combination $\alpha_{\varphi_P^*} + \beta_{\varphi_P^*}$ as a function of the sample size for each prior. In general, the probabilities of both kind of errors and their linear combination decrease when the sample size increases.

n	$\alpha_{\varphi_P^*}$		
	$\gamma = 1, \delta = 1$	$\gamma = 1, \delta = 3$	$\gamma = 10, \delta = 10$
10	0.10938	0.18262	0.34375
20	0.11532	0.06357	0.26318
30	0.09874	0.05743	0.20049
40	0.08069	0.04864	0.15386
50	0.06491	0.04013	0.20264
60	0.05189	0.03269	0.15500
70	0.04139	0.04735	0.18822
80	0.05666	0.03749	0.14564
90	0.04460	0.02973	0.17024
100	0.03520	0.02362	0.13321
150	0.02715	0.01914	0.12053
200	0.02813	0.02063	0.10364
250	0.02666	0.02007	0.08750
300	0.02418	0.01855	0.09390

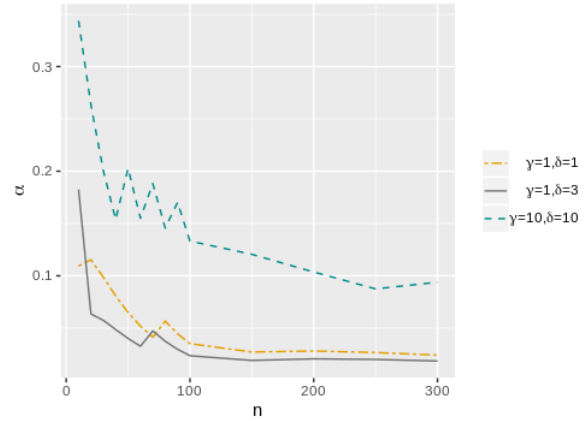


Figure 3.14: Optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as function of n , with $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$.

Table 3.6: Optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as function of n , with $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$.

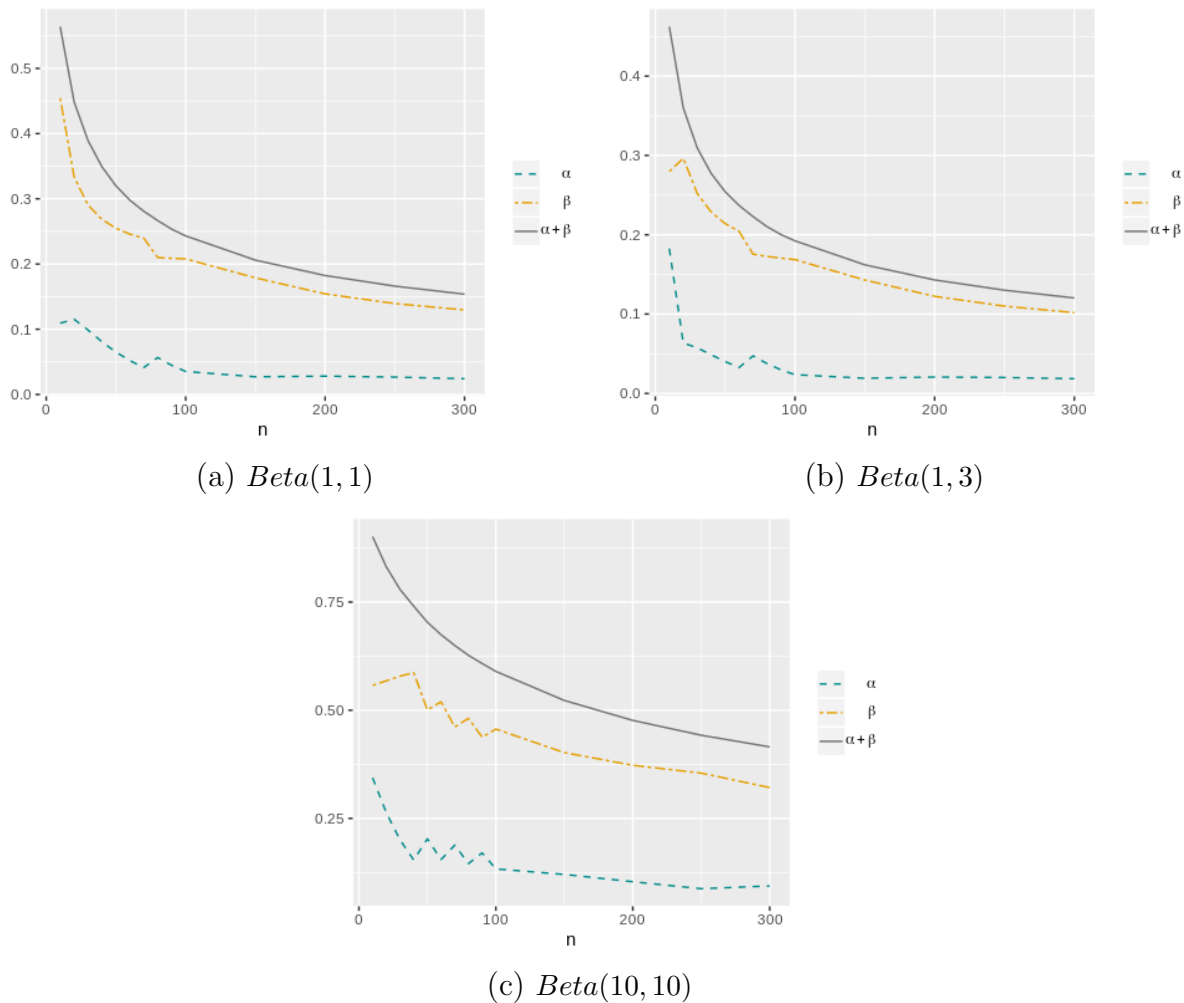


Figure 3.15: Optimal averaged error probabilities ($\alpha_{\varphi_P^*}$, $\beta_{\varphi_P^*}$ and $\alpha_{\varphi_P^*} + \beta_{\varphi_P^*}$) as function of n , with $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$.

3.1.3 Relationship between e -value and P -value

Following the ideas from [Diniz *et al.* \(2012\)](#) who illustrate how the frequentist p -value and the e -value have a straight mathematical relationship, we also want to establish a possible relationship between the P -value with the results from FBST. For that purpose, we simulated 1000 binomial random samples with parameter $\theta = 0.5$ to be used as observations to calculate the evidence and the significance indices. So, in [Figure 3.16](#) are plotted, for each prior, the e -value versus P -value for different sample sizes. Note that, for the smaller sample sizes, the e -values tend to be higher than the P -values, which is more evident when the prior is more informative, that is, when $\theta \sim \text{Beta}(10, 10)$. For larger sample sizes, the two values are almost the same.

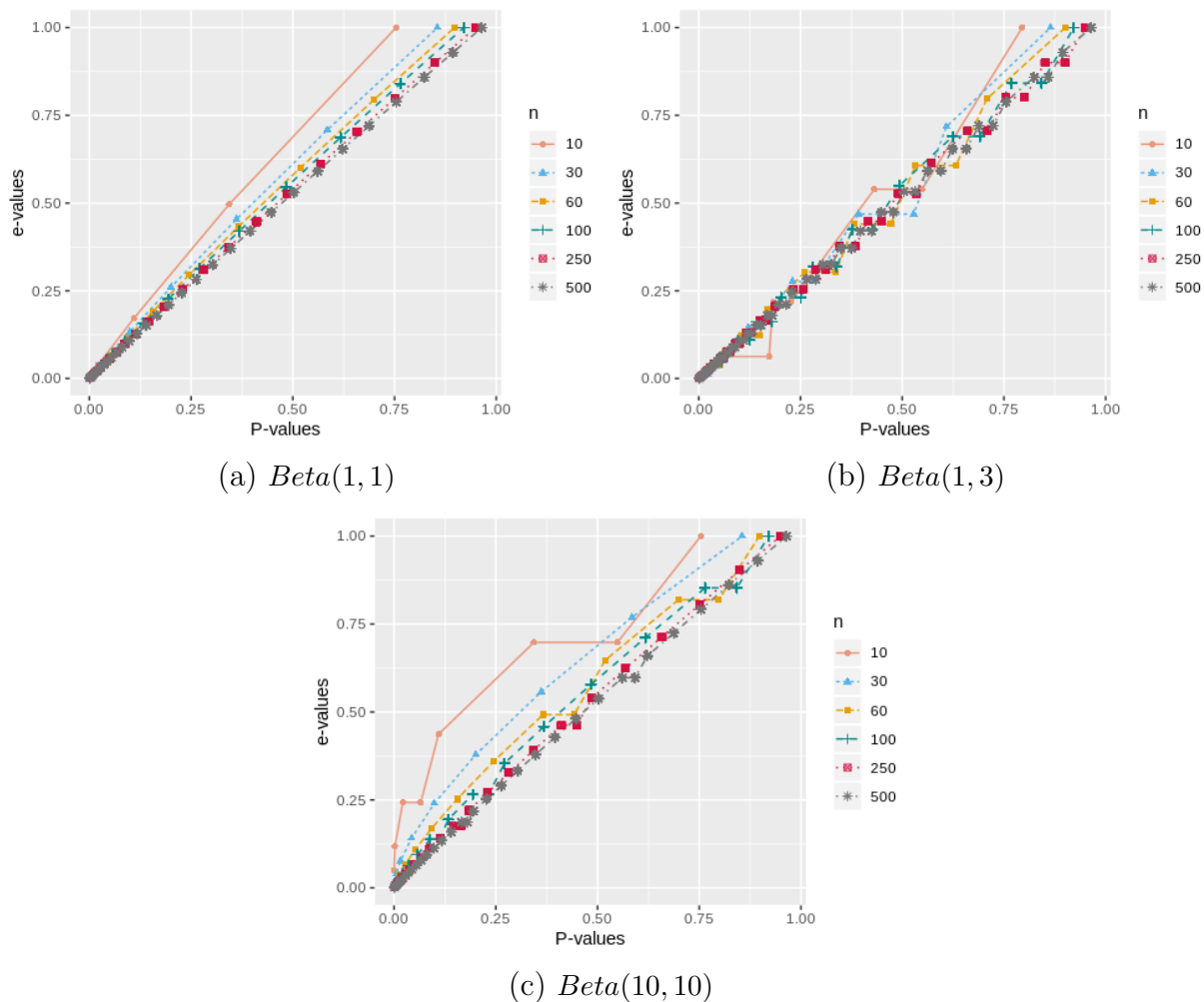


Figure 3.16: Relationship between e -value and P -value as function of n in proportion hypothesis testing, with $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$ and $\theta \sim \text{Beta}(10, 10)$.

These samples, that were simulated under \mathbf{H} , were also used to build the histograms in Figure 3.17 for a single sample size $n = 60$. Here is difficult to see a particular distribution for the simulated observations which may occur due to the discrete nature of X^* .

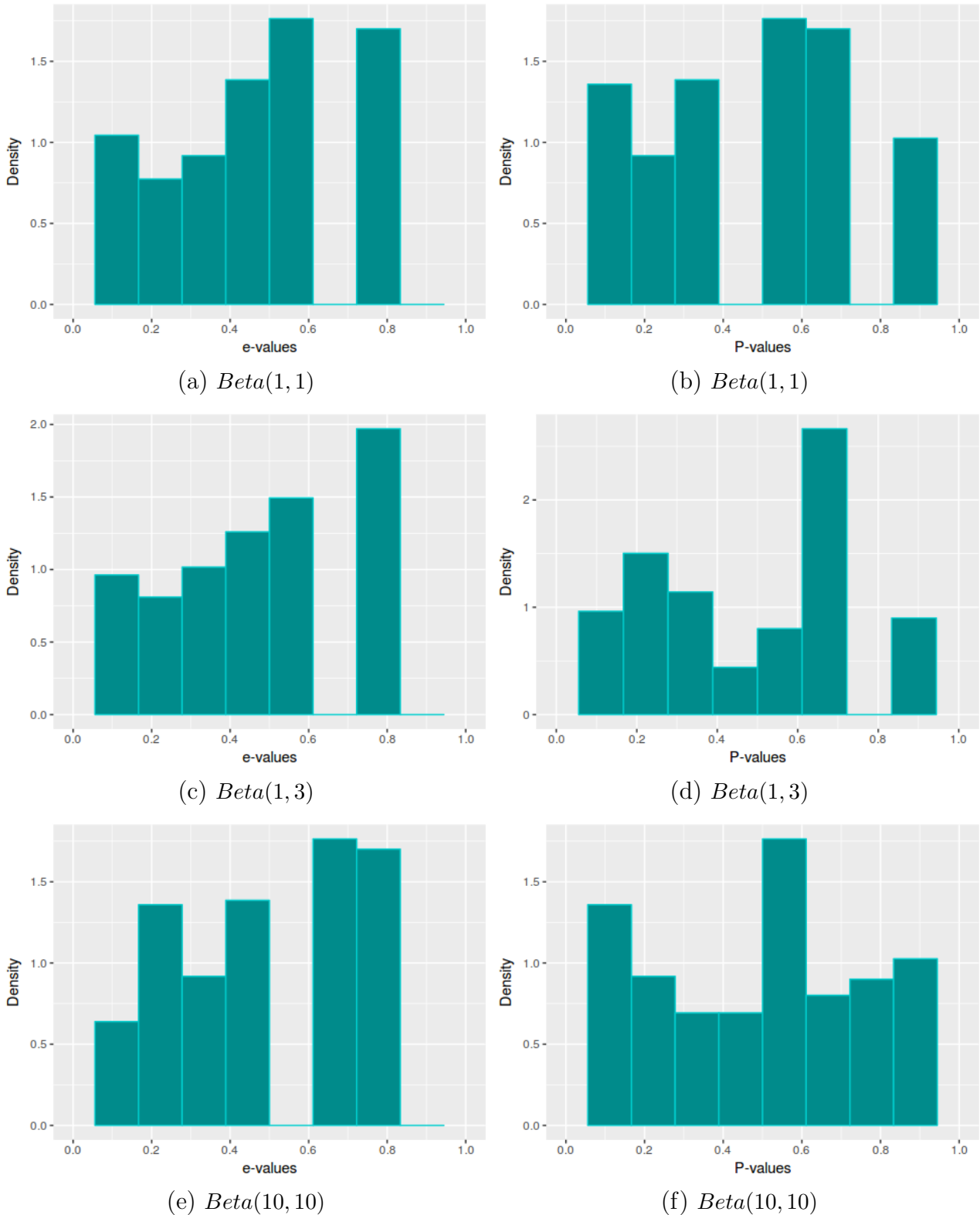


Figure 3.17: *e-value and P-value densities under \mathbf{H} in proportion hypothesis testing, with $\theta \sim Beta(1, 1)$, $\theta \sim Beta(1, 3)$ and $\theta \sim Beta(10, 10)$. Sample size $n = 60$.*

Tables 3.7, 3.8 and 3.9 summarize the optimal averaged error probabilities from e -value and P -value as well as the cutoff values k^* as functions of n for the three priors. It can be seen that the values corresponding to both kinds of errors are exactly the same in all cases, which was already evident comparing Figures 3.13 and 3.15.

n	$\alpha_{\varphi_P^*}$	$\alpha_{\varphi_e^*}^*$	k_{inf}^*	\bar{k}^*	k_{sup}^*	$\beta_{\varphi_P^*}$	$\beta_{\varphi_e^*}^*$
10	0.10938	0.10938	0.03910	0.10530	0.17150	0.45455	0.45454
20	0.11532	0.11532	0.06080	0.11115	0.16150	0.33333	0.33333
30	0.09874	0.09874	0.05930	0.09585	0.13240	0.29032	0.29032
40	0.08069	0.08069	0.05170	0.07860	0.10550	0.26829	0.26829
50	0.06491	0.06491	0.04320	0.06340	0.08360	0.25490	0.25490
60	0.05189	0.05189	0.03550	0.05075	0.06600	0.24590	0.24590
70	0.04139	0.04139	0.02890	0.04055	0.05220	0.23944	0.23943
80	0.05666	0.05666	0.04140	0.05570	0.07000	0.20988	0.20987
90	0.04460	0.04460	0.03290	0.04390	0.05490	0.20879	0.20879
100	0.03520	0.03520	0.02610	0.03465	0.04320	0.20792	0.20792
150	0.02715	0.02715	0.02130	0.02685	0.03240	0.17881	0.17881
200	0.02813	0.02813	0.02290	0.02790	0.03290	0.15423	0.15423
250	0.02666	0.02666	0.02220	0.02645	0.03070	0.13944	0.13944
300	0.02418	0.02418	0.02050	0.02405	0.02760	0.12957	0.12957

Table 3.7: Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim \text{Beta}(1, 1)$.

n	$\alpha_{\varphi_P^*}$	$\alpha_{\varphi_e^*}^*$	k_{inf}^*	\bar{k}^*	k_{sup}^*	$\beta_{\varphi_P^*}$	$\beta_{\varphi_e^*}^*$
10	0.18262	0.18262	0.06270	0.14015	0.21760	0.27972	0.27972
20	0.06357	0.06357	0.02540	0.05025	0.07510	0.29644	0.29644
30	0.05743	0.05743	0.02830	0.04835	0.06840	0.25202	0.25201
40	0.04864	0.04864	0.02660	0.04220	0.05780	0.22859	0.22859
50	0.04013	0.04013	0.02340	0.03550	0.04760	0.21408	0.21408
60	0.03269	0.03269	0.01990	0.02925	0.03860	0.20420	0.20420
70	0.04735	0.04735	0.03130	0.04350	0.05570	0.17548	0.17548
80	0.03749	0.03749	0.02520	0.03460	0.04400	0.17267	0.17267
90	0.02973	0.02973	0.02020	0.02750	0.03480	0.17046	0.17046
100	0.02362	0.02362	0.01620	0.02190	0.02760	0.16868	0.16868
150	0.01914	0.01914	0.01420	0.01815	0.02210	0.14295	0.14295
200	0.02063	0.02063	0.01620	0.01985	0.02350	0.12240	0.12240
250	0.02007	0.02007	0.01620	0.01945	0.02270	0.11004	0.11004
300	0.01855	0.01855	0.01530	0.01805	0.02080	0.10177	0.10177

Table 3.8: Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim \text{Beta}(1, 3)$.

Finally, we choose a single observation $\bar{x}_0^* = 0.6$ to calculate the e -value and P -value for the three priors, all with the purpose of comparing the decision that can be made by performing the test with each of these measures. The cells in **boldface** represent the cases when \mathbf{H} is rejected. Tables 3.10 and 3.11 display these values considering the minimum k^*

n	$\alpha_{\varphi_P}^*$	$\alpha_{\varphi_e}^*$	k_{inf}^*	\bar{k}^*	k_{sup}^*	$\beta_{\varphi_P}^*$	$\beta_{\varphi_e}^*$
10	0.34375	0.34375	0.43720	0.56755	0.69790	0.55742	0.55742
20	0.26318	0.26318	0.32010	0.41395	0.50780	0.56812	0.56812
30	0.20049	0.20049	0.24000	0.30925	0.37850	0.57884	0.57884
40	0.15386	0.15386	0.18260	0.23460	0.28660	0.58669	0.58669
50	0.20264	0.20264	0.21950	0.27285	0.32620	0.50071	0.50070
60	0.15500	0.15500	0.16940	0.21085	0.25230	0.51954	0.51953
70	0.18822	0.18822	0.19640	0.23910	0.28180	0.46140	0.46140
80	0.14564	0.14564	0.15350	0.18740	0.22130	0.48125	0.48124
90	0.17024	0.17024	0.17440	0.20940	0.24440	0.43772	0.43772
100	0.13321	0.13321	0.13770	0.16590	0.19410	0.45653	0.45653
150	0.12053	0.12053	0.12070	0.14165	0.16260	0.40205	0.40205
200	0.10364	0.10364	0.10250	0.11850	0.13450	0.37284	0.37284
250	0.08750	0.08750	0.08610	0.09855	0.11100	0.35465	0.35464
300	0.09390	0.09390	0.09150	0.10335	0.11520	0.32134	0.32133

Table 3.9: Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim \text{Beta}(10, 10)$.

and maximum k^* , respectively. The results show that the decision is the same whatever the index is used and also is the same for both k^* values, then, any $k^* \in [k_{\text{inf}}^*, k_{\text{sup}}^*]$ could be taken in order to perform a hypothesis test. On the other hand, it is important to note that the decision changes with the prior. In the $\theta \sim \text{Beta}(10, 10)$ case, the null hypothesis is more rejected than in the $\theta \sim \text{Beta}(1, 1)$ case. This is because the posterior has smaller variance even though it is centered closer to $\mathbf{H} : \theta = 0.5$. In contrast, when $\theta \sim \text{Beta}(1, 3)$, the null hypothesis is less rejected, in the first place, because the prior differs with the information brought by \bar{x}_0^* , and furthermore, because the posterior is centered closer to $\mathbf{H} : \theta = 0.5$. Figure 3.18 illustrates these statements.

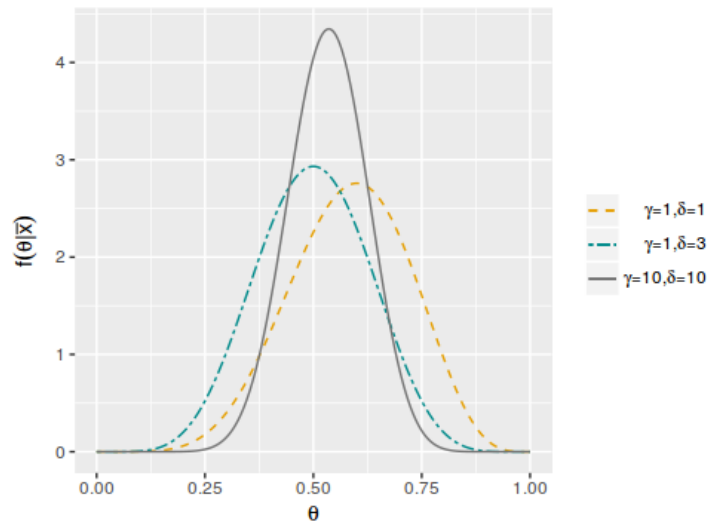


Figure 3.18: Posterior distributions with $\theta \sim \text{Beta}(1, 1)$, $\theta \sim \text{Beta}(1, 3)$, $\theta \sim \text{Beta}(10, 10)$, $\bar{x}_0^* = 0.6$ and $n = 10$

$\gamma = 1, \delta = 1, \bar{x}_0^* = 0.6$					$\gamma = 1, \delta = 3, \bar{x}_0^* = 0.6$				$\gamma = 10, \delta = 10, \bar{x}_0^* = 0.6$			
n	k_{inf}^*	ev	$\alpha_{\varphi_P^*}$	Pv	k_{inf}^*	ev	$\alpha_{\varphi_P^*}$	Pv	k_{inf}^*	ev	$\alpha_{\varphi_P^*}$	Pv
10	0.0391	0.4966	0.1094	0.5488	0.0627	1.0000	0.1826	0.7949	0.4372	0.6979	0.3438	0.5488
20	0.0608	0.3523	0.1153	0.2632	0.0254	0.6592	0.0636	0.5435	0.3201	0.5078	0.2632	0.2632
30	0.0593	0.2602	0.0987	0.2005	0.0283	0.4687	0.0574	0.5280	0.2400	0.3786	0.2005	0.2005
40	0.0517	0.1963	0.0807	0.1539	0.0266	0.3453	0.0486	0.2917	0.1826	0.2867	0.1539	0.1539
50	0.0432	0.1499	0.0649	0.1189	0.0234	0.2596	0.0401	0.2206	0.2195	0.2194	0.2026	0.1189
60	0.0355	0.1156	0.0519	0.0925	0.0199	0.1978	0.0327	0.1688	0.1694	0.1693	0.1550	0.0925
70	0.0289	0.0898	0.0414	0.0722	0.0313	0.1521	0.0474	0.1771	0.1964	0.1315	0.1882	0.0722
80	0.0414	0.0701	0.0567	0.0567	0.0252	0.1179	0.0375	0.1011	0.1535	0.1027	0.1456	0.0748
90	0.0329	0.0549	0.0446	0.0446	0.0202	0.0918	0.0297	0.1074	0.1744	0.0804	0.1702	0.0446
100	0.0261	0.0432	0.0352	0.0352	0.0162	0.0719	0.0236	0.0619	0.1377	0.0633	0.1332	0.0352
150	0.0213	0.0135	0.0272	0.0111	0.0142	0.0221	0.0191	0.0191	0.1207	0.0198	0.1205	0.0111
200	0.0229	0.0044	0.0281	0.0036	0.0162	0.0071	0.0206	0.0062	0.1025	0.0064	0.1036	0.0036
250	0.0222	0.0015	0.0267	0.0012	0.0162	0.0024	0.0201	0.0028	0.0861	0.0021	0.0875	0.0012
300	0.0205	0.0005	0.0242	0.0004	0.0153	0.0008	0.0185	0.0007	0.0915	0.0007	0.0939	0.0004

Table 3.10: Cutoff values k_{inf}^* , $ev(0.5; \bar{x}_0^*)$, $\alpha_{\varphi_P^*}$ and P -value (\bar{x}_0^*) as functions of n , with $\theta \sim Beta(1, 1)$, $\theta \sim Beta(1, 3)$ and $\theta \sim Beta(10, 10)$.

$\gamma = 1, \delta = 1, \bar{x}_0^* = 0.6$					$\gamma = 1, \delta = 3, \bar{x}_0^* = 0.6$				$\gamma = 10, \delta = 10, \bar{x}_0^* = 0.6$			
n	k_{sup}^*	ev	$\alpha_{\varphi_P^*}$	Pv	k_{sup}^*	ev	$\alpha_{\varphi_P^*}$	Pv	k_{sup}^*	ev	$\alpha_{\varphi_P^*}$	Pv
10	0.1715	0.4966	0.1094	0.5488	0.2176	1.0000	0.1826	0.7949	0.6979	0.6979	0.3438	0.5488
20	0.1615	0.3523	0.1153	0.2632	0.0751	0.6592	0.0636	0.5435	0.5078	0.5078	0.2632	0.2632
30	0.1324	0.2602	0.0987	0.2005	0.0684	0.4687	0.0574	0.5280	0.3785	0.3786	0.2005	0.2005
40	0.1055	0.1963	0.0807	0.1539	0.0578	0.3453	0.0486	0.2917	0.2866	0.2867	0.1539	0.1539
50	0.0836	0.1499	0.0649	0.1189	0.0476	0.2596	0.0401	0.2206	0.3262	0.2194	0.2026	0.1189
60	0.0660	0.1156	0.0519	0.0925	0.0386	0.1978	0.0327	0.1688	0.2523	0.1693	0.1550	0.0925
70	0.0522	0.0898	0.0414	0.0722	0.0557	0.1521	0.0474	0.1771	0.2818	0.1315	0.1882	0.0722
80	0.0700	0.0701	0.0567	0.0567	0.0440	0.1179	0.0375	0.1011	0.2213	0.1027	0.1456	0.0748
90	0.0549	0.0549	0.0446	0.0446	0.0348	0.0918	0.0297	0.1074	0.2444	0.0804	0.1702	0.0446
100	0.0432	0.0432	0.0352	0.0352	0.0276	0.0719	0.0236	0.0619	0.1941	0.0633	0.1332	0.0352
150	0.0324	0.0135	0.0272	0.0111	0.0221	0.0221	0.0191	0.0191	0.1626	0.0198	0.1205	0.0111
200	0.0329	0.0044	0.0281	0.0036	0.0235	0.0071	0.0206	0.0062	0.1345	0.0064	0.1036	0.0036
250	0.0307	0.0015	0.0267	0.0012	0.0227	0.0024	0.0201	0.0028	0.1110	0.0021	0.0875	0.0012
300	0.0276	0.0005	0.0242	0.0004	0.0208	0.0008	0.0185	0.0007	0.1152	0.0007	0.0939	0.0004

Table 3.11: Cutoff values k_{sup}^* , $ev(0.5; \bar{x}_0^*)$, $\alpha_{\varphi_P^*}$ and P -value (\bar{x}_0^*) as functions of n , with $\theta \sim Beta(1, 1)$, $\theta \sim Beta(1, 3)$ and $\theta \sim Beta(10, 10)$.

3.2 Adaptive significance levels in normal mean hypothesis testing

Consider that X_1, \dots, X_n are *c.i.i.d* $Normal(\theta, \sigma^2)$ given $\theta \in \mathbb{R}$ ($\sigma^2 > 0$ known), and define $X = (X_1, \dots, X_n)$. Let $\bar{X} = \sum_{i=1}^n X_i/n$ be a sufficient statistic for θ , then, $\bar{X}|\theta \sim Normal(\theta, \sigma^2/n)$. Suppose also that $\theta \sim Normal(m, v^2)$ ($m \in \mathbb{R}$ and $v^2 > 0$). Thus, the

posterior distribution of θ given that $\bar{X} = \bar{x}$ can be calculated as

$$\begin{aligned}
 f(\theta|x) &= f(\theta|\bar{x}) \propto f(X = x|\theta) g(\theta) \\
 &\propto \exp \left\{ -\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2} \right\} \exp \left\{ -\frac{(\theta - m)^2}{2v^2} \right\} \\
 &\propto \exp \left\{ -\frac{1}{2} \left(\theta^2 \left[\frac{n}{\sigma^2} + \frac{1}{v^2} \right] - 2\theta \left[\frac{n\bar{x}}{\sigma^2} + \frac{m}{v^2} \right] \right) \right\} \\
 &\propto \exp \left\{ -\frac{1}{2} \left[\frac{n}{\sigma^2} + \frac{1}{v^2} \right] \left(\theta^2 - 2\theta \frac{\left[\frac{n\bar{x}}{\sigma^2} + \frac{m}{v^2} \right]}{\left[\frac{n}{\sigma^2} + \frac{1}{v^2} \right]} + \left[\frac{\frac{n\bar{x}}{\sigma^2} + \frac{m}{v^2}}{\left[\frac{n}{\sigma^2} + \frac{1}{v^2} \right]} \right]^2 \right) \right\} \\
 &\propto \exp \left\{ -\frac{1}{2 \left(\frac{\sigma^2 v^2}{\sigma^2 + nv^2} \right)} \left(\theta - \frac{\sigma^2 m + nv^2 \bar{x}}{\sigma^2 + nv^2} \right)^2 \right\}, \theta \in \mathbb{R}, \sigma^2 > 0, m \in \mathbb{R}, v^2 > 0.
 \end{aligned}$$

Therefore, the posterior distribution is

$$\theta|\bar{x} \sim Normal \left(\frac{\sigma^2 m + nv^2 \bar{x}}{\sigma^2 + nv^2}, \frac{\sigma^2 v^2}{\sigma^2 + nv^2} \right).$$

The hypotheses to be tested are given by

$$\mathbf{H} : \theta = \theta_0$$

$$\mathbf{A} : \theta \neq \theta_0.$$

3.2.1 Evidence index: e -value

With $\Theta_{\mathbf{H}} = \{\theta_0\}$, the tangential set to the null hypothesis and the evidence in favor of \mathbf{H} are respectively

$$T_{\bar{x}_0} = \left\{ \theta \in \Theta : f(\theta|\bar{x}_0) > \sup_{\mathbf{H}} f(\theta|\bar{x}_0) \right\} = \left\{ \theta \in \Theta : f(\theta|\bar{x}_0) > f(\theta_0|\bar{x}_0) \right\},$$

$$ev(\mathbf{H}; \bar{x}_0) = 1 - P(\theta \in T_{\bar{x}_0} | \bar{x}_0).$$

Let $\varphi_e(\bar{x})$ be the test given by

$$\varphi_e(\bar{x}) = \begin{cases} 0 & \text{if } ev(\mathbf{H}; \bar{x}) > k \\ 1 & \text{if } ev(\mathbf{H}; \bar{x}) \leq k. \end{cases}$$

Let denote by M the mean of the posterior distribution $\theta|\bar{x}$. Figure 3.19 show the representation of the tangential set to the null hypothesis when $M > \theta_0$ and $M < \theta_0$.

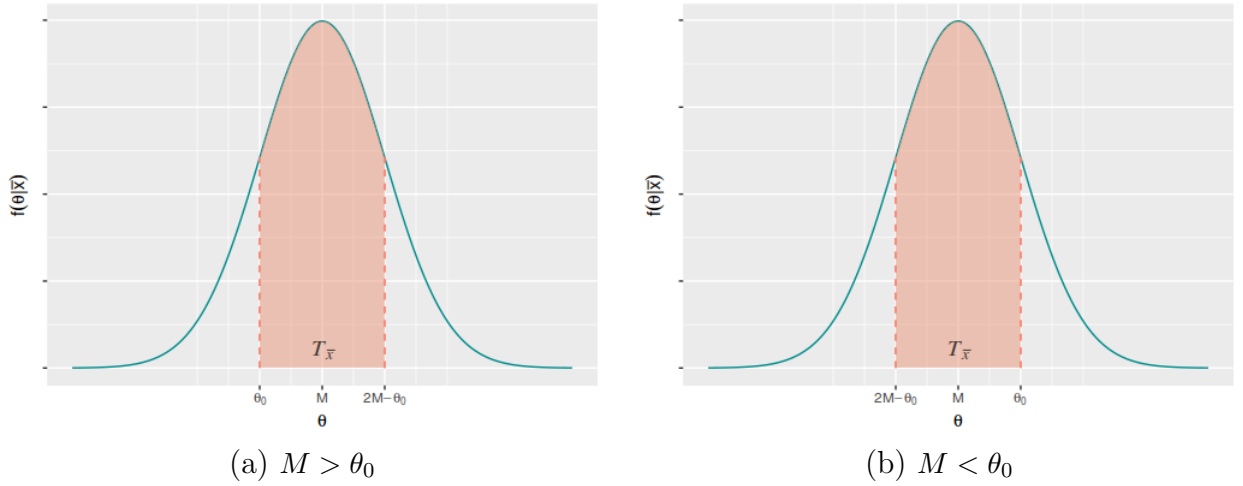


Figure 3.19: Tangential set to $\mathbf{H} : \theta = \theta_0$.

Because of the normal distribution symmetry, the evidence in favor of \mathbf{H} when $M > \theta_0$ and $M < \theta_0$ is equivalent, and can be expressed as

$$\begin{aligned} ev(\mathbf{H}; \bar{x}_0) &= 1 - P(\theta \in T_{\bar{x}_0} | \bar{x}_0) \\ &= 1 - P\left(\theta_0 \leq \theta \leq 2\left(\frac{\sigma^2 m + nv^2 \bar{x}_0}{\sigma^2 + nv^2}\right) - \theta_0 \mid \bar{x}_0\right) \\ &= 1 - P\left(\frac{\theta_0 - \left(\frac{\sigma^2 m + nv^2 \bar{x}_0}{\sigma^2 + nv^2}\right)}{\frac{\sigma v}{\sqrt{\sigma^2 + nv^2}}} \leq Z \leq -\frac{\left(\theta_0 - \left(\frac{\sigma^2 m + nv^2 \bar{x}_0}{\sigma^2 + nv^2}\right)\right)}{\frac{\sigma v}{\sqrt{\sigma^2 + nv^2}}} \mid \bar{x}_0\right) \\ &= 2\Phi\left(\frac{\left|\theta_0 - \left(\frac{\sigma^2 m + nv^2 \bar{x}_0}{\sigma^2 + nv^2}\right)\right|}{\frac{\sigma v}{\sqrt{\sigma^2 + nv^2}}}\right) \\ &= 2\Phi\left(-\left|\frac{\sigma(\theta_0 - m)}{v\sqrt{\sigma^2 + nv^2}} + \frac{nv(\theta_0 - \bar{x}_0)}{\sigma\sqrt{\sigma^2 + nv^2}}\right|\right) \end{aligned}$$

$$= 2\Phi\left(-\frac{|\sigma^2(\theta_0 - m) + nv^2(\theta_0 - \bar{x}_0)|}{\sigma v\sqrt{\sigma^2 + nv^2}}\right), \quad (3.1)$$

where Φ is the standard normal cumulative distribution function.

Hence, the power function and the averaged error probabilities can be calculated as follows

$$\begin{aligned} \pi_{\varphi_e}(\theta) &= P(\{\bar{X} \in \Omega^* : \varphi_e(\bar{X}) = 1\}|\theta) \\ &= P(\{\bar{X} \in \Omega^* : ev(\mathbf{H}; \bar{X}) \leq k\}|\theta) \\ &= P\left(2\Phi\left(-\frac{|\sigma^2(\theta_0 - m) + nv^2(\theta_0 - \bar{X})|}{\sigma v\sqrt{\sigma^2 + nv^2}}\right) \leq k \mid \theta\right) \\ &= P\left(-|\sigma^2(\theta_0 - m) + nv^2(\theta_0 - \bar{X})| \leq \sigma v\sqrt{\sigma^2 + nv^2} \Phi^{-1}\left(\frac{k}{2}\right) \mid \theta\right) \\ &= P\left(-\left|\frac{\sigma^2(\theta_0 - m)}{nv^2\sigma/\sqrt{n}} + \frac{(\theta_0 - \bar{X})}{\sigma/\sqrt{n}}\right| \leq \frac{\sigma v\sqrt{\sigma^2 + nv^2} \Phi^{-1}\left(\frac{k}{2}\right)}{nv^2\sigma/\sqrt{n}} \mid \theta\right) \\ &= P\left(\left|\frac{\sigma(\theta_0 - m)}{\sqrt{n}v^2} - \frac{(\bar{X} - \theta)}{\sigma/\sqrt{n}} - \frac{(\theta - \theta_0)}{\sigma/\sqrt{n}}\right| \geq -\frac{\sqrt{\sigma^2 + nv^2} \Phi^{-1}\left(\frac{k}{2}\right)}{\sqrt{n}v} \mid \theta\right) \\ &= 1 - P\left(\left|\frac{\sigma(\theta_0 - m)}{\sqrt{n}v^2} - Z - \frac{(\theta - \theta_0)}{\sigma/\sqrt{n}}\right| \leq \underbrace{-\frac{\sqrt{\sigma^2 + nv^2} \Phi^{-1}\left(\frac{k}{2}\right)}{\sqrt{n}v}}_Q \mid \theta\right) \\ &= 1 - P\left(Q \leq -\frac{\sigma(\theta_0 - m)}{\sqrt{n}v^2} + Z + \frac{(\theta - \theta_0)}{\sigma/\sqrt{n}} \leq -Q \mid \theta\right) \\ &= 1 - P\left(\underbrace{Q + \frac{\sigma(\theta_0 - m)}{\sqrt{n}v^2} - \frac{(\theta - \theta_0)}{\sigma/\sqrt{n}}}_{z_1(\theta)} \leq Z \leq \underbrace{-Q + \frac{\sigma(\theta_0 - m)}{\sqrt{n}v^2} - \frac{(\theta - \theta_0)}{\sigma/\sqrt{n}}}_{z_2(\theta)} \mid \theta\right) \\ &= 1 - [\Phi(z_2(\theta)) - \Phi(z_1(\theta))]. \end{aligned}$$

where Ω^* is the space induced by the sufficient statistic.

$$\begin{aligned}
\alpha_{\varphi_e} &= P(\{\bar{X} \in \Omega^* : \varphi_e(\bar{X}) = 1\} | \mathbf{H}) \\
&= P(\{\bar{X} \in \Omega^* : ev(\mathbf{H}; \bar{X}) \leq k\} | \mathbf{H}) \\
&= 1 - P\left(\underbrace{\frac{\sqrt{\sigma^2 + nv^2} \Phi^{-1}\left(\frac{k}{2}\right)}{\sqrt{n}v} + \frac{\sigma(\theta_0 - m)}{\sqrt{n}v^2}}_{z_1^*(\theta)} \leq Z \leq \underbrace{-\frac{\sqrt{\sigma^2 + nv^2} \Phi^{-1}\left(\frac{k}{2}\right)}{\sqrt{n}v} + \frac{\sigma(\theta_0 - m)}{\sqrt{n}v^2}}_{z_2^*(\theta)}\right) \\
&= 1 - [\Phi(z_2^*(\theta)) - \Phi(z_1^*(\theta)) | \theta = \theta_0].
\end{aligned}$$

$$\begin{aligned}
\beta_{\varphi_e} &= E_{\theta} [1 - \pi_{\varphi_e}(\theta) | \mathbf{A}] \\
&= E_{\theta} [\Phi(z_2(\theta)) - \Phi(z_1(\theta)) | \mathbf{A}] \\
&= \int_{\mathbf{A}} [\Phi(z_2(\theta)) - \Phi(z_1(\theta))] g(\theta) d\theta \\
&= \int_{\mathbf{A}} [\Phi(z_2(\theta)) - \Phi(z_1(\theta))] \frac{1}{\sqrt{2\pi}v^2} \exp\left\{-\frac{(\theta - m)^2}{2v^2}\right\} d\theta.
\end{aligned}$$

The averaged type-II error probability can be approximated by Monte Carlo methods (Robert and Casella, 2005).

Taking the particular case in which $\sigma^2 = 10$ and $m = 0$, then $\bar{X} | \theta \sim Normal(\theta, 10/n)$ and the prior will be $\theta \sim Normal(0, v^2)$, therefore $\theta | \bar{x} \sim Normal\left(\frac{nv^2\bar{x}}{10+nv^2}, \frac{10v^2}{10+nv^2}\right)$. Take also the hypotheses

$$\mathbf{H} : \theta = 0$$

$$\mathbf{A} : \theta \neq 0.$$

Consider three prior distributions for θ , two that can be seen as informative priors, $\theta \sim Normal(0, 0.1)$ and $\theta \sim Normal(0, 1)$, the first one providing more information than the second one, and a “vague” prior, $\theta \sim Normal(0, 100)$ (Figure 3.20).

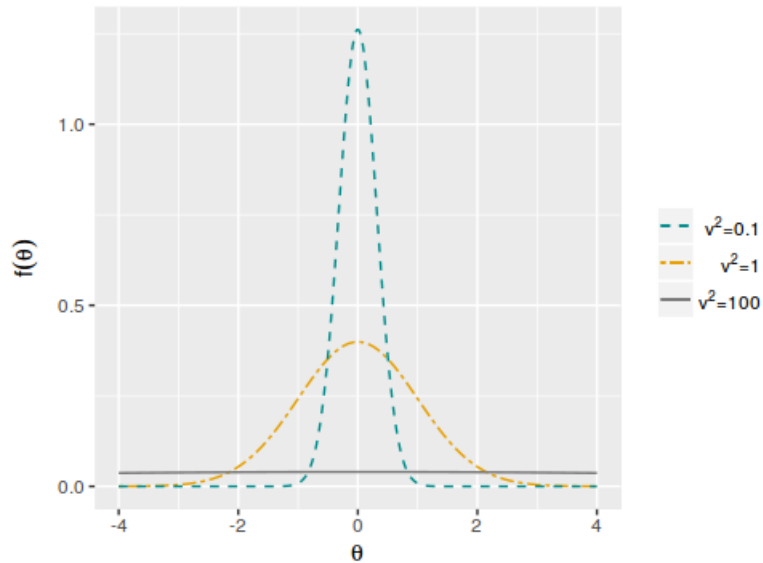
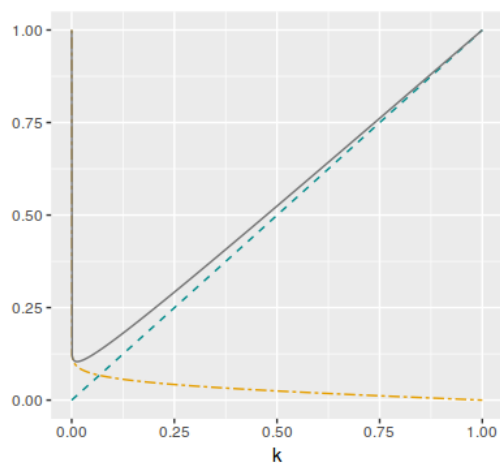


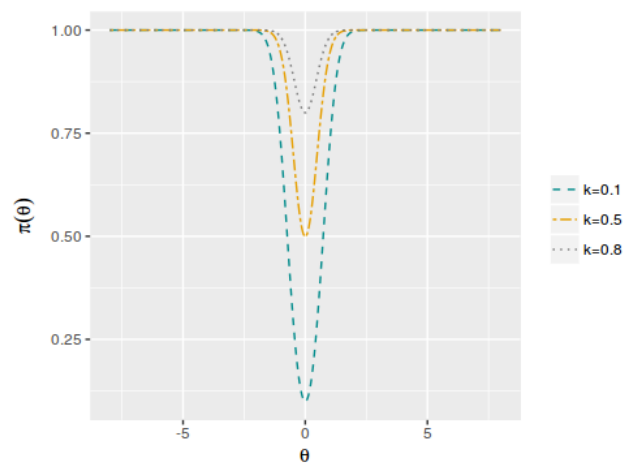
Figure 3.20: Prior distributions $\theta \sim \text{Normal}(0, 0.1)$, $\theta \sim \text{Normal}(0, 1)$ and $\theta \sim \text{Normal}(0, 100)$.

In Figure 3.21 for the three priors and a sample size $n = 50$ the averaged error probabilities as a function of k (Figures 3.21.(a), 3.21.(c), 3.21.(e)), as well as the power function for a fixed k (Figures 3.21.(b), 3.21.(d), 3.21.(f)) are plotted.

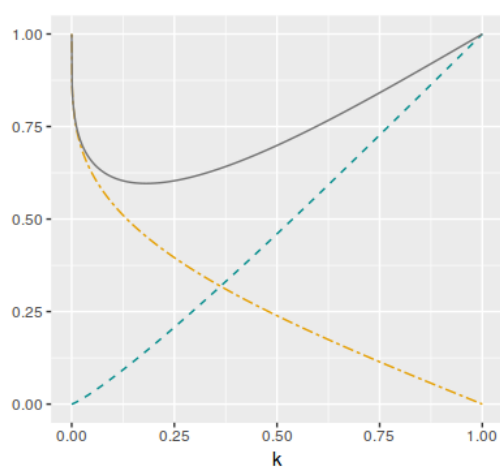
Again, the procedure previously described to find the k^* values is replicated to different sample sizes and considering each prior distribution. In Figures 3.22.(a), 3.22.(c) and 3.22.(e) the results are presented. In this example, where the distribution of X is continuous, the continuity of the curve that represents the relation between k^* and n can be noted, making more evident that by increasing the sample size, the k^* value decreases until approaching zero. Figures 3.22.(b), 3.22.(d) and 3.22.(f) show the optimal averaged error probabilities ($\alpha_{\varphi_e}^*$, $\beta_{\varphi_e}^*$ and $\alpha_{\varphi_e}^* + \beta_{\varphi_e}^*$) as functions of n . Once more, by increasing the sample size, the probabilities of both kind of errors and their linear combination decrease.



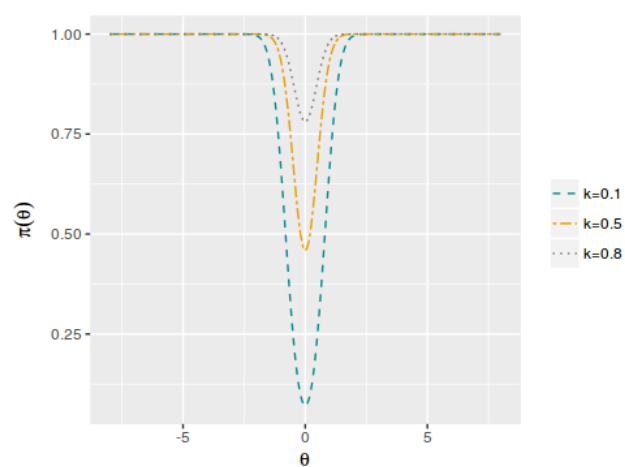
(a) $Normal(0, 100)$



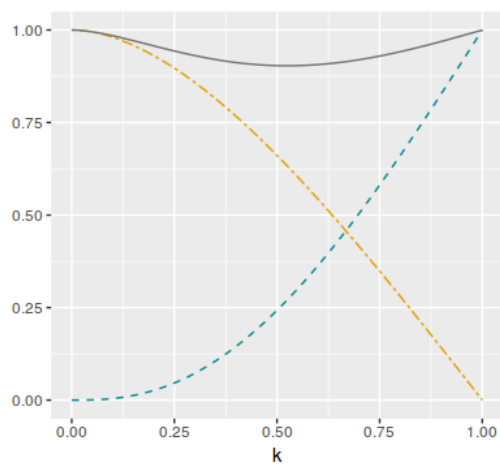
(b) $Normal(0, 100)$



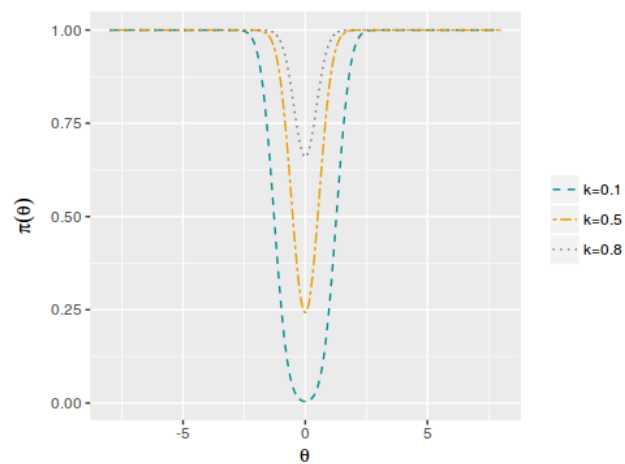
(c) $Normal(0, 1)$



(d) $Normal(0, 1)$



(e) $Normal(0, 0.1)$



(f) $Normal(0, 0.1)$

Figure 3.21: (a), (c) and (e) Averaged error probabilities (α_{φ_e} , β_{φ_e} and $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k ; (b), (d) and (f) Power function when $k = 0.1$, $k = 0.5$ and $k = 0.8$. Sample size $n = 50$.

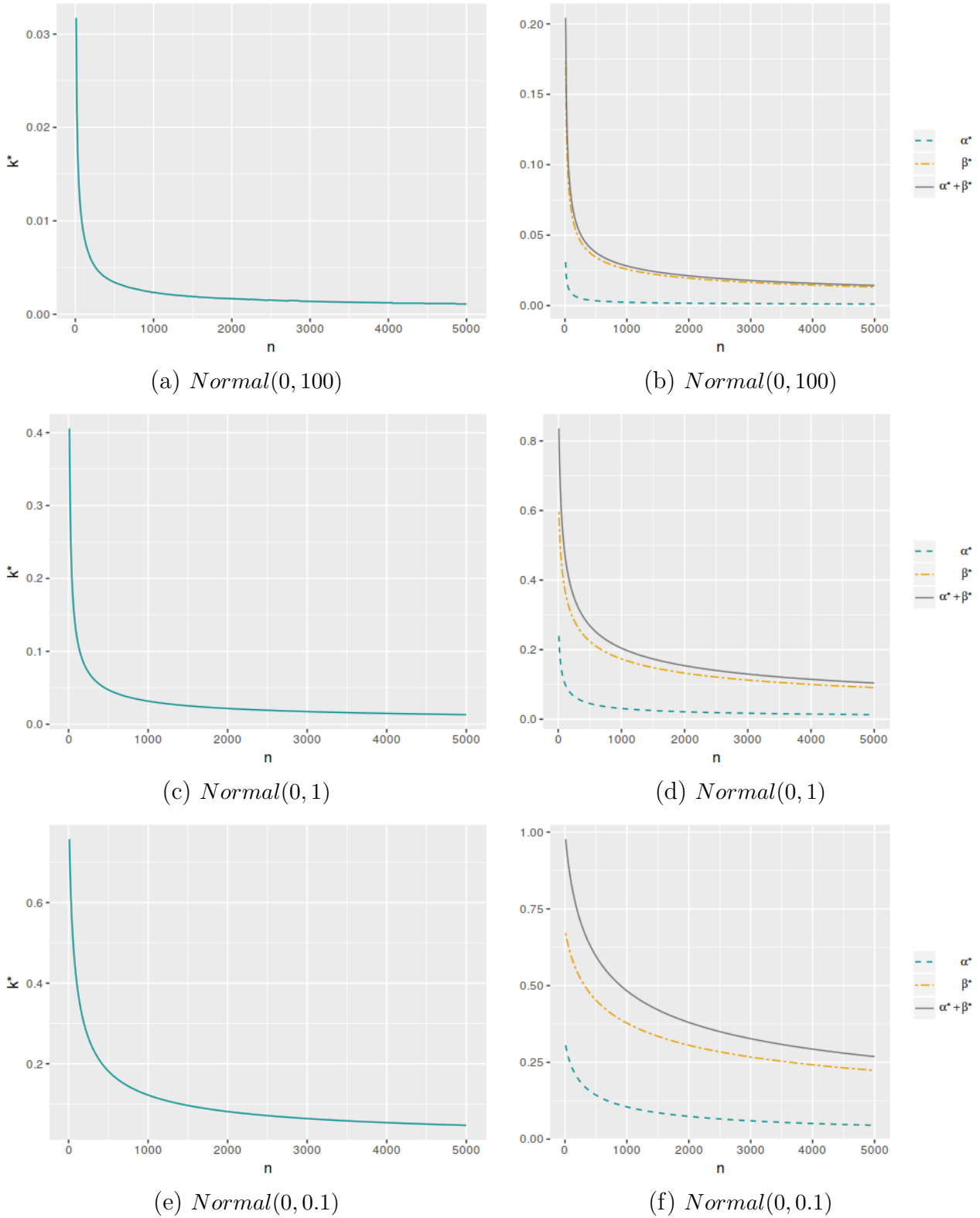


Figure 3.22: (a), (c) and (e) Cutoff values k^* for $ev(0; \bar{x}_0)$ as function of n ; (b), (d) and (f) Optimal averaged error probabilities $(\alpha_{\varphi_e^*}^*, \beta_{\varphi_e^*}^*$ and $\alpha_{\varphi_e^*}^* + \beta_{\varphi_e^*}^*)$ as function of n .

Finally, Table 3.12 and Figure 3.23 display the comparisons of k^* as a function of the sample size for each of the predetermined priors. It is possible to perceive the differences in the results between priors, and therefore, to identify their relevance at the moment of setting the cutoff value for $ev(\mathbf{H}; \bar{x}_0)$. It can also be observed that, insofar as the prior is less informative, the k^* value is smaller.

n	k^*		
	$v^2 = 100$	$v^2 = 1$	$v^2 = 0.1$
10	0.03174	0.40574	0.75754
50	0.01320	0.18178	0.52456
100	0.00892	0.12234	0.40574
150	0.00705	0.09651	0.33928
200	0.00588	0.08142	0.29552
250	0.00513	0.07131	0.26404
300	0.00459	0.06398	0.24006
350	0.00417	0.05838	0.22113
400	0.00387	0.05395	0.20565
450	0.00362	0.05033	0.19274
500	0.00342	0.04732	0.18178
1000	0.00233	0.03174	0.12234
1500	0.00188	0.02533	0.09651
2000	0.00166	0.02168	0.08142

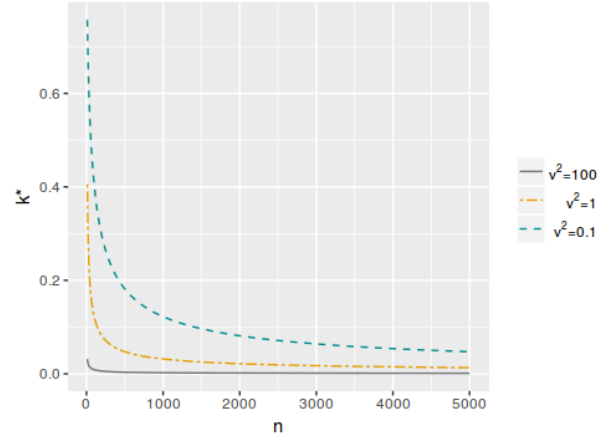


Figure 3.23: Cutoff values k^* for $ev(0; \bar{x}_0)$ as function of n , with $\theta \sim \text{Normal}(0, 100)$, $\theta \sim \text{Normal}(0, 1)$ and $\theta \sim \text{Normal}(0, 0.1)$.

Table 3.12: Cutoff values k^* for $ev(0; \bar{x}_0)$ as function of n , with $\theta \sim \text{Normal}(0, 100)$, $\theta \sim \text{Normal}(0, 1)$ and $\theta \sim \text{Normal}(0, 0.1)$.

3.2.2 Significance index: P -value

Taking the same particular case where $\sigma^2 = 10$, $m = 0$ and also the hypotheses

$$\mathbf{H} : \theta = 0$$

$$\mathbf{A} : \theta \neq 0$$

want to be tested, the prior predictive distributions are $\bar{X}|\mathbf{H} \sim \text{Normal}(0, \sigma^2/n)$ and $\bar{X}|\mathbf{A} \sim \text{Normal}\left(0, \frac{\sigma^2 + nv^2}{n}\right)$ (see Appendix A). Thus,

$$\begin{aligned} f_{\mathbf{H}}(\bar{x}) &= \int_{\mathbf{H}} f(\bar{x}|\theta) g_{\mathbf{H}}(\theta) d\theta \\ &= f(\bar{x}|\theta = 0) \end{aligned}$$

$$= \frac{\sqrt{n}}{\sqrt{2\sigma^2\pi}} \exp \left\{ -\frac{n\bar{x}^2}{2\sigma^2} \right\}.$$

and

$$\begin{aligned} f_{\mathbf{A}}(\bar{x}) &= \int_{\mathbf{A}} f(\bar{x}|\theta) g_{\mathbf{A}}(\theta) d\theta \\ &= \int_{\mathbf{A}} f(\bar{x}|\theta) g(\theta) d\theta \\ &= f(\bar{x}) \\ &= \frac{\sqrt{n}}{\sqrt{2\pi(\sigma^2 + nv^2)}} \exp \left\{ -\frac{n\bar{x}^2}{2(\sigma^2 + nv^2)} \right\}. \end{aligned}$$

Let $\varphi_P^*(x^*)$ be the test given by

$$\varphi_P^*(\bar{x}) = \begin{cases} 0 & \text{if } \text{BF}(\bar{x}) > 1 \\ 1 & \text{if } \text{BF}(\bar{x}) \leq 1. \end{cases}$$

As stated in (2.6), the P -value is calculated as

$$P\text{-value}(\bar{x}_0) = P(\{\bar{X} \in \Omega^* : \text{BF}(\bar{X}) \leq \text{BF}(\bar{x}_0)\} | \mathbf{H}).$$

So, from the inequality

$$\begin{aligned} \text{BF}(\bar{x}) &\leq \text{BF}(\bar{x}_0) \\ \frac{f_{\mathbf{H}}(\bar{x})}{f_{\mathbf{A}}(\bar{x})} &\leq \frac{f_{\mathbf{H}}(\bar{x}_0)}{f_{\mathbf{A}}(\bar{x}_0)} \\ \frac{\frac{\sqrt{n}}{\sqrt{2\pi\sqrt{\sigma^2}}} \exp \left\{ -\frac{n\bar{x}^2}{2\sigma^2} \right\}}{\frac{\sqrt{n}}{\sqrt{2\pi\sqrt{\sigma^2 + nv^2}}} \exp \left\{ -\frac{n\bar{x}^2}{2(\sigma^2 + nv^2)} \right\}} &\leq \frac{\frac{\sqrt{n}}{\sqrt{2\pi\sqrt{\sigma^2}}} \exp \left\{ -\frac{n\bar{x}_0^2}{2\sigma^2} \right\}}{\frac{\sqrt{n}}{\sqrt{2\pi\sqrt{\sigma^2 + nv^2}}} \exp \left\{ -\frac{n\bar{x}_0^2}{2(\sigma^2 + nv^2)} \right\}} \\ \exp \left\{ -\frac{n\bar{x}^2}{2\sigma^2} + \frac{n\bar{x}^2}{2(\sigma^2 + nv^2)} \right\} &\leq \exp \left\{ -\frac{n\bar{x}_0^2}{2\sigma^2} + \frac{n\bar{x}_0^2}{2(\sigma^2 + nv^2)} \right\} \\ \frac{n\bar{x}^2}{2\sigma^2} - \frac{n\bar{x}^2}{2(\sigma^2 + nv^2)} &\geq \frac{n\bar{x}_0^2}{2\sigma^2} - \frac{n\bar{x}_0^2}{2(\sigma^2 + nv^2)} \\ \bar{x}^2 &\geq \bar{x}_0^2, \end{aligned}$$

an expression for the P -value can be specified by

$$\begin{aligned}
P\text{-value}(\bar{x}_0) &= 1 - P\left(|\bar{X}| \leq \sqrt{\bar{x}_0^2} \mid \mathbf{H}\right) \\
&= 1 - P\left(-\bar{x}_0 \leq \bar{X} \leq \bar{x}_0 \mid \mathbf{H}\right) \\
&= 1 - P\left(-\frac{\bar{x}_0}{\sqrt{\frac{\sigma^2}{n}}} \leq Z \leq \frac{\bar{x}_0}{\sqrt{\frac{\sigma^2}{n}}}\right) \\
&= 2\Phi\left(-\frac{\sqrt{n}|\bar{x}_0|}{\sigma}\right). \tag{3.2}
\end{aligned}$$

Considering the sample points \bar{x} with Bayes factors smaller than or equal to $b/a = 1$, i.e.,

$$\begin{aligned}
\text{BF}(\bar{x}) &\leq 1 \\
\frac{f_{\mathbf{H}}(\bar{x})}{f_{\mathbf{A}}(\bar{x})} &\leq 1 \\
\frac{\frac{\sqrt{n}}{\sqrt{2\pi}\sqrt{\sigma^2}} \exp\left\{-\frac{n\bar{x}^2}{2\sigma^2}\right\}}{\frac{\sqrt{n}}{\sqrt{2\pi}\sqrt{\sigma^2 + nv^2}} \exp\left\{-\frac{n\bar{x}^2}{2(\sigma^2 + nv^2)}\right\}} &\leq 1 \\
\exp\left\{-\frac{n\bar{x}^2}{2\sigma^2} + \frac{n\bar{x}^2}{2(\sigma^2 + nv^2)}\right\} &\leq \sqrt{\frac{\sigma^2}{\sigma^2 + nv^2}} \\
\frac{n\bar{x}^2}{\sigma^2} - \frac{n\bar{x}^2}{\sigma^2 + nv^2} &\geq -\log\left(\frac{\sigma^2}{\sigma^2 + nv^2}\right) \\
\bar{x}^2 &\geq -\log\left(\frac{\sigma^2}{\sigma^2 + nv^2}\right) \frac{\sigma^2(\sigma^2 + nv^2)}{n^2v^2},
\end{aligned}$$

the optimal averaged type I error probability can be calculated as follows

$$\begin{aligned}
\alpha_{\varphi_P^*} &= P\left(\{\bar{X} \in \Omega^* : \varphi_P^*(\bar{X}) = 1\} \mid \mathbf{H}\right) \\
&= P\left(\{\bar{X} \in \Omega^* : \text{BF}(\bar{X}) \leq 1\} \mid \mathbf{H}\right) \\
&= 1 - P\left(|\bar{X}| \leq \sqrt{-\log\left(\frac{\sigma^2}{\sigma^2 + nv^2}\right) \frac{\sigma^2(\sigma^2 + nv^2)}{n^2v^2}} \mid \theta = 0\right)
\end{aligned}$$

$$= 2 \Phi \left(-\frac{\sqrt{n}}{\sigma} \sqrt{-\log \left(\frac{\sigma^2}{\sigma^2 + nv^2} \right) \frac{\sigma^2 (\sigma^2 + nv^2)}{n^2 v^2}} \right).$$

Analogously, the optimal averaged type II error probability can be expressed as

$$\begin{aligned} \beta_{\varphi_P^*} &= P(\{\bar{X} \in \Omega^* : \varphi_P^*(\bar{X}) = 0\} | \mathbf{A}) \\ &= P(\{\bar{X} \in \Omega^* : \text{BF}(\bar{X}) > 1\} | \mathbf{A}) \\ &= P \left(|\bar{X}| \leq \sqrt{-\log \left(\frac{\sigma^2}{\sigma^2 + nv^2} \right) \frac{\sigma^2 (\sigma^2 + nv^2)}{n^2 v^2}} \mid \mathbf{A} \right) \\ &= 1 - 2 \Phi \left(-\frac{\sqrt{-\log \left(\frac{\sigma^2}{\sigma^2 + nv^2} \right) \frac{\sigma^2 (\sigma^2 + nv^2)}{n^2 v^2}}}{\sqrt{\frac{\sigma^2}{n} + v^2}} \right). \end{aligned}$$

The optimal adaptive significance level for each prior is shown in Table 3.13 and Figure 3.24. In this example the $\alpha_{\varphi_P^*}$ value presents the same behavior as in the first one, it is higher when the prior is more informative, i.e., when $\theta \sim \text{Normal}(0, 0.1)$.

n	$\alpha_{\varphi_P^*}$		
	$v^2 = 100$	$v^2 = 1$	$v^2 = 0.1$
10	0.03085	0.23903	0.30587
50	0.01257	0.14256	0.27007
100	0.00854	0.10436	0.23903
150	0.00682	0.08548	0.21654
200	0.00582	0.07378	0.19924
250	0.00515	0.06566	0.18539
300	0.00465	0.05960	0.17397
350	0.00428	0.05487	0.16434
400	0.00397	0.05106	0.15608
450	0.00372	0.04789	0.14889
500	0.00351	0.04522	0.14256
1000	0.00241	0.03085	0.10436
1500	0.00193	0.02462	0.08548
2000	0.00165	0.02096	0.07378

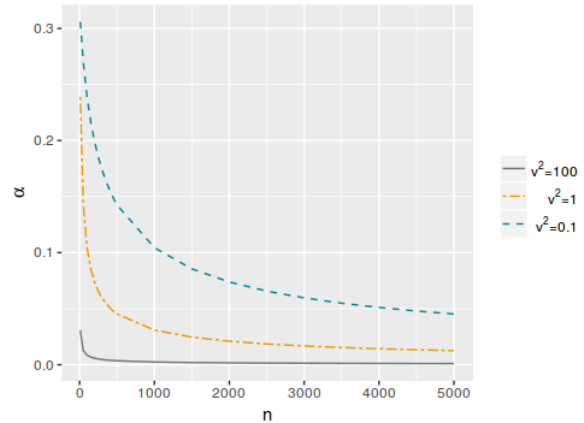


Figure 3.24: Optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as function of n , with $\theta \sim \text{Normal}(0, 100)$, $\theta \sim \text{Normal}(0, 1)$ and $\theta \sim \text{Normal}(0, 0.1)$.

Table 3.13: Optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as function of n , with $\theta \sim \text{Normal}(0, 100)$, $\theta \sim \text{Normal}(0, 1)$ and $\theta \sim \text{Normal}(0, 0.1)$.

A tendency to decrease when the sample size increases can be observed in Figure 3.25 where the optimal adaptive significance level $\alpha_{\varphi_P^*}$, the optimal adaptive type-II averaged error probability and the minimized linear combination $\alpha_{\varphi_P^*} + \beta_{\varphi_P^*}$ are plotted for each prior.

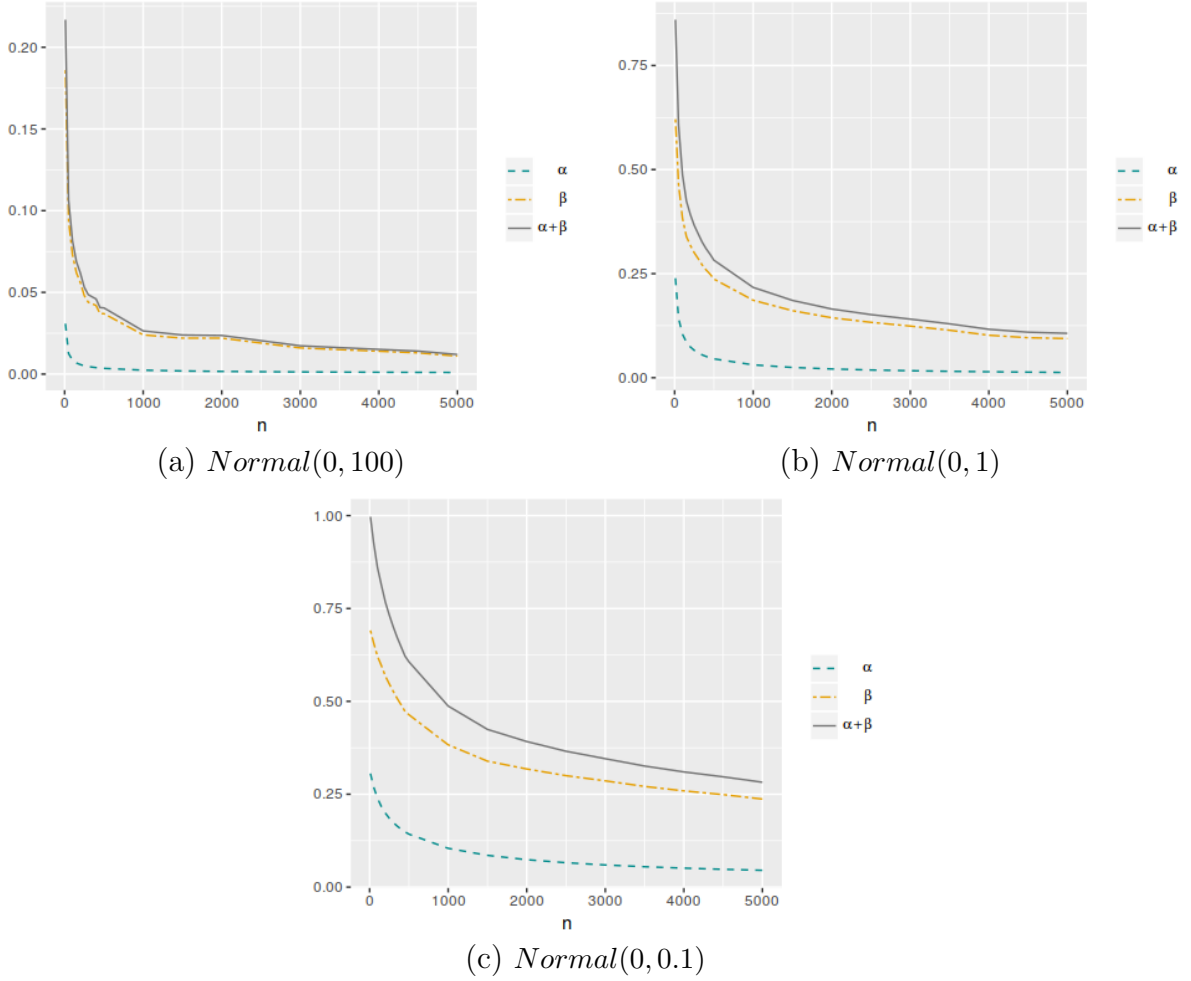


Figure 3.25: Optimal averaged error probabilities ($\alpha_{\varphi_P^*}$, $\beta_{\varphi_P^*}$ and $\alpha_{\varphi_P^*} + \beta_{\varphi_P^*}$) as function of n , with $\theta \sim Normal(0, 0.1)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 100)$.

3.2.3 Relationship between e -value and P -value

In order to establish a possible functional relationship between the e -value and the P -value, without loss of generality, consider the particular case in which $m = 0$ and $\theta_0 = 0$. Therefore, the e -value in (3.1) can be rewritten as

$$ev(0; \bar{x}_0) = 2 \Phi \left(- \frac{v}{\sqrt{\frac{\sigma^2}{n} + v^2}} \frac{\sqrt{n} |\bar{x}_0|}{\sigma} \right). \quad (3.3)$$

In that way, the e -value in (3.3) can be defined as a function of the P -value in (3.2)

$$ev(0; \bar{x}_0) = 2 \Phi \left(\frac{v}{\sqrt{\frac{\sigma^2}{n} + v^2}} \Phi^{-1} \left(\frac{P\text{-value}(\bar{x}_0)}{2} \right) \right). \quad (3.4)$$

Then the cutoff value k^* can also be denoted as a function of the optimal averaged type-I error probability $\alpha_{\varphi_P^*}$ as follows:

$$k_{\varphi_P^*}^* = 2 \Phi \left(\frac{v}{\sqrt{\frac{\sigma^2}{n} + v^2}} \Phi^{-1} \left(\frac{\alpha_{\varphi_P^*}}{2} \right) \right). \tag{3.5}$$

This means that, in this particular case, through the P -value it is also possible to provide a cutoff value for the evidence in the FBST. Figure 3.26 shows for each prior the e -value versus P -value for different sample sizes. In this case, we have the functional relationship given by (3.4), so a simulation is not required, it is enough to plot $ev(p)$ for $0 \leq p \leq 1$. Here it is seen more clearly that, when the prior is more informative, the e -values are higher than the P -values for the smaller sample sizes. Again, for larger sample sizes, the two values are very close.

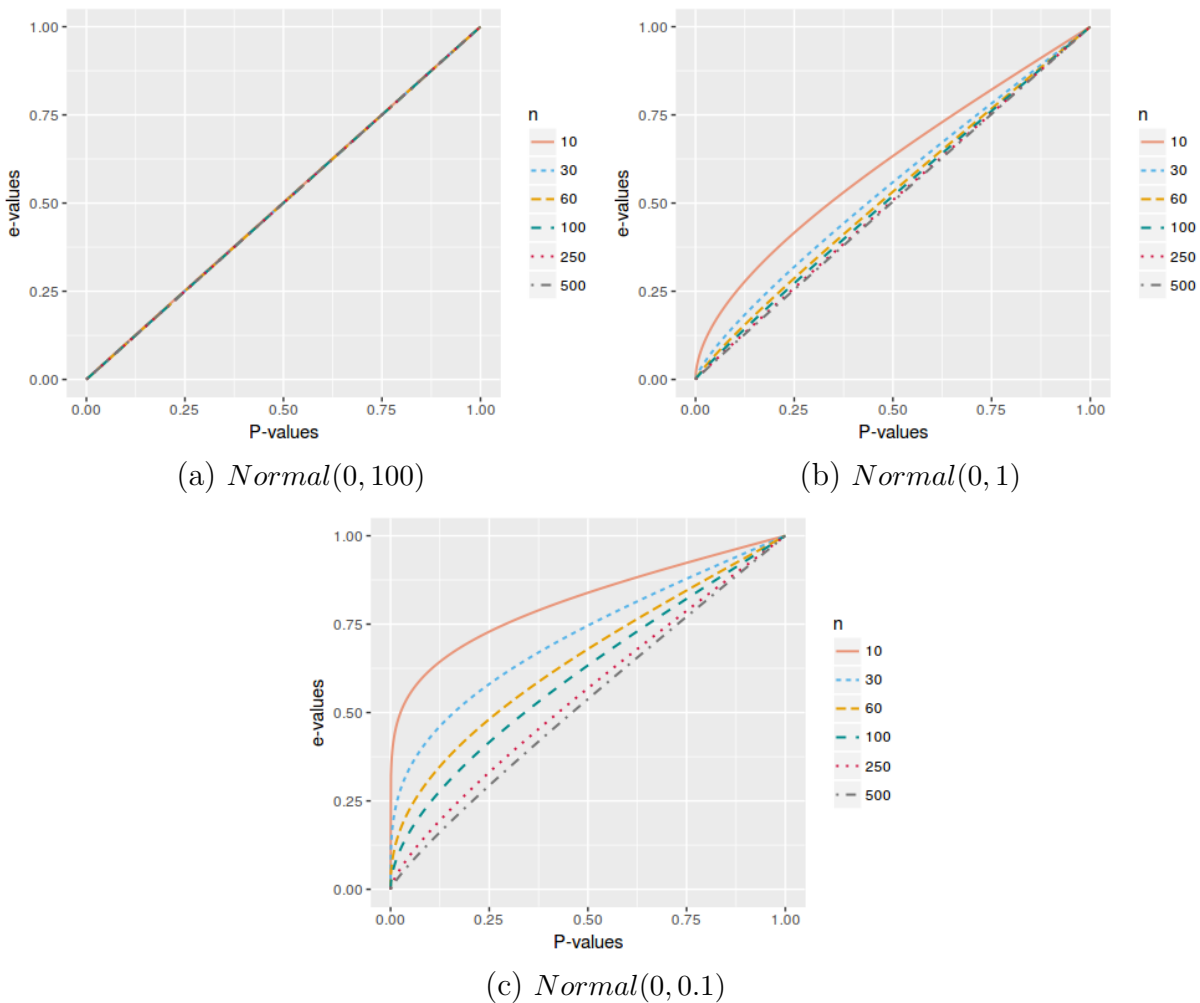


Figure 3.26: Relationship between e -value and P -value as function of n in normal mean hypothesis testing, with $\theta \sim Normal(0, 100)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 0.1)$.

The histograms in Figure 3.27, made with simulated samples for a single sample size $n = 100$, show that under \mathbf{H} , the indices have a uniform distribution. However, the e -value apparently begins to leave this distribution when the prior is more informative.

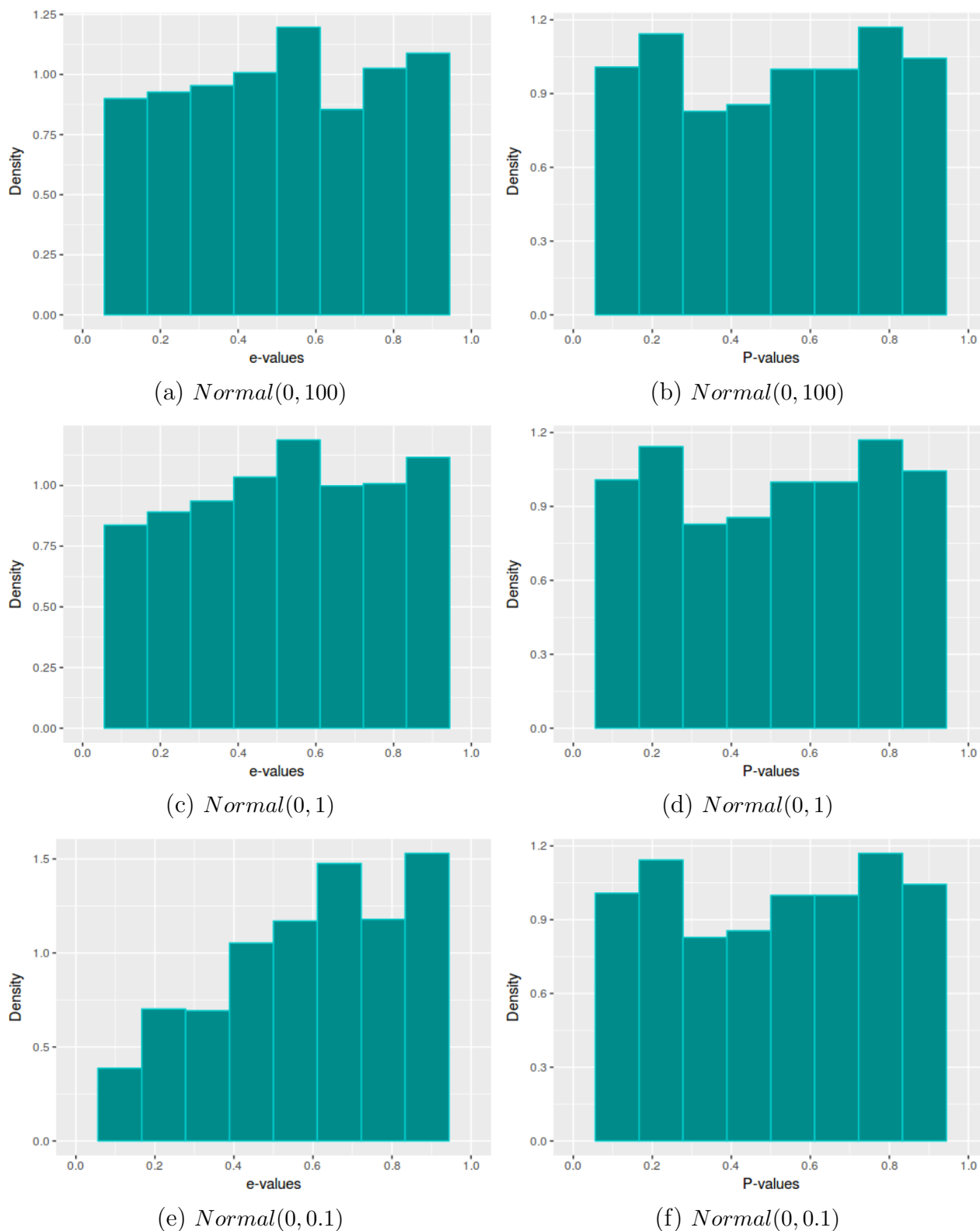


Figure 3.27: e -value and P -value densities under \mathbf{H} in normal mean hypothesis testing, with $\theta \sim Normal(0, 100)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 0.1)$. Sample size $n = 100$.

The optimal averaged error probabilities from e -values and P -values as well as the cutoff values for the three priors are shown in Tables 3.14, 3.15 and 3.16. Note that the $k_{\varphi_P}^*$ and k^* values are almost the same in all cases and the probabilities of both kinds of errors are very similar, which was already visible comparing Figures 3.22 (b), (d), (f) and 3.25. In this example, it was possible to make comparisons between exact cutoff values ($k_{\varphi_P}^*$) and cutoff values obtained by simulation (k^*), however, in more complex cases, such as those presented in the next chapter, it will only be possible to obtain these values via simulation.

n	$\alpha_{\varphi_P}^*$	$\alpha_{\varphi_e}^*$	$k_{\varphi_P}^*$	k^*	$\beta_{\varphi_P}^*$	$\beta_{\varphi_e}^*$
10	0.03085	0.03086	0.03169	0.03174	0.17340	0.17343
50	0.01257	0.01323	0.01266	0.01320	0.08560	0.09102
100	0.00854	0.00835	0.00858	0.00892	0.06400	0.06771
150	0.00682	0.00681	0.00684	0.00705	0.05520	0.05679
200	0.00582	0.00497	0.00583	0.00588	0.04950	0.05022
250	0.00515	0.00604	0.00515	0.00513	0.04550	0.04567
300	0.00465	0.00290	0.00466	0.00459	0.04240	0.04232
350	0.00428	0.00660	0.00428	0.00417	0.04040	0.03971
400	0.00397	0.00604	0.00398	0.00387	0.03770	0.03757
450	0.00372	0.00375	0.00373	0.00362	0.03490	0.03578
500	0.00351	0.00285	0.00352	0.00342	0.03390	0.03426
1000	0.00241	0.00310	0.00241	0.00233	0.02380	0.02577
1500	0.00193	0.00184	0.00193	0.00188	0.01920	0.02189
2000	0.00165	0.00094	0.00165	0.00166	0.01800	0.01945

Table 3.14: Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim Normal(0, 100)$.

n	$\alpha_{\varphi_P}^*$	$\alpha_{\varphi_e}^*$	$k_{\varphi_P}^*$	k^*	$\beta_{\varphi_P}^*$	$\beta_{\varphi_e}^*$
10	0.23903	0.23437	0.40510	0.40574	0.60100	0.59622
50	0.14256	0.12809	0.18071	0.18178	0.45460	0.45272
100	0.10436	0.11664	0.12150	0.12234	0.38100	0.37792
150	0.08548	0.08055	0.09589	0.09651	0.33640	0.33507
200	0.07378	0.06341	0.08101	0.08142	0.30850	0.30597
250	0.06566	0.05105	0.07107	0.07131	0.28830	0.28440
300	0.05960	0.04917	0.06387	0.06398	0.27190	0.26749
350	0.05487	0.05947	0.05836	0.05838	0.25810	0.25373
400	0.05106	0.05768	0.05397	0.05395	0.24560	0.24220
450	0.04789	0.05634	0.05038	0.05033	0.23550	0.23235
500	0.04522	0.04484	0.04738	0.04732	0.22600	0.22379
1000	0.03085	0.03086	0.03169	0.03174	0.17340	0.17343
1500	0.02462	0.03369	0.02510	0.02533	0.14880	0.14846
2000	0.02096	0.02618	0.02129	0.02168	0.13300	0.13255

Table 3.15: Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim Normal(0, 1)$.

n	$\alpha_{\varphi_P^*}$	$\alpha_{\varphi_e^*}^*$	$k_{\varphi_P^*}^*$	k^*	$\beta_{\varphi_P^*}$	$\beta_{\varphi_e^*}^*$
10	0.30587	0.23878	0.75753	0.75754	0.67660	0.67136
50	0.27007	0.26420	0.52428	0.52456	0.63740	0.63308
100	0.23903	0.23437	0.40510	0.40574	0.60100	0.59622
150	0.21654	0.23344	0.33845	0.33928	0.57190	0.56712
200	0.19924	0.23323	0.29457	0.29552	0.54870	0.54314
250	0.18539	0.20918	0.26303	0.26404	0.52700	0.52286
300	0.17397	0.18696	0.23903	0.24006	0.50690	0.50533
350	0.16434	0.17326	0.22005	0.22113	0.49060	0.48990
400	0.15608	0.15115	0.20457	0.20565	0.47600	0.47620
450	0.14889	0.13221	0.19167	0.19274	0.46450	0.46389
500	0.14256	0.12809	0.18071	0.18178	0.45460	0.45272
1000	0.10436	0.11664	0.12150	0.12234	0.38100	0.37792
1500	0.08548	0.08055	0.09589	0.09651	0.33640	0.33507
2000	0.07378	0.06341	0.08101	0.08142	0.30850	0.30597

Table 3.16: Optimal averaged error probabilities and cutoff values k^* as functions of n , with $\theta \sim Normal(0,0.1)$.

Lastly, this time we choose a single observation $\bar{x}_0 = 0.3$ to calculate the e -value and P -value for the three priors. The results are presented in Table 3.17. The cells in **boldface** represent the cases when \mathbf{H} is rejected. It can be seen that the decision is the same whatever the index is used. Here, when the prior is more informative, i.e., in the $\theta \sim Normal(0,0.1)$ case, the null hypothesis is more easily rejected. This is because the corresponding posterior has smaller variance, as seen in Figure 3.28. Observe that, unlike the e -value, the P -value does not depend on the hyperparameter v^2 , as the expression in (3.2) shows, so it takes the same value regardless of the prior.

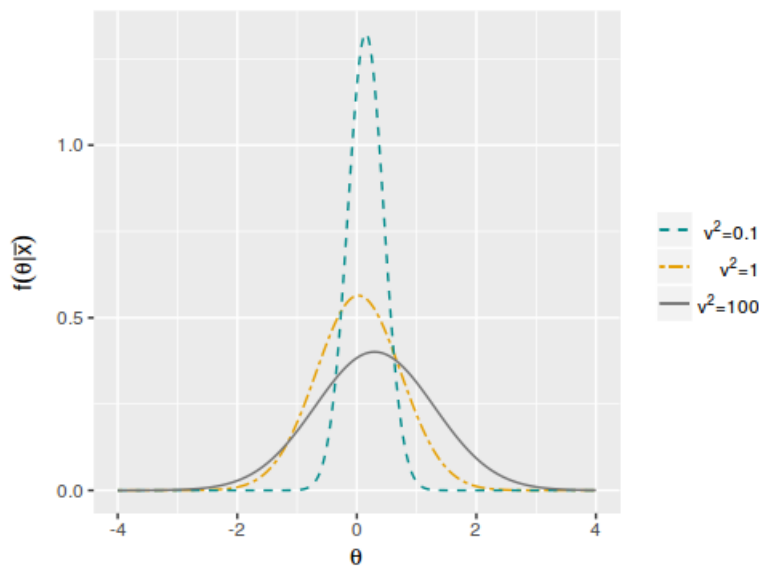


Figure 3.28: Posterior distributions $\theta \sim Normal(0,0.1)$, $\theta \sim Normal(0,1)$, $\theta \sim Normal(0,100)$, $\bar{x}_0 = 0.3$ and $n = 10$.

$v^2 = 100, \bar{x}_0 = 0.3$					$v^2 = 1, \bar{x}_0 = 0.3$				$v^2 = 0.1, \bar{x}_0 = 0.3$			
n	k^*	ev	$\alpha_{\varphi_P^*}$	Pv	k^*	ev	$\alpha_{\varphi_P^*}$	Pv	k^*	ev	$\alpha_{\varphi_P^*}$	Pv
10	0.0317	0.7653	0.0309	0.7642	0.4057	0.8320	0.2390	0.7642	0.7575	0.9279	0.3059	0.7642
50	0.0132	0.5028	0.0126	0.5023	0.1818	0.5403	0.1426	0.5023	0.5246	0.6985	0.2701	0.5023
100	0.0089	0.3430	0.0085	0.3428	0.1223	0.3657	0.1044	0.3428	0.4057	0.5023	0.2390	0.3428
150	0.0070	0.2454	0.0068	0.2453	0.0965	0.2606	0.0855	0.2453	0.3393	0.3681	0.2165	0.2453
200	0.0059	0.1798	0.0058	0.1797	0.0814	0.1904	0.0738	0.1797	0.2955	0.2733	0.1992	0.1797
250	0.0051	0.1337	0.0051	0.1336	0.0713	0.1413	0.0657	0.1336	0.2640	0.2049	0.1854	0.1336
300	0.0046	0.1004	0.0047	0.1003	0.0640	0.1060	0.0596	0.1003	0.2401	0.1547	0.1740	0.1003
350	0.0042	0.0760	0.0043	0.0759	0.0584	0.0801	0.0549	0.0759	0.2211	0.1175	0.1643	0.0759
400	0.0039	0.0578	0.0040	0.0578	0.0540	0.0609	0.0511	0.0578	0.2056	0.0897	0.1561	0.0578
450	0.0036	0.0442	0.0037	0.0442	0.0503	0.0465	0.0479	0.0442	0.1927	0.0687	0.1489	0.0442
500	0.0034	0.0339	0.0035	0.0339	0.0473	0.0357	0.0452	0.0339	0.1818	0.0528	0.1426	0.0339
1000	0.0023	0.0027	0.0024	0.0027	0.0317	0.0028	0.0309	0.0027	0.1223	0.0042	0.1044	0.0027
1500	0.0019	0.0002	0.0019	0.0002	0.0253	0.0003	0.0246	0.0002	0.0965	0.0004	0.0855	0.0002
2000	0.0017	0.0000	0.0016	0.0000	0.0217	0.0000	0.0210	0.0000	0.0814	0.0000	0.0738	0.0000

Table 3.17: Cutoff values k^* , $ev(0; \bar{x}_0)$, $\alpha_{\varphi_P^*}$ and P -value(\bar{x}_0) as functions of n , with $\theta \sim Normal(0, 100)$, $\theta \sim Normal(0, 1)$ and $\theta \sim Normal(0, 0.1)$.

Adaptive Significance Levels in Linear Regression Models

Linear models are probably the most used statistical models to establish the influence of a set of covariates on a response variable. In that sense, the proper identification of the relevant variables in the model is an important issue in many scientific researches. Such identification can be done through hypothesis-testing procedures involving the respective regression coefficients. In the conjugate Bayesian analysis of the normal linear regression model, it is possible to obtain expressions for the posterior distributions of the parameters and their respective marginals. Therefore, given a model, the FBST can be used for testing if one or more of its regression coefficients are null, which is the basis of one possible model-selection procedure. In this chapter, we present the proposed methodology to determine the adaptive cutoff value k^* for evidence extended to the context of linear regression models under a Bayesian approach. Emphasis is on analysis of two formulations of the normal linear model using conjugate priors. Specifically, the unknown variance model and the known variance model are considered. Likewise, all the developments to calculate the adaptive significance levels using P -value are described. Additional proofs and explanations about the models can be found in [O’Hagan and Forster \(2004\)](#) and [Banerjee \(2008\)](#).

Consider the usual multiple linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{4.1}$$

where $\mathbf{y} = (y_1, \dots, y_n)^\top$ is an $n \times 1$ vector of y_i observations, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ is an

$n \times p$ matrix of covariates, also called the design matrix, with $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip-1})^\top$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top$ is a $p \times 1$ vector of parameters (regression coefficients) and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$ an $n \times 1$ vector of random errors. The elements of $\boldsymbol{\varepsilon}$ are assumed to have zero mean, to be uncorrelated and to have common variance σ^2 , which is an additional parameter.

If we further assume that the elements of $\boldsymbol{\varepsilon}$ are jointly normally distributed, i.e. $N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, then the model is described as the *normal linear model*. The model says simply that the conditional distribution of \mathbf{y} given parameters $(\boldsymbol{\theta}, \sigma^2)$ is the multivariate normal distribution $N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n)$.

Therefore, the likelihood becomes

$$f(\mathbf{y}|\boldsymbol{\theta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) \right\}. \quad (4.2)$$

We can write the quadratic form $(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})$ in the exponent of (4.2) in various ways. Just expanding,

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) = \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{X} \boldsymbol{\theta} - \boldsymbol{\theta}^\top \mathbf{X}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{X} \boldsymbol{\theta} + \mathbf{y}^\top \mathbf{y}.$$

If the $p \times p$ matrix $\mathbf{X}^\top \mathbf{X}$ is non-singular we can go on to complete the square to give the likelihood

$$f(\mathbf{y}|\boldsymbol{\theta}, \sigma^2) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \left[ks^2 + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top \mathbf{X}^\top \mathbf{X} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right] \right\}, \quad (4.3)$$

where $\hat{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ is the classical Maximum Likelihood or Least Squares estimator of $\boldsymbol{\theta}$, and $ks^2 = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\theta}})$ is the residual sum of squares, with $k = n - p$. Notice that $(\hat{\boldsymbol{\theta}}, s^2)$ is the sufficient statistic value with respect to the linear model considered.

4.1 Unknown-variance model

The natural conjugate prior distribution of $(\boldsymbol{\theta}, \sigma^2)$ is a p -variate Normal-Inverse Gamma distribution with hyperparameters \mathbf{m}_0 , \mathbf{V}_0 , a_0 and b_0 , denoted by $(\boldsymbol{\theta}, \sigma^2) \sim N_p IG(\mathbf{m}_0, \mathbf{V}_0, a_0, b_0)$:

$$g(\boldsymbol{\theta}, \sigma^2) = \frac{(b_0)^{a_0}}{(2\pi)^{p/2} |\mathbf{V}_0|^{1/2} \Gamma(a_0)} (\sigma^2)^{-(a_0 + \frac{p}{2} + 1)} \exp \left\{ -\frac{1}{2\sigma^2} [(\boldsymbol{\theta} - \mathbf{m}_0)^\top \mathbf{V}_0^{-1} (\boldsymbol{\theta} - \mathbf{m}_0) + 2b_0] \right\}, \quad (4.4)$$

such that, the conditional prior distributions of $\boldsymbol{\theta}$ given σ^2 is

$$g(\boldsymbol{\theta} | \sigma^2) = (2\pi)^{-p/2} |\mathbf{V}_0|^{-1/2} (\sigma^2)^{-p/2} \exp \left\{ -\frac{1}{2\sigma^2} [(\boldsymbol{\theta} - \mathbf{m}_0)^\top \mathbf{V}_0^{-1} (\boldsymbol{\theta} - \mathbf{m}_0)] \right\}, \quad (4.5)$$

and the prior marginal distribution of σ^2 is

$$g(\sigma^2) = \frac{(b_0)^{a_0}}{\Gamma(a_0)} (\sigma^2)^{-(a_0+1)} \exp \left\{ -\frac{b_0}{\sigma^2} \right\}, \quad (4.6)$$

denoted, respectively, as

$$\boldsymbol{\theta} | \sigma^2 \sim N_p(\mathbf{m}_0, \sigma^2 \mathbf{V}_0), \quad \sigma^2 \sim IG(a_0, b_0). \quad (4.7)$$

Both distributions are equivalent to the new pair

$$g(\sigma^2 | \boldsymbol{\theta}) = \frac{\left(b_0 + \frac{(\boldsymbol{\theta} - \mathbf{m}_0)^\top \mathbf{V}_0^{-1} (\boldsymbol{\theta} - \mathbf{m}_0)}{2} \right)^{\left(a_0 + \frac{p}{2} \right)}}{\Gamma\left(a_0 + \frac{p}{2} \right)} (\sigma^2)^{-(a_0 + \frac{p}{2} + 1)} \times \exp \left\{ -\frac{1}{2\sigma^2} [(\boldsymbol{\theta} - \mathbf{m}_0)^\top \mathbf{V}_0^{-1} (\boldsymbol{\theta} - \mathbf{m}_0) + 2b_0] \right\}, \quad (4.8)$$

and

$$g(\boldsymbol{\theta}) = \frac{(2b_0)^{a_0} \Gamma\left(a_0 + \frac{p}{2} \right)}{\pi^{p/2} |\mathbf{V}_0|^{1/2} \Gamma(a_0)} \left\{ (\boldsymbol{\theta} - \mathbf{m}_0)^\top \mathbf{V}_0^{-1} (\boldsymbol{\theta} - \mathbf{m}_0) + 2b_0 \right\}^{-\left(a_0 + \frac{p}{2} \right)} \propto \left\{ 1 + (\boldsymbol{\theta} - \mathbf{m}_0)^\top (2b_0 \mathbf{V}_0)^{-1} (\boldsymbol{\theta} - \mathbf{m}_0) \right\}^{-\left(a_0 + \frac{p}{2} \right)}. \quad (4.9)$$

The density in (4.9) is a p -variate t distribution with $2a_0$ degrees of freedom and hyper-

parameters \mathbf{m}_0 and $(b_0/a_0)\mathbf{V}_0$. Then these distributions are denoted by

$$\sigma^2|\boldsymbol{\theta} \sim IG\left(a_0 + \frac{p}{2}, b_0 + \frac{(\boldsymbol{\theta} - \mathbf{m}_0)^\top \mathbf{V}_0^{-1}(\boldsymbol{\theta} - \mathbf{m}_0)}{2}\right), \quad \boldsymbol{\theta} \sim t_p\left(2a_0; \mathbf{m}_0, \frac{b_0}{a_0}\mathbf{V}_0\right). \quad (4.10)$$

Now suppose that the $N_p IG(\mathbf{m}_0, \mathbf{V}_0, a_0, b_0)$ distribution (4.4) is adopted as the prior distribution for $(\boldsymbol{\theta}, \sigma^2)$. Combining it with the likelihood (4.2) gives the posterior distribution

$$f(\boldsymbol{\theta}, \sigma^2|\mathbf{y}) \propto (\sigma^2)^{-(a_0 + \frac{n}{2} + \frac{p}{2} + 1)} \exp\left\{-\frac{1}{2\sigma^2} [(\boldsymbol{\theta} - \mathbf{m}^*)^\top \mathbf{V}^{*-1}(\boldsymbol{\theta} - \mathbf{m}^*) + 2b_1]\right\}, \quad (4.11)$$

where

$$\mathbf{V}^* = (\mathbf{V}_0^{-1} + \mathbf{X}^\top \mathbf{X})^{-1}, \quad \mathbf{m}^* = \mathbf{V}^* (\mathbf{V}_0^{-1}\mathbf{m}_0 + \mathbf{X}^\top \mathbf{y}),$$

$$a_1 = a_0 + \frac{n}{2}, \quad b_1 = b_0 + \frac{\mathbf{m}_0^\top \mathbf{V}_0^{-1}\mathbf{m}_0 + \mathbf{y}^\top \mathbf{y} - \mathbf{m}^{*\top} \mathbf{V}^{*-1}\mathbf{m}^*}{2}.$$

If $\mathbf{X}^\top \mathbf{X}$ is non-singular, we can write

$$\mathbf{m}^* = \mathbf{V}^* (\mathbf{V}_0^{-1}\mathbf{m}_0 + \mathbf{X}^\top \mathbf{X}\hat{\boldsymbol{\theta}}).$$

Therefore, the posterior distribution of $(\boldsymbol{\theta}, \sigma^2)$ is

$$(\boldsymbol{\theta}, \sigma^2)|\mathbf{y} \sim N_p IG(\mathbf{m}^*, \mathbf{V}^*, a_1, b_1). \quad (4.12)$$

Consequently,

$$\boldsymbol{\theta}|\sigma^2, \mathbf{y} \sim N_p(\mathbf{m}^*, \sigma^2\mathbf{V}^*), \quad \sigma^2|\mathbf{y} \sim IG(a_1, b_1), \quad (4.13)$$

and this is equivalent to,

$$\sigma^2|\boldsymbol{\theta}, \mathbf{y} \sim IG\left(a_1 + \frac{p}{2}, b_1 + \frac{(\boldsymbol{\theta} - \mathbf{m}^*)^\top \mathbf{V}^{*-1}(\boldsymbol{\theta} - \mathbf{m}^*)}{2}\right), \quad (4.14)$$

$$\boldsymbol{\theta}|\mathbf{y} \sim t_p\left(2a_1; \mathbf{m}^*, \frac{b_1}{a_1}\mathbf{V}^*\right). \quad (4.15)$$

4.1.1 Conditional distributions

Consider now conditional distributions given partial specification of $\boldsymbol{\theta}$. First let $\boldsymbol{\theta}^\top = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)$, and consider distributions conditional on $\boldsymbol{\theta}_2$. Suppose that $(\boldsymbol{\theta}, \sigma^2) \sim N_p IG(\mathbf{m}_0, \mathbf{V}_0, a_0, b_0)$. Corresponding distributions result if we change a_0 to a_1 , b_0 to b_1 , \mathbf{m}_0 to \mathbf{m}^* and \mathbf{V}_0 to \mathbf{V}^* . If $\boldsymbol{\theta}_1$ has s elements and $\boldsymbol{\theta}_2$ has r elements, write

$$\mathbf{m}_0 = \begin{bmatrix} \mathbf{m}_{01} \\ \mathbf{m}_{02} \end{bmatrix}, \quad \mathbf{V}_0 = \begin{bmatrix} \mathbf{V}_{011} & \mathbf{V}_{012} \\ \mathbf{V}_{021} & \mathbf{V}_{022} \end{bmatrix},$$

where \mathbf{m}_{01} is $s \times 1$, \mathbf{V}_{011} is $s \times s$, \mathbf{m}_{02} is $r \times 1$, \mathbf{V}_{022} is $r \times r$, with $r = p - s$. Now since $\boldsymbol{\theta}$ given σ^2 is distributed as $N_p(\mathbf{m}_0, \sigma^2 \mathbf{V}_0)$, using general results on multivariate normal distributions (see [Mardia *et al.* \(1979\)](#)), we have the following distributions:

$$\boldsymbol{\theta}_2 | \sigma^2 \sim N_r(\mathbf{m}_{02}, \sigma^2 \mathbf{V}_{022}), \quad (4.16)$$

$$(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2, \sigma^2) \sim N_s(\mathbf{m}_{01.2}(\boldsymbol{\theta}_2), \sigma^2 \mathbf{V}_{011.2}), \quad (4.17)$$

where $\mathbf{m}_{01.2}(\boldsymbol{\theta}_2) = \mathbf{m}_{01} + \mathbf{V}_{012} \mathbf{V}_{022}^{-1} (\boldsymbol{\theta}_2 - \mathbf{m}_{02})$ and $\mathbf{V}_{011.2} = \mathbf{V}_{011} - \mathbf{V}_{012} \mathbf{V}_{022}^{-1} \mathbf{V}_{021}$.

From (4.16) and the $IG(a_0, b_0)$ distribution of σ^2 we have the distribution

$$(\boldsymbol{\theta}_2, \sigma^2) \sim N_r IG(\mathbf{m}_{02}, \mathbf{V}_{022}, a_0, b_0) \quad (4.18)$$

and hence

$$\boldsymbol{\theta}_2 \sim t_r \left(2a_0; \mathbf{m}_{02}, \frac{b_0}{a_0} \mathbf{V}_{022} \right), \quad (4.19)$$

$$\sigma^2 | \boldsymbol{\theta}_2 \sim IG \left(a_0 + \frac{r}{2}, b_0 + \frac{(\boldsymbol{\theta}_2 - \mathbf{m}_{02})^\top \mathbf{V}_{022}^{-1} (\boldsymbol{\theta}_2 - \mathbf{m}_{02})}{2} \right), \quad (4.20)$$

Now (4.17) and (4.20) together give

$$(\boldsymbol{\theta}_1, \sigma^2 | \boldsymbol{\theta}_2) \sim N_s IG \left(\mathbf{m}_{01.2}(\boldsymbol{\theta}_2), \mathbf{V}_{011.2}, a_0 + \frac{r}{2}, b_0 + \frac{(\boldsymbol{\theta}_2 - \mathbf{m}_{02})^\top \mathbf{V}_{022}^{-1} (\boldsymbol{\theta}_2 - \mathbf{m}_{02})}{2} \right) \quad (4.21)$$

and finally

$$\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2 \sim t_s \left(2a_0 + r; \mathbf{m}_{01.2}(\boldsymbol{\theta}_2), \frac{2b_0 + (\boldsymbol{\theta}_2 - \mathbf{m}_{02})^\top \mathbf{V}_{022}^{-1} (\boldsymbol{\theta}_2 - \mathbf{m}_{02})}{2a_0 + r} \mathbf{V}_{011.2} \right). \quad (4.22)$$

4.1.2 Marginal distributions

The following results are defined in Banerjee (2008). The model (4.1) expresses \mathbf{y} as a sum of $\mathbf{X}\boldsymbol{\theta}$ and $\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is distributed as $N_n(0, \sigma^2 \mathbf{I}_n)$ given σ^2 , and $\boldsymbol{\theta}$ is distributed as $N_p(\mathbf{m}_0, \sigma^2 \mathbf{V}_0)$ also given σ^2 . Therefore the distribution of \mathbf{y} given σ^2 can be calculated as

$$\begin{aligned} f(\mathbf{y} | \sigma^2) &= \int f(\mathbf{y} | \boldsymbol{\theta}, \sigma^2) f(\boldsymbol{\theta} | \sigma^2) d\boldsymbol{\theta} = \int N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n) \times N_p(\mathbf{m}_0, \sigma^2 \mathbf{V}_0) d\boldsymbol{\theta} \\ &= N_n(\mathbf{X}\mathbf{m}_0, \sigma^2 (\mathbf{I}_n + \mathbf{X}\mathbf{V}_0\mathbf{X}^\top)). \end{aligned} \quad (4.23)$$

With the prior distribution of σ^2 being $IG(a_0, b_0)$, it follows that the joint distribution of (\mathbf{y}, σ^2) is $N_n IG(\mathbf{X}\mathbf{m}_0, \sigma^2 (\mathbf{I}_n + \mathbf{X}\mathbf{V}_0\mathbf{X}^\top), a_0, b_0)$. So the marginal distribution of $f(\mathbf{y})$ is obtained by integrating the $N_n IG$ as follows

$$\begin{aligned} f(\mathbf{y}) &= \int f(\mathbf{y} | \sigma^2) f(\sigma^2) d\sigma^2 = \int N_n(\mathbf{X}\mathbf{m}_0, \sigma^2 (\mathbf{I}_n + \mathbf{X}\mathbf{V}_0\mathbf{X}^\top)) \times IG(a_0, b_0) d\sigma^2 \\ &= \int N_n IG(\mathbf{X}\mathbf{m}_0, \sigma^2 (\mathbf{I}_n + \mathbf{X}\mathbf{V}_0\mathbf{X}^\top), a_0, b_0) d\sigma^2 \\ &= t_n \left(2a_0; \mathbf{X}\mathbf{m}_0, \frac{b_0}{a_0} (\mathbf{I}_n + \mathbf{X}\mathbf{V}_0\mathbf{X}^\top) \right). \end{aligned} \quad (4.24)$$

Rewriting this result in another way reveals another useful property of the NIG distribution

$$\begin{aligned} f(\mathbf{y}) &= \int f(\mathbf{y} | \boldsymbol{\theta}, \sigma^2) f(\boldsymbol{\theta}, \sigma^2) d\boldsymbol{\theta} d\sigma^2 \\ &= \int N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n) \times N_p IG(\mathbf{m}_0, \mathbf{V}_0, a_0, b_0) d\boldsymbol{\theta} d\sigma^2 \\ &= t_n \left(2a_0; \mathbf{X}\mathbf{m}_0, \frac{b_0}{a_0} (\mathbf{I}_n + \mathbf{X}\mathbf{V}_0\mathbf{X}^\top) \right). \end{aligned} \quad (4.25)$$

The computation of $f(\mathbf{y})$ could also be carried out in terms of the $N_p IG$ distribution

parameters more directly as

$$\begin{aligned}
 f(\mathbf{y}) &= \int f(\mathbf{y}|\boldsymbol{\theta}, \sigma^2) f(\boldsymbol{\theta}, \sigma^2) d\boldsymbol{\theta} d\sigma^2 \\
 &= \int N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n) \times N_p IG(\mathbf{m}_0, \mathbf{V}_0, a_0, b_0) d\boldsymbol{\theta} d\sigma^2 \\
 &= \frac{(b_0)^{a_0} \Gamma\left(a_0 + \frac{n}{2}\right) |\mathbf{V}^*|^{1/2}}{(2\pi)^{n/2} \Gamma(a_0) |\mathbf{V}_0|^{1/2}} \left[b_0 + \frac{\mathbf{m}_0^\top \mathbf{V}_0^{-1} \mathbf{m}_0 + \mathbf{y}^\top \mathbf{y} - \mathbf{m}^{*\top} \mathbf{V}_0^{-1} \mathbf{m}^*}{2} \right]^{-(a_0+n/2)} \\
 &= \frac{(b_0)^{a_0} \Gamma(a_1) |\mathbf{V}^*|^{1/2}}{(2\pi)^{n/2} (b_1)^{a_1} \Gamma(a_0) |\mathbf{V}_0|^{1/2}}. \tag{4.26}
 \end{aligned}$$

4.1.3 Prior predictive densities in regression-coefficient hypothesis testing

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$, with $\boldsymbol{\theta}_1 = (\theta_1, \dots, \theta_s)^\top$ and $\boldsymbol{\theta}_2 = (\theta_{s+1}, \dots, \theta_p)^\top$. Then, $\mathbf{Y}|\boldsymbol{\xi} \sim N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n)$ where $\boldsymbol{\xi} \in \Xi$ is the parameter vector such that $\boldsymbol{\xi} = (\boldsymbol{\theta}^\top, \sigma^2)^\top = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \sigma^2)^\top$.

We are interested about testing the hypotheses

$$\mathbf{H} : \boldsymbol{\theta}_2 = \mathbf{0}$$

$$\mathbf{A} : \boldsymbol{\theta}_2 \neq \mathbf{0}.$$

The prior predictive density under \mathbf{H} is

$$\begin{aligned}
 f_{\mathbf{H}}(\mathbf{y}) &= \int_{\mathbf{H}} f(\mathbf{y}|\boldsymbol{\xi}) dP_{\mathbf{H}}(\boldsymbol{\xi}) \\
 &= \int_{\mathbf{H}} f(\mathbf{y}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) g_{\mathbf{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 d\sigma^2,
 \end{aligned}$$

where $g_{\mathbf{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2)$ is the prior density under the subset of smaller dimensionality \mathbf{H} calculated as

$$\begin{aligned}
 g_{\mathbf{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) &= g_{\mathbf{H}}(\boldsymbol{\theta}_1, \sigma^2) = \frac{g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \mathbb{1}(\boldsymbol{\theta}_2 = \mathbf{0})}{\int_{\mathbf{H}} g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 d\sigma^2} \\
 &= \frac{g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \mathbb{1}(\boldsymbol{\theta}_2 = \mathbf{0})}{\int_{\boldsymbol{\theta}_1 \in \mathbb{R}^s, \sigma^2 \in \mathbb{R}_+} g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2) d\boldsymbol{\theta}_1 d\sigma^2}
 \end{aligned}$$

$$= g(\boldsymbol{\theta}_1, \sigma^2 | \boldsymbol{\theta}_2 = \mathbf{0}).$$

Thus, $f_{\mathbf{H}}(\mathbf{y})$ is given by

$$\begin{aligned} f_{\mathbf{H}}(\mathbf{y}) &= \int_{\mathbf{H}} f(\mathbf{y} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) g(\boldsymbol{\theta}_1, \sigma^2 | \boldsymbol{\theta}_2 = \mathbf{0}) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 d\sigma^2 \\ &= \int_{\boldsymbol{\theta}_1 \in \mathbb{R}^s, \sigma^2 \in \mathbb{R}_+} f(\mathbf{y} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2) g(\boldsymbol{\theta}_1, \sigma^2 | \boldsymbol{\theta}_2 = \mathbf{0}) d\boldsymbol{\theta}_1 d\sigma^2 \\ &= \int_{\boldsymbol{\theta}_1 \in \mathbb{R}^s, \sigma^2 \in \mathbb{R}_+} N_n(\mathbf{X}\mathbf{C}\boldsymbol{\theta}_1, \sigma^2 \mathbf{I}_n) \times \\ &\quad N_s IG \left(\mathbf{m}_{01.2}(\mathbf{0}), \mathbf{V}_{011.2}, a_0 + \frac{r}{2}, b_0 + \frac{\mathbf{m}_{02}^\top (\mathbf{V}_{022})^{-1} \mathbf{m}_{02}}{2} \right) d\boldsymbol{\theta}_1 d\sigma^2 \\ &= t_n \left(2 \left(a_0 + \frac{r}{2} \right); \mathbf{X}\mathbf{C}\mathbf{m}_{01.2}(\mathbf{0}), \frac{b_0 + \frac{\mathbf{m}_{02}^\top (\mathbf{V}_{022})^{-1} \mathbf{m}_{02}}{2}}{\left(a_0 + \frac{r}{2} \right)} \left(\mathbf{I}_n + (\mathbf{X}\mathbf{C})\mathbf{V}_{011.2}(\mathbf{X}\mathbf{C})^\top \right) \right), \end{aligned} \tag{4.27}$$

where $\mathbf{C}_{(s+r) \times s} = [\mathbf{I}_s, \mathbf{0}_{s \times r}]^\top$.

The prior predictive density under \mathbf{A} can be obtained from the result in (4.25) as follows

$$\begin{aligned} f_{\mathbf{A}}(\mathbf{y}) &= \int_{\mathbf{A}} f(\mathbf{y} | \boldsymbol{\xi}) dP_{\mathbf{A}}(\boldsymbol{\xi}) \\ &= \int_{\mathbf{A}} f(\mathbf{y} | \boldsymbol{\theta}, \sigma^2) g_{\mathbf{A}}(\boldsymbol{\theta}, \sigma^2) d\boldsymbol{\theta} d\sigma^2 \\ &= \int_{\mathbf{A}} f(\mathbf{y} | \boldsymbol{\theta}, \sigma^2) g(\boldsymbol{\theta}, \sigma^2) d\boldsymbol{\theta} d\sigma^2 \\ &= \int_{\mathbf{A}} N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n) \times N_p IG(\mathbf{m}_0, \mathbf{V}_0, a_0, b_0) d\boldsymbol{\theta} d\sigma^2 \\ &= t_n \left(2a_0; \mathbf{X}\mathbf{m}_0, \frac{b_0}{a_0} (\mathbf{I}_n + \mathbf{X}\mathbf{V}_0\mathbf{X}^\top) \right). \end{aligned} \tag{4.28}$$

It is necessary to consider the particular case where $\mathbf{X} = \mathbf{1}_n$ (the dimensionality of the parameter space is $d = \dim(\boldsymbol{\Xi}) = 2$). Therefore, consider the linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \tag{4.29}$$

here $\mathbf{y} = (y_1, \dots, y_n)^\top$ is an $n \times 1$ vector of y_i observations, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\top$ is the $n \times 1$

vector of random errors and $\theta = \theta_1$ is the regression coefficient. Then, $\mathbf{Y}|\boldsymbol{\xi} \sim N_n(\mathbf{1}_n\theta_1, \sigma^2\mathbf{I}_n)$ where $\boldsymbol{\xi} \in \Xi$ is the parameter vector such that $\boldsymbol{\xi} = (\theta_1, \sigma^2)^\top$. The hypotheses of interest are

$$\mathbf{H} : \theta_1 = 0$$

$$\mathbf{A} : \theta_1 \neq 0.$$

The prior predictive density under \mathbf{H} is

$$\begin{aligned} f_{\mathbf{H}}(\mathbf{y}) &= \int_{\mathbf{H}} f(\mathbf{y}|\boldsymbol{\xi}) dP_{\mathbf{H}}(\boldsymbol{\xi}) \\ &= \int_{\mathbf{H}} f(\mathbf{y}|\theta_1, \sigma^2) g_{\mathbf{H}}(\theta_1, \sigma^2) d\theta_1 d\sigma^2, \end{aligned}$$

$g_{\mathbf{H}}(\theta_1, \sigma^2)$ is calculated as

$$g_{\mathbf{H}}(\theta_1, \sigma^2) = \frac{g(\theta_1, \sigma^2) \mathbb{1}(\theta_1 = 0)}{\int_{\mathbf{H}} g(\theta_1, \sigma^2) d\theta_1 d\sigma^2} = \frac{g(\theta_1, \sigma^2) \mathbb{1}(\theta_1 = 0)}{\int_{\sigma^2 \in \mathbb{R}_+} g(\theta_1 = 0, \sigma^2) d\sigma^2} = g(\sigma^2|\theta_1 = 0).$$

So $f_{\mathbf{H}}(\mathbf{y})$ is given by

$$\begin{aligned} f_{\mathbf{H}}(\mathbf{y}) &= \int_{\mathbf{H}} f(\mathbf{y}|\theta_1, \sigma^2) g(\sigma^2|\theta_1 = 0) d\theta_1 d\sigma^2 \\ &= \int_{\sigma^2 \in \mathbb{R}_+} f(\mathbf{y}|\theta_1 = 0, \sigma^2) g(\sigma^2|\theta_1 = 0) d\sigma^2 \\ &= \int_{\sigma^2 \in \mathbb{R}_+} N_n(\mathbf{0}, \sigma^2\mathbf{I}_n) \times IG\left(a_0 + \frac{1}{2}, b_0 + \frac{\mathbf{m}_0^2 \mathbf{V}_0^{-1}}{2}\right) d\sigma^2 \\ &= \frac{\Gamma\left(a_1 + \frac{1}{2}\right) \left[b_0 + \frac{\mathbf{m}_0^2 \mathbf{V}_0^{-1}}{2}\right]^{(a_0+1/2)}}{(2\pi)^{n/2} \Gamma\left(a_0 + \frac{1}{2}\right) \left[b_0 + \frac{\mathbf{y}^\top \mathbf{y} + \mathbf{m}_0^2 \mathbf{V}_0^{-1}}{2}\right]^{(a_1+1/2)}}, \end{aligned} \tag{4.30}$$

and the prior predictive density under \mathbf{A} in this case will be

$$\begin{aligned} f_{\mathbf{A}}(\mathbf{y}) &= \int_{\mathbf{A}} f(\mathbf{y}|\boldsymbol{\xi}) dP_{\mathbf{A}}(\boldsymbol{\xi}) \\ &= \int_{\mathbf{A}} f(\mathbf{y}|\theta_1, \sigma^2) g_{\mathbf{A}}(\theta_1, \sigma^2) d\theta_1 d\sigma^2 \end{aligned}$$

$$\begin{aligned}
&= \int_{\mathbf{A}} f(\mathbf{y}|\theta_1, \sigma^2) g(\theta_1, \sigma^2) d\theta_1 d\sigma^2 \\
&= \int_{\mathbf{A}} N_n(\mathbf{1}_n\theta_1, \sigma^2\mathbf{I}_n) \times NIG(\mathbf{m}_0, \mathbf{V}_0, a_0, b_0) d\theta_1 d\sigma^2 \\
&= t_n \left(2a_0; \mathbf{1}_n\mathbf{m}_0, \frac{b_0}{a_0} (\mathbf{I}_n + \mathbf{1}_n\mathbf{V}_0\mathbf{1}_n^\top) \right). \tag{4.31}
\end{aligned}$$

4.2 Known-variance model

Suppose the residual error variance σ^2 is known. We now regard σ^2 as fixed in the likelihood (4.2), and denote it by $f(\mathbf{y}|\boldsymbol{\theta})$. Only $\boldsymbol{\theta}$ is unknown, and we require a suitable prior distribution $g(\boldsymbol{\theta})$. Using also (4.3), the likelihood simplifies to

$$f(\mathbf{y}|\boldsymbol{\theta}) \propto \exp \left\{ - \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \right)^\top \mathbf{X}^\top \mathbf{X} \left(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}} \right) / (2\sigma^2) \right\}, \tag{4.32}$$

and therefore the natural conjugate prior family is the family of multivariate normal distributions. Suppose therefore that $\boldsymbol{\theta}$ has the $N_p(\mathbf{m}_0, \mathbf{W}_0)$ prior distribution

$$g(\boldsymbol{\theta}) \propto \exp \left\{ - \frac{(\boldsymbol{\theta} - \mathbf{m}_0)^\top \mathbf{W}_0^{-1} (\boldsymbol{\theta} - \mathbf{m}_0)}{2} \right\}. \tag{4.33}$$

Then $f(\boldsymbol{\theta}|\mathbf{y}) \propto f(\mathbf{y}|\boldsymbol{\theta})g(\boldsymbol{\theta}) \propto \exp(-Q/2)$, where

$$\begin{aligned}
Q &= \sigma^{-2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top \mathbf{X}^\top \mathbf{X}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \mathbf{m}_0)^\top \mathbf{W}_0^{-1}(\boldsymbol{\theta} - \mathbf{m}_0) \\
&= \boldsymbol{\theta}^\top (\mathbf{W}_0^{-1} + \sigma^{-2}\mathbf{X}^\top \mathbf{X})\boldsymbol{\theta} + \boldsymbol{\theta}^\top (\mathbf{W}_0^{-1}\mathbf{m}_0 + \sigma^{-2}\mathbf{X}^\top \mathbf{y}) + (\mathbf{W}_0^{-1}\mathbf{m}_0 + \sigma^{-2}\mathbf{X}^\top \mathbf{y})\boldsymbol{\theta} + R_1 \\
&= (\boldsymbol{\theta} - \mathbf{m}^*)^\top (\mathbf{W}^*)^{-1}(\boldsymbol{\theta} - \mathbf{m}^*) + R_2,
\end{aligned}$$

in which

$$\mathbf{m}^* = (\mathbf{W}_0^{-1} + \sigma^{-2}\mathbf{X}^\top \mathbf{X})^{-1}(\mathbf{W}_0^{-1}\mathbf{m}_0 + \sigma^{-2}\mathbf{X}^\top \mathbf{y}), \tag{4.34}$$

$$\mathbf{W}^* = (\mathbf{W}_0^{-1} + \sigma^{-2}\mathbf{X}^\top \mathbf{X})^{-1} \tag{4.35}$$

and R_1, R_2 are constants. Therefore,

$$f(\boldsymbol{\theta}|\mathbf{y}) \propto \exp \left\{ - \frac{(\boldsymbol{\theta} - \mathbf{m}^*)^\top \mathbf{W}^{*-1}(\boldsymbol{\theta} - \mathbf{m}^*)}{2} \right\}, \tag{4.36}$$

i.e. the posterior distribution of $\boldsymbol{\theta}$ is $N_p(\mathbf{m}^*, \mathbf{W}^*)$.

4.2.1 Prior predictive densities in regression coefficients hypothesis testing

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$. If $\boldsymbol{\theta}_1$ has s elements and $\boldsymbol{\theta}_2$ has r elements, write

$$\mathbf{m}_0 = \begin{bmatrix} \mathbf{m}_{01} \\ \mathbf{m}_{02} \end{bmatrix}, \quad \mathbf{W}_0 = \begin{bmatrix} \mathbf{W}_{011} & \mathbf{W}_{012} \\ \mathbf{W}_{021} & \mathbf{W}_{022} \end{bmatrix},$$

where \mathbf{m}_{01} is $s \times 1$, \mathbf{W}_{011} is $s \times s$, \mathbf{m}_{02} is $r \times 1$, \mathbf{W}_{022} is $r \times r$. Then

$$\boldsymbol{\theta}_1 \sim N_s(\mathbf{m}_{01}, \mathbf{W}_{011}), \quad (4.37)$$

$$\boldsymbol{\theta}_2 \sim N_r(\mathbf{m}_{02}, \mathbf{W}_{022}), \quad (4.38)$$

Using general results on multivariate normal distributions,

$$\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2 \sim N_s(\mathbf{m}_{01.2}(\boldsymbol{\theta}_2), \mathbf{W}_{011.2}), \quad (4.39)$$

where $\mathbf{m}_{01.2}(\boldsymbol{\theta}_2) = \mathbf{m}_{01} + \mathbf{W}_{012} \mathbf{W}_{022}^{-1}(\boldsymbol{\theta}_2 - \mathbf{m}_{02})$ and $\mathbf{W}_{011.2} = \mathbf{W}_{011} - \mathbf{W}_{012} \mathbf{W}_{022}^{-1} \mathbf{W}_{021}$.

A corresponding distribution results if we change \mathbf{m}_0 to \mathbf{m}^* and \mathbf{W}_0 to \mathbf{W}^* .

We are interested in testing the hypotheses

$$\mathbf{H} : \boldsymbol{\theta}_2 = \mathbf{0}$$

$$\mathbf{A} : \boldsymbol{\theta}_2 \neq \mathbf{0}.$$

The prior predictive density under \mathbf{H} is

$$\begin{aligned} f_{\mathbf{H}}(\mathbf{y}) &= \int_{\mathbf{H}} f(\mathbf{y} | \boldsymbol{\theta}) dP_{\mathbf{H}}(\boldsymbol{\theta}) \\ &= \int_{\mathbf{H}} f(\mathbf{y} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) g_{\mathbf{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 \end{aligned}$$

where $g_{\mathbf{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is the prior density under the subset of smaller dimensionality \mathbf{H} calculated

as

$$\begin{aligned}
g_{\mathbf{H}}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) &= g_{\mathbf{H}}(\boldsymbol{\theta}_1) = \frac{g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \mathbb{1}(\boldsymbol{\theta}_2 = \mathbf{0})}{\int_{\mathbf{H}} g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2} \\
&= \frac{g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \mathbb{1}(\boldsymbol{\theta}_2 = \mathbf{0})}{\int_{\boldsymbol{\theta}_1 \in \mathbb{R}^s} g(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}) d\boldsymbol{\theta}_1} \\
&= g(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2 = \mathbf{0}).
\end{aligned}$$

Thus, $f_{\mathbf{H}}(\mathbf{y})$ is given by

$$\begin{aligned}
f_{\mathbf{H}}(\mathbf{y}) &= \int_{\mathbf{H}} f(\mathbf{y} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) g(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2 = \mathbf{0}) d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_2 \\
&= \int_{\boldsymbol{\theta}_1 \in \mathbb{R}^s} f(\mathbf{y} | \boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}) g(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2 = \mathbf{0}) d\boldsymbol{\theta}_1 \\
&= \int_{\boldsymbol{\theta}_1 \in \mathbb{R}^s} N_n(\mathbf{X}\mathbf{C}\boldsymbol{\theta}_1, \sigma^2 \mathbf{I}_n) \times N_s(\mathbf{m}_{01.2}(\mathbf{0}), \mathbf{W}_{011.2}) d\boldsymbol{\theta}_1 \\
&= N_n\left(\mathbf{X}\mathbf{C}\mathbf{m}_{01.2}(\mathbf{0}), \left(\sigma^2 \mathbf{I}_n + (\mathbf{X}\mathbf{C})\mathbf{W}_{011.2}(\mathbf{X}\mathbf{C})^\top\right)\right). \tag{4.40}
\end{aligned}$$

where $\mathbf{C}_{(s+r) \times s} = [\mathbf{I}_s, \mathbf{0}_{s \times r}]^\top$.

The prior predictive density under \mathbf{A} can be obtained as follows

$$\begin{aligned}
f_{\mathbf{A}}(\mathbf{y}) &= \int_{\mathbf{A}} f(\mathbf{y} | \boldsymbol{\theta}) dP_{\mathbf{A}}(\boldsymbol{\theta}) \\
&= \int_{\mathbf{A}} f(\mathbf{y} | \boldsymbol{\theta}) g_{\mathbf{A}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int_{\mathbf{A}} f(\mathbf{y} | \boldsymbol{\theta}) g(\boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \int_{\mathbf{A}} N_n(\mathbf{X}\boldsymbol{\theta}, \sigma^2 \mathbf{I}_n) \times N_p(\mathbf{m}_0, \mathbf{W}_0) d\boldsymbol{\theta} \\
&= N_n\left(\mathbf{X}\mathbf{m}_0, \left(\sigma^2 \mathbf{I}_n + \mathbf{X}\mathbf{W}_0\mathbf{X}^\top\right)\right). \tag{4.41}
\end{aligned}$$

When $\mathbf{X} = \mathbf{1}_n$ ($d = 1$) and considering the regression model in (4.29), the hypotheses of interest are

$$\mathbf{H} : \theta_1 = 0$$

$$\mathbf{A} : \theta_1 \neq 0.$$

Therefore, the prior predictive density under \mathbf{H} is

$$\begin{aligned}
 f_{\mathbf{H}}(\mathbf{y}) &= \int_{\mathbf{H}} f(\mathbf{y}|\theta_1) dP_{\mathbf{H}}(\theta_1) \\
 &= \int_{\mathbf{H}} f(\mathbf{y}|\theta_1) g_{\mathbf{H}}(\theta_1) d\theta_1 \\
 &= f(\mathbf{y}|\theta_1 = 0)
 \end{aligned} \tag{4.42}$$

and here the prior predictive density under \mathbf{A} will be

$$\begin{aligned}
 f_{\mathbf{A}}(\mathbf{y}) &= \int_{\mathbf{A}} f(\mathbf{y}|\theta_1) dP_{\mathbf{A}}(\theta_1) \\
 &= \int_{\mathbf{A}} f(\mathbf{y}|\theta_1) g_{\mathbf{A}}(\theta_1) d\theta_1 \\
 &= \int_{\mathbf{A}} f(\mathbf{y}|\theta_1) g(\theta_1) d\theta_1 \\
 &= \int_{\mathbf{A}} N_n(\mathbf{1}_n\theta_1, \sigma^2\mathbf{I}_n) \times N(\mathbf{m}_0, \mathbf{W}_0) d\theta_1 \\
 &= N_n(\mathbf{1}_n\mathbf{m}_0, (\sigma^2\mathbf{I}_n + \mathbf{1}_n\mathbf{W}_0\mathbf{1}_n^\top)).
 \end{aligned} \tag{4.43}$$

4.3 Adaptive significance levels in regression coefficients hypothesis testing

4.3.1 Evidence index: e -value

As stated in Definition 5, to calculate the evidence in the FBST, first, we need to perform an optimization process, where the maximum of the posterior density restricted to the subset of the parameter space that represents the null hypothesis \mathbf{H} is obtained. This maximum value is used to find the tangential set to the null hypothesis, formed by the points of the parameter space with posterior density greater than the maximum over \mathbf{H} . Below, the procedure to calculate the tangential set and the evidence for the two prior formulation considered is presented.

4.3.1.1 Unknown-variance model

The tangential set to the null hypothesis \mathbf{H} is defined as

$$T_{\mathbf{y}_0} = \left\{ \boldsymbol{\xi} \in \Xi : f(\boldsymbol{\xi}|\mathbf{y}_0) > \sup_{\mathbf{H}} f(\boldsymbol{\xi}|\mathbf{y}_0) \right\} = \left\{ (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \in \Xi : f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2|\mathbf{y}_0) > \sup_{\mathbf{H}} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2|\mathbf{y}_0) \right\}.$$

Knowing that the posterior distribution of $(\boldsymbol{\theta}_1, \sigma^2)$ given $\boldsymbol{\theta}_2$ is an s -variate Normal-Inverse Gamma, as stated in (4.21), the point under \mathbf{H} for which the posterior attains its maximum value can be calculated as follows:

$$\begin{aligned} \arg \sup_{\mathbf{H}} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2|\mathbf{y}_0) &= \arg \sup_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2|\mathbf{y}_0) \\ &= \arg \sup_{\boldsymbol{\theta}_1, \sigma^2} \frac{f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2|\mathbf{y}_0)}{\int_{\boldsymbol{\theta}_1 \in \mathbb{R}^s, \sigma^2 \in \mathbb{R}_+} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}, \sigma^2|\mathbf{y}_0) d\boldsymbol{\theta}_1 d\sigma^2} \\ &= \arg \sup_{\boldsymbol{\theta}_1, \sigma^2} f(\boldsymbol{\theta}_1, \sigma^2|\boldsymbol{\theta}_2 = \mathbf{0}, \mathbf{y}_0) \\ &= \text{Mode} [f(\boldsymbol{\theta}_1, \sigma^2|\boldsymbol{\theta}_2 = \mathbf{0}, \mathbf{y}_0)] \\ &= \left(\mathbf{m}^*_{1.2}(\mathbf{0}), \mathbf{0}, \frac{b_1 + \frac{(\mathbf{m}^*_2)^\top (\mathbf{V}^*_{22})^{-1} (\mathbf{m}^*_2)}{2}}{\left(a_1 + \frac{r}{2}\right) + 1 + \frac{s}{2}} \right) \\ &= (\hat{\boldsymbol{\theta}}_1, \mathbf{0}, \hat{\sigma}^2). \end{aligned}$$

Thus, we get the tangential set

$$T_{\mathbf{y}_0} = \left\{ (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2) \in \Xi : f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \sigma^2|\mathbf{y}_0) > f(\hat{\boldsymbol{\theta}}_1, \mathbf{0}, \hat{\sigma}^2|\mathbf{y}_0) \right\}.$$

In the case of the model in (4.29), the tangential set to the null hypothesis can be expressed as

$$T_{\mathbf{y}_0} = \left\{ \boldsymbol{\xi} \in \Xi : f(\boldsymbol{\xi}|\mathbf{y}_0) > \sup_{\mathbf{H}} f(\boldsymbol{\xi}|\mathbf{y}_0) \right\} = \left\{ (\theta_1, \sigma^2) \in \Xi : f(\theta_1, \sigma^2|\mathbf{y}_0) > \sup_{\mathbf{H}} f(\theta_1, \sigma^2|\mathbf{y}_0) \right\}.$$

Here, the posterior distribution of σ^2 given $\boldsymbol{\theta}$ is an Inverse Gamma, as stated in (4.14).

Then, the point under \mathbf{H} for which the posterior attains its maximum is obtained as follows

$$\begin{aligned}
\arg \sup_{\mathbf{H}} f(\theta_1, \sigma^2 | \mathbf{y}_0) &= \arg \sup_{\theta_1=0, \sigma^2} f(\theta_1 = 0, \sigma^2 | \mathbf{y}_0) \\
&= \arg \sup_{\sigma^2} \frac{f(\theta_1 = 0, \sigma^2 | \mathbf{y}_0)}{\int_{\sigma^2 \in \mathbb{R}_+} f(\theta_1 = 0, \sigma^2 | \mathbf{y}_0) d\sigma^2} \\
&= \arg \sup_{\sigma^2} f(\sigma^2 | \theta_1 = 0, \mathbf{y}_0) \\
&= \text{Mode} [f(\sigma^2 | \theta_1 = 0, \mathbf{y}_0)] \\
&= \left(0, \frac{b_1 + \frac{(\mathbf{m}^*)^2 (\mathbf{V}^*)^{-1}}{2}}{\left(a_1 + \frac{1}{2}\right) + 1} \right) \\
&= (0, \hat{\sigma}^2).
\end{aligned}$$

Therefore, we get

$$T_{\mathbf{y}_0} = \left\{ (\theta_1, \sigma^2) \in \Xi : f(\theta_1, \sigma^2 | \mathbf{y}_0) > f(0, \hat{\sigma}^2 | \mathbf{y}_0) \right\}. \quad (4.44)$$

For the unknown-variance model, once the tangential set is defined, we perform an integration as a second step, where the evidence in favor \mathbf{H} is calculated as the complement of the posterior probability of $T_{\mathbf{y}_0}$. That is,

$$ev(\mathbf{H}; \mathbf{y}_0) = 1 - P(\boldsymbol{\xi} \in T_{\mathbf{y}_0} | \mathbf{y}_0).$$

Generating M samples from the posterior distribution of $\boldsymbol{\xi}$, say $\boldsymbol{\xi}^{(1)}, \dots, \boldsymbol{\xi}^{(M)}$, specified by (4.12), we estimate the evidence by Monte Carlo simulation through the expression

$$1 - \frac{1}{M} \sum_{j=1}^M \mathbb{1}(\boldsymbol{\xi}^{(j)} \in T_{\mathbf{y}_0}).$$

4.3.1.2 Known-variance model

The tangential set to the null hypothesis is given by

$$T_{\mathbf{y}_0} = \left\{ \boldsymbol{\theta} \in \Theta : f(\boldsymbol{\theta}|\mathbf{y}_0) > \sup_{\mathbf{H}} f(\boldsymbol{\theta}|\mathbf{y}_0) \right\} = \left\{ (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \Theta : f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}_0) > \sup_{\mathbf{H}} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}_0) \right\}.$$

In this case we know that the posterior distribution of $\boldsymbol{\theta}_1$ given $\boldsymbol{\theta}_2$ is an s -variate Normal distribution, as stated in (4.39). So, the point under \mathbf{H} for which the posterior attains its maximum value will be

$$\begin{aligned} \arg \sup_{\mathbf{H}} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}_0) &= \arg \sup_{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}|\mathbf{y}_0) \\ &= \arg \sup_{\boldsymbol{\theta}_1} \frac{f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}|\mathbf{y}_0)}{\int_{\boldsymbol{\theta}_1 \in \mathbb{R}^s} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 = \mathbf{0}|\mathbf{y}_0) d\boldsymbol{\theta}_1} \\ &= \arg \sup_{\boldsymbol{\theta}_1} f(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2 = \mathbf{0}, \mathbf{y}_0) \\ &= \text{Mode} [f(\boldsymbol{\theta}_1|\boldsymbol{\theta}_2 = \mathbf{0}, \mathbf{y}_0)] \\ &= (\mathbf{m}^*_{1.2}(\boldsymbol{\theta}_2 = \mathbf{0}), \mathbf{0}) \\ &= (\hat{\boldsymbol{\theta}}_1, \mathbf{0}). \end{aligned}$$

Then, we get the tangential set

$$T_{\mathbf{y}_0} = \left\{ (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \in \Theta : f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}_0) > f(\hat{\boldsymbol{\theta}}_1, \mathbf{0}|\mathbf{y}_0) \right\}.$$

From this definition for $T_{\mathbf{y}_0}$, we have that

$$\begin{aligned} f(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2|\mathbf{y}_0) &> f(\hat{\boldsymbol{\theta}}_1, \mathbf{0}|\mathbf{y}_0) \\ \frac{(2\pi)^{-p/2}}{|\mathbf{W}^*|^{1/2}} \exp \left\{ -\frac{(\boldsymbol{\theta} - \mathbf{m}^*)^\top \mathbf{W}^{*-1}(\boldsymbol{\theta} - \mathbf{m}^*)}{2} \right\} &> f(\hat{\boldsymbol{\theta}}_1, \mathbf{0}|\mathbf{y}_0) \\ (\boldsymbol{\theta} - \mathbf{m}^*)^\top \mathbf{W}^{*-1}(\boldsymbol{\theta} - \mathbf{m}^*) &< -2 \log \left\{ f(\hat{\boldsymbol{\theta}}_1, \mathbf{0}|\mathbf{y}_0) |\mathbf{W}^*|^{1/2} (2\pi)^{p/2} \right\}, \end{aligned}$$

and since $(\boldsymbol{\theta} - \mathbf{m}^*)^\top \mathbf{W}^{*-1}(\boldsymbol{\theta} - \mathbf{m}^*) \sim \chi_p^2$ (see [Mardia *et al.* \(1979\)](#)), the evidence in favor

of \mathbf{H} can be calculated directly as

$$\begin{aligned} ev(\mathbf{H}; \mathbf{y}_0) &= 1 - P(\boldsymbol{\theta} \in T_{\mathbf{y}_0} | \mathbf{y}_0) \\ &= 1 - P\left(\chi_p^2 < -2 \log \left\{ f(\hat{\boldsymbol{\theta}}_1, \mathbf{0} | \mathbf{y}_0) |\mathbf{W}^*|^{1/2} (2\pi)^{p/2} \right\}\right). \end{aligned}$$

When we have the model in (4.29), the tangential set takes the form

$$T_{\mathbf{y}_0} = \left\{ \theta_1 \in \Theta : f(\theta_1 | \mathbf{y}_0) > \sup_{\mathbf{H}} f(\theta_1 | \mathbf{y}_0) \right\} = \{ \theta \in \Theta : f(\theta_1 | \mathbf{y}_0) > f(\theta_1 = 0 | \mathbf{y}_0) \},$$

and it follows that

$$\begin{aligned} & f(\theta_1 | \mathbf{y}_0) > f(\theta_1 = 0 | \mathbf{y}_0) \\ \frac{(2\pi)^{-1/2}}{W^{*1/2}} \exp \left\{ -\frac{(\theta_1 - m^*)^\top W^{*-1} (\theta_1 - m^*)}{2} \right\} &> f(\theta_1 = 0 | \mathbf{y}_0) \\ & (\theta_1 - m^*)^\top W^{*-1} (\theta_1 - m^*) < -2 \log \left\{ f(\theta_1 = 0 | \mathbf{y}_0) W^{*1/2} (2\pi)^{p/2} \right\}. \end{aligned}$$

Since $(\theta_1 - m^*)^\top W^{*-1} (\theta_1 - m^*) \sim \chi_1^2$, the evidence in favor of \mathbf{H} is

$$\begin{aligned} ev(\mathbf{H}; \mathbf{y}_0) &= 1 - P(\theta_1 \in T_{\mathbf{y}_0} | \mathbf{y}_0) \\ &= 1 - P\left(\chi_1^2 < -2 \log \left\{ f(\theta_1 = 0 | \mathbf{y}_0) W^{*1/2} (2\pi)^{1/2} \right\}\right). \end{aligned}$$

Now, in order to calculate the averaged error probabilities for the unknown variance model and the known variance model, consider the test such that

$$\varphi_e(\mathbf{y}) = \begin{cases} 0 & \text{if } ev(\mathbf{H}; \mathbf{y}) > k \\ 1 & \text{if } ev(\mathbf{H}; \mathbf{y}) \leq k. \end{cases}$$

The averaged error probabilities, expressed in terms of the predictive densities defined in (4.27) and (4.28) (or by (4.30) and (4.31) when $\mathbf{X} = \mathbf{1}_n$) in the unknown variance case, and also in (4.40) and (4.41) (or by (4.42) and (4.43) when $\mathbf{X} = \mathbf{1}_n$) in the known variance

case, can be estimated by Monte Carlo simulation through the expressions

$$\alpha_{\varphi_e} = \int_{\mathbf{y} \in \Psi_e} f_{\mathbf{H}}(\mathbf{y}) d\mathbf{y} \quad \text{and} \quad \beta_{\varphi_e} = \int_{\mathbf{y} \notin \Psi_e} f_{\mathbf{A}}(\mathbf{y}) d\mathbf{y}.$$

where Ψ_e is the set

$$\Psi_e = \{\mathbf{y} \in \Omega : ev(\mathbf{H}; \mathbf{y}) \leq k\}.$$

So the adaptive cutoff value k^* for $ev(\mathbf{H}; \mathbf{y})$ will be the k that minimizes $a\alpha_{\varphi_e} + b\beta_{\varphi_e}$.

Finally, consider the test

$$\varphi_e^*(\mathbf{y}) = \begin{cases} 0 & \text{if } ev(\mathbf{H}; \mathbf{y}) > k^* \\ 1 & \text{if } ev(\mathbf{H}; \mathbf{y}) \leq k^*. \end{cases}$$

Then the optimal averaged error probabilities that depend on the sample size will be

$$\alpha_{\varphi_e^*}^* = \int_{\mathbf{y} \in \Psi_e^*} f_{\mathbf{H}}(\mathbf{y}) d\mathbf{y} \quad \text{and} \quad \beta_{\varphi_e^*}^* = \int_{\mathbf{y} \notin \Psi_e^*} f_{\mathbf{A}}(\mathbf{y}) d\mathbf{y}.$$

where Ψ_e^* is the set defined by

$$\Psi_e^* = \{\mathbf{y} \in \Omega : ev(\mathbf{H}; \mathbf{y}) \leq k^*\}.$$

4.3.2 Significance index: P -value

This procedure is extended to both previous models only changing the respective prior predictive densities. Let $\varphi_P^*(\mathbf{y})$ be the test given by

$$\varphi_P^*(\mathbf{y}) = \begin{cases} 0 & \text{if } BF(\mathbf{y}) > \frac{b}{a} \\ 1 & \text{if } BF(\mathbf{y}) \leq \frac{b}{a}. \end{cases}$$

To obtain the P -value, define the set Ψ_P of sample points \mathbf{y} for which the Bayes factors

are smaller than or equal to the Bayes factor of the observed sample point \mathbf{y}_0 , that is

$$\Psi_P = \{\mathbf{y} \in \Omega : \text{BF}(\mathbf{y}) \leq \text{BF}(\mathbf{y}_0)\}.$$

Then, the P -value is the integral of the predictive density under \mathbf{H} in Ψ_P

$$P\text{-value}(\mathbf{y}_0) = \int_{\Psi_P} f_{\mathbf{H}}(\mathbf{y}) d\mathbf{y}.$$

Defining the set Ψ_P^* of sample points \mathbf{y} with Bayes factors smaller than or equal to b/a , i.e.,

$$\Psi_P^* = \left\{ \mathbf{y} \in \Omega : \text{BF}(\mathbf{y}) \leq \frac{b}{a} \right\},$$

the optimal averaged error probabilities, that can be approximated by Monte Carlo simulation, are given by

$$\alpha_{\varphi_P^*} = \int_{\mathbf{y} \in \Psi_P^*} f_{\mathbf{H}}(\mathbf{y}) d\mathbf{y} \quad \text{and} \quad \beta_{\varphi_P^*} = \int_{\mathbf{y} \notin \Psi_P^*} f_{\mathbf{A}}(\mathbf{y}) d\mathbf{y}.$$

4.4 Simulation study

We developed a simulation study considering two models, both under the two formulations of conjugate prior distributions described in the previous sections. The first model is

$$\mathbf{y} = \mathbf{X}\theta + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad (4.45)$$

where $\mathbf{X} = \mathbf{1}_n$ and $\theta = \theta_1$. For the unknown-variance model, we selected the parameters a_0 and b_0 of the Inverse Gamma prior distribution for σ^2 such that the resulting distribution had a mean equal to 1 and a variance equal to 100, those values being $a_0 = 2.01$ and $b_0 = 1.01$. Next, we specified the parameters in the Normal prior for θ given σ^2 as $m_0 = 0$ and $V_0 = 1$. On the other hand, in the known variance case, we supposed that the residual error variance was $\sigma^2 = 1$ and the parameters in the Normal prior for θ were $m_0 = 0$ and $W_0 = 1$. The

hypotheses to be tested were

$$\mathbf{H} : \theta_1 = 0$$

$$\mathbf{A} : \theta_1 \neq 0.$$

The second model studied was

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n), \quad (4.46)$$

where $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^\top$ is an $n \times p$ matrix of covariates with $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip-1})^\top$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top)^\top$ is the $p \times 1$ vector of coefficients. In the unknown-variance analysis, the parameters of the Inverse Gamma distribution were selected in the same way as in model (4.45). Additionally, in the Normal prior for $\boldsymbol{\theta}$ given σ^2 , $\mathbf{m}_0 = (\mathbf{0}^\top, \mathbf{0}^\top)^\top$ and $\mathbf{V}_0 = \mathbf{I}_p$ were taken as parameters, assuming that $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ could be treated as independent. Finally, in the known-variance formulation, we supposed that the residual error variance was $\sigma^2 = 1$ and the parameters in the Normal prior for $\boldsymbol{\theta}$ were $\mathbf{m}_0 = (\mathbf{0}^\top, \mathbf{0}^\top)^\top$ and $\mathbf{W}_0 = \mathbf{I}_p$, being $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ also treated as independent. This time, the hypotheses of interest were

$$\mathbf{H} : \boldsymbol{\theta}_2 = \mathbf{0}$$

$$\mathbf{A} : \boldsymbol{\theta}_2 \neq \mathbf{0}.$$

The averaged error probabilities α_{φ_e} , β_{φ_e} , $\alpha_{\varphi_e^*}$, $\beta_{\varphi_e^*}$, $\alpha_{\varphi_p^*}$ and $\beta_{\varphi_p^*}$, were calculated using the Monte Carlo method with values generated from the following distributions:

- **Unknown-variance analysis**

- **Model (4.45) under \mathbf{H}**

$$\begin{aligned} \theta_1^{(j)} &= 0 \\ \sigma^{2(j)} | \theta_1^{(j)} = 0 &\sim IG \left(a_0 + \frac{1}{2}, b_0 + \frac{(\theta_1^{(j)} - \mathbf{m}_0)^\top \mathbf{V}_0^{-1} (\theta_1^{(j)} - \mathbf{m}_0)}{2} \right) \\ \mathbf{Y}^{(j)} | \sigma^{2(j)}, \theta_1^{(j)} &\sim N_n(\mathbf{1}_n \theta_1^{(j)}, \sigma^{2(j)} \mathbf{I}_n). \end{aligned}$$

– Model (4.45) under A

$$\begin{aligned}\sigma^{2(j)} &\sim IG(a_0, b_0) \\ \theta_1^{(j)} | \sigma^{2(j)} &\sim N(\mathbf{m}_0, \sigma^{2(j)} \mathbf{V}_0) \\ \mathbf{Y}^{(j)} | \sigma^{2(j)}, \theta_1^{(j)} &\sim N_n(\mathbf{1}_n \theta_1^{(j)}, \sigma^{2(j)} \mathbf{I}_n).\end{aligned}$$

– Model (4.46) under H

$$\begin{aligned}\theta_2^{(j)} &= \mathbf{0} \\ \theta_1^{(j)} | \theta_2^{(j)} = \mathbf{0} &\sim t_s \left(2a_0 + 1; \mathbf{m}_{01.2}(\theta_2^{(j)}), \frac{2b_0 + (\theta_2^{(j)} - \mathbf{m}_{02})^\top \mathbf{V}_{022}^{-1} (\theta_2^{(j)} - \mathbf{m}_{02})}{2a_0 + 1} \mathbf{V}_{011.2} \right) \\ \sigma^{2(j)} | \theta_1^{(j)}, \theta_2^{(j)} = \mathbf{0} &\sim IG \left(a_0 + 1, b_0 + \frac{(\theta^{(j)} - \mathbf{m}_0)^\top \mathbf{V}_0^{-1} (\theta^{(j)} - \mathbf{m}_0)}{2} \right) \\ \mathbf{Y}^{(j)} | \sigma^{2(j)}, \theta_1^{(j)}, \theta_2^{(j)} = \mathbf{0} &\sim N_n(\mathbf{X}\theta^{(j)}, \sigma^{2(j)} \mathbf{I}_n).\end{aligned}$$

– Model (4.46) under A

$$\begin{aligned}\sigma^{2(j)} &\sim IG(a_0, b_0) \\ \theta^{(j)} | \sigma^{2(j)} &\sim N_p(\mathbf{m}_0, \sigma^{2(j)} \mathbf{V}_0) \\ \mathbf{Y}^{(j)} | \sigma^{2(j)}, \theta^{(j)} &\sim N_n(\mathbf{X}\theta^{(j)}, \sigma^{2(j)} \mathbf{I}_n).\end{aligned}$$

• Known-variance analysis

– Model (4.45) under H

$$\begin{aligned}\sigma^{2(j)} &= 1 \\ \theta_1^{(j)} &= 0 \\ \mathbf{Y}^{(j)} | \theta_1^{(j)} = 0 &\sim N_n(\mathbf{1}_n \theta_1^{(j)}, \sigma^{2(j)} \mathbf{I}_n).\end{aligned}$$

– Model (4.45) under A

$$\begin{aligned}\sigma^{2(j)} &= 1 \\ \theta_1^{(j)} &\sim N(\mathbf{m}_0, W_0) \\ \mathbf{Y}^{(j)} | \theta_1^{(j)} &\sim N_n(\mathbf{1}_n \theta_1^{(j)}, \sigma^{2(j)} \mathbf{I}_n).\end{aligned}$$

– Model (4.46) under H

– Model (4.46) under A

$$\sigma^{2(j)} = 1$$

$$\boldsymbol{\theta}_2^{(j)} = \mathbf{0}$$

$$\boldsymbol{\theta}_1^{(j)} | \boldsymbol{\theta}_2^{(j)} = \mathbf{0} \sim N_s \left(\mathbf{m}_{01.2}(\boldsymbol{\theta}_2^{(j)}), \mathbf{W}_{011.2} \right)$$

$$\mathbf{Y}^{(j)} | \boldsymbol{\theta}_1^{(j)}, \boldsymbol{\theta}_2^{(j)} = \mathbf{0} \sim N_n(\mathbf{X}\boldsymbol{\theta}^{(j)}, \sigma^{2(j)}\mathbf{I}_n).$$

$$\sigma^{2(j)} = 1$$

$$\boldsymbol{\theta}^{(j)} \sim N_p(\mathbf{m}_0, \mathbf{W}_0)$$

$$\mathbf{Y}^{(j)} | \boldsymbol{\theta}^{(j)} \sim N_n(\mathbf{X}\boldsymbol{\theta}^{(j)}, \sigma^{2(j)}\mathbf{I}_n).$$

Then, $\mathbf{y}^{(j)} = (y_1^{(j)}, \dots, y_n^{(j)})$ is a random sample of the conditional distribution of \mathbf{Y} , $j = 1 \dots M$.

In a first stage, we considered the model (4.45) where $\theta = \theta_1$ and the model (4.46) with $\boldsymbol{\theta} = (\theta_1, \theta_2)^\top$. Note that the dimensionality of the parameter space in the two models is different when the prior changes, so for model (4.45), the dimensionality will be $d = 2$ in the unknown-variance case and $d = 1$ in the known-variance case. For model (4.46) the dimensionality will be $d = 3$ in the unknown-variance case and $d = 2$ in the known-variance case. Samples of size $M = 1000$ were generated for each model under the respective hypotheses and also for different sample sizes between $n = 10$ and $n = 5000$. In model (4.46) the covariate x_{i1} , $i = 1 \dots n$, was generated from a standard normal distribution. The random generation of variables from the Inverse Gamma distribution and the Multivariate Normal and t distributions were made respectively through the packages `LaplaceDemon` (Statisticat and LLC., 2018) and `mnormt` (Azzalini and Genz, 2016) both available in R software (R Core Team, 2017). Finally, to obtain the adaptive values of k^* , $\alpha_{\varphi_P^*}$ and $\beta_{\varphi_P^*}$, the two types of errors were considered as equally severe, that is, $a = b = 1$.

Figures 4.1 and 4.2 show the averaged error probabilities for the FBST as functions of k for a sample size $n = 100$. This was replicated for various sample sizes in order to numerically find the corresponding k^* value that minimizes $\alpha_{\varphi_e} + \beta_{\varphi_e}$. Tables 4.1-4.4 and Figures 4.3-4.6 present the k^* and $\alpha_{\varphi_P^*}$ values as function of n for each model under the two formulations of conjugate prior distributions. As it can be seen, both values have a decreasing trend when the sample size increases. In the case of the cutoff value for the evidence, it is possible to notice the differences in the results when the dimensionality of the parameter space change. Then, the k^* value depends not only on the sample size but also on

the dimensionality of the parameter space, more specifically, it is greater when d is higher. However, this does not occur with $\alpha_{\varphi_P^*}$ which maintains almost the same values even if d increases. The results in the unknown-variance and the known-variance analysis do not differ considerably when the models have the same dimension, the k^* and $\alpha_{\varphi_P^*}$ values being a little higher in the unknown-variance case. On the other hand, Figures 4.7-4.10 illustrate that in all these models, the optimal averaged error probabilities and their linear combination also decrease with increasing sample size.

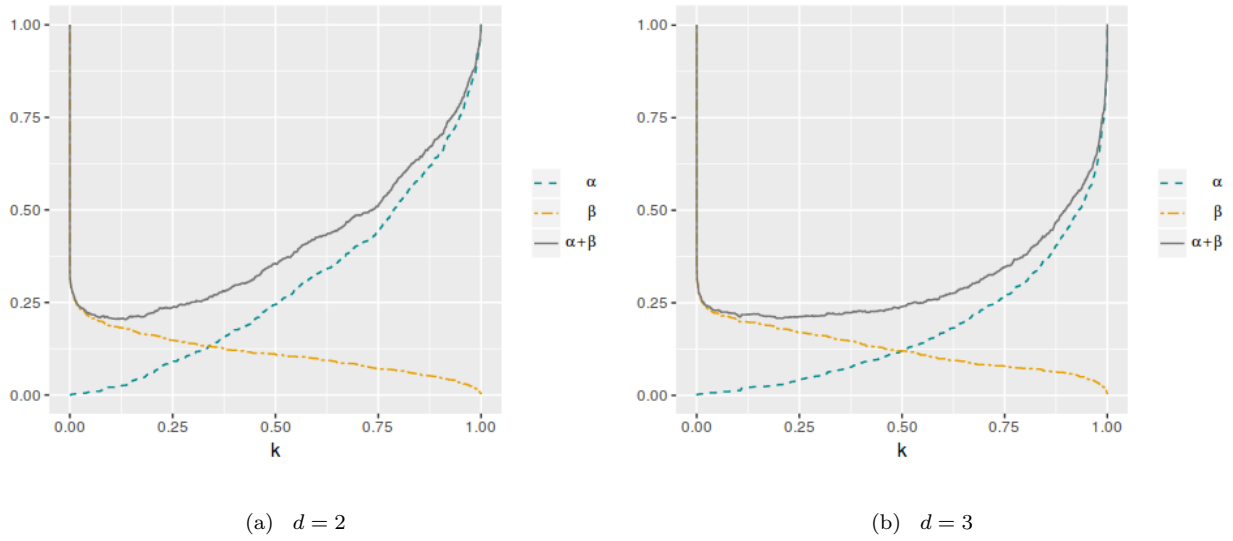


Figure 4.1: Unknown-variance model averaged error probabilities (α_{φ_e} , β_{φ_e} and $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k . Sample size $n = 100$.

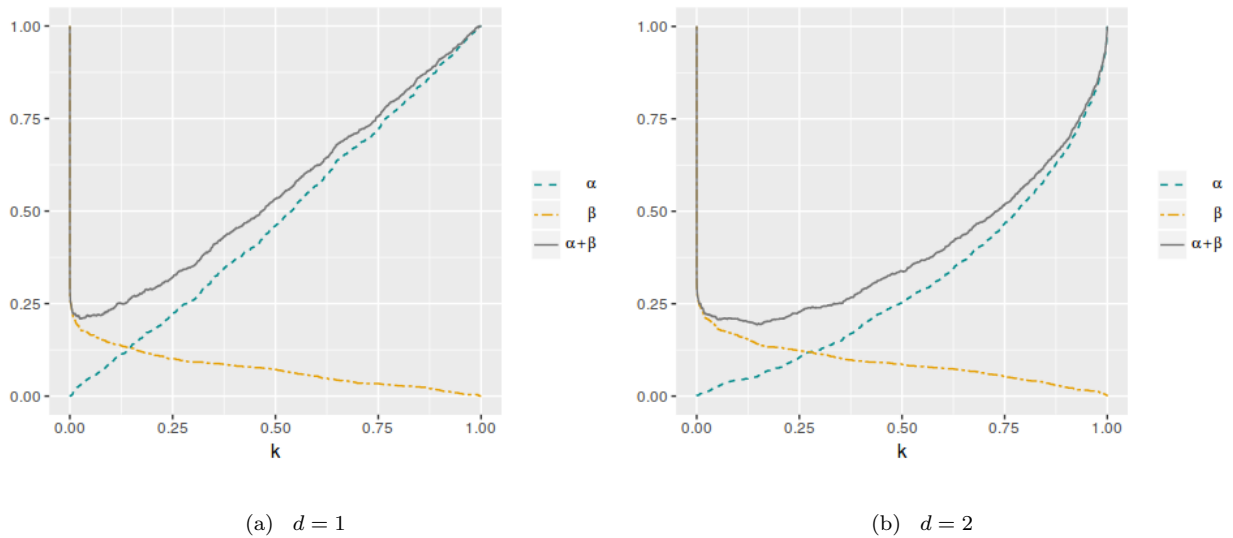


Figure 4.2: Known-variance model averaged error probabilities (α_{φ_e} , β_{φ_e} and $\alpha_{\varphi_e} + \beta_{\varphi_e}$) as function of k . Sample size $n = 100$.

n	k^*	
	$d = 2$	$d = 3$
10	0.32530	0.51220
50	0.12534	0.22442
100	0.11705	0.21081
150	0.10889	0.19735
200	0.10092	0.18416
250	0.09323	0.17132
300	0.08587	0.15894
350	0.07893	0.14713
400	0.07243	0.13598
450	0.06641	0.12560
500	0.06091	0.11606
1000	0.03035	0.06689
1500	0.02223	0.07086
2000	0.01892	0.07173

Table 4.1: Unknown-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of n , with $d = 2$ and $d = 3$.

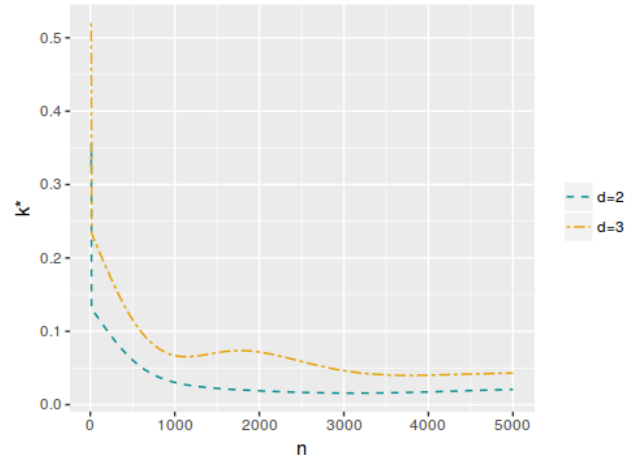


Figure 4.3: Unknown-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of n , with $d = 2$ and $d = 3$.

n	k^*	
	$d = 1$	$d = 2$
10	0.10040	0.35260
50	0.05166	0.11262
100	0.04447	0.10473
150	0.03776	0.09698
200	0.03179	0.08946
250	0.02667	0.08226
300	0.02244	0.07544
350	0.01905	0.06904
400	0.01639	0.06311
450	0.01429	0.05767
500	0.01264	0.05274
1000	0.00649	0.02823
1500	0.00622	0.02954
2000	0.00610	0.03000

Table 4.2: Known-variance cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of n , with $d = 1$ and $d = 2$.

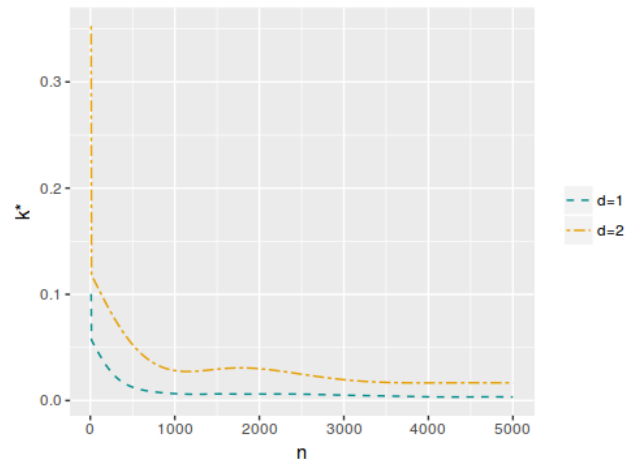


Figure 4.4: Known-variance cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of n , with $d = 1$ and $d = 2$.

n	$\alpha_{\varphi_P^*}$	
	$d = 2$	$d = 3$
10	0.12400	0.09200
50	0.04515	0.04327
100	0.03899	0.03775
150	0.03327	0.03252
200	0.02817	0.02772
250	0.02380	0.02341
300	0.02018	0.01963
350	0.01732	0.01642
400	0.01513	0.01376
450	0.01353	0.01163
500	0.01241	0.01002
1000	0.00941	0.00683
1500	0.00827	0.00398
2000	0.00681	0.00524

Table 4.3: Unknown-variance model optimal averaged type I error probability ($\alpha_{\varphi_P^*}$) as a function of n , with $d = 2$ and $d = 3$.

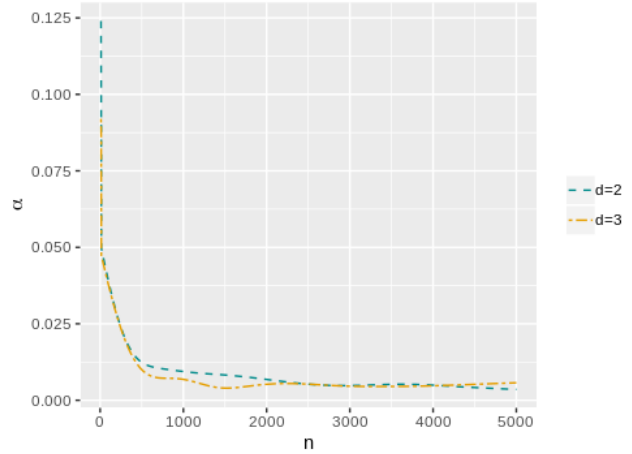


Figure 4.5: Unknown-variance model optimal averaged type I error probability ($\alpha_{\varphi_P^*}$) as a function of n , with $d = 2$ and $d = 3$.

n	$\alpha_{\varphi_P^*}$	
	$d = 1$	$d = 2$
10	0.10200	0.10100
50	0.04137	0.04136
100	0.03766	0.03683
150	0.03407	0.03248
200	0.03067	0.02841
250	0.02752	0.02469
300	0.02467	0.02135
350	0.02213	0.01844
400	0.01990	0.01595
450	0.01798	0.01388
500	0.01635	0.01221
1000	0.01011	0.00684
1500	0.00836	0.00421
2000	0.00606	0.00507

Table 4.4: Known-variance model optimal averaged type I error probability ($\alpha_{\varphi_P^*}$) as a function of n , with $d = 1$ and $d = 2$.

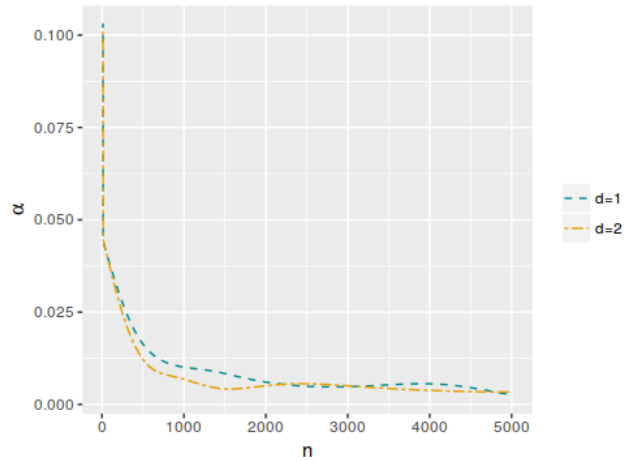
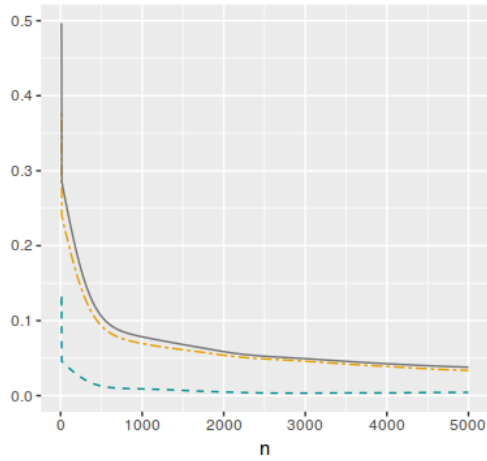
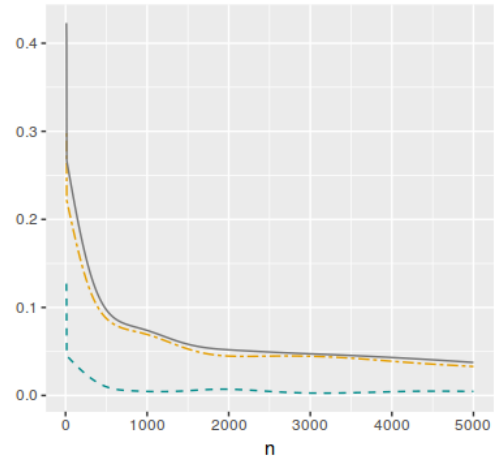


Figure 4.6: Known-variance model optimal averaged type I error probability ($\alpha_{\varphi_P^*}$) as a function of n , with $d = 1$ and $d = 2$.

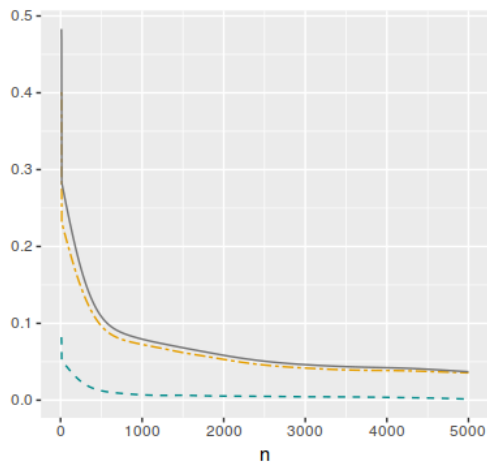


(a) $d = 2$

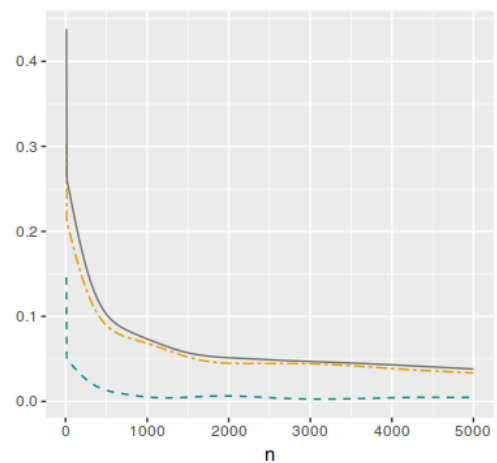


(b) $d = 3$

Figure 4.7: *Unknown-variance model optimal averaged error probabilities ($\alpha_{\varphi_e^*}^*$, $\beta_{\varphi_e^*}^*$ and $\alpha_{\varphi_e^*}^* + \beta_{\varphi_e^*}^*$) as functions of n .*



(a) $d = 1$



(b) $d = 2$

Figure 4.8: *Known-variance model optimal averaged error probabilities ($\alpha_{\varphi_e^*}^*$, $\beta_{\varphi_e^*}^*$ and $\alpha_{\varphi_e^*}^* + \beta_{\varphi_e^*}^*$) as functions of n .*

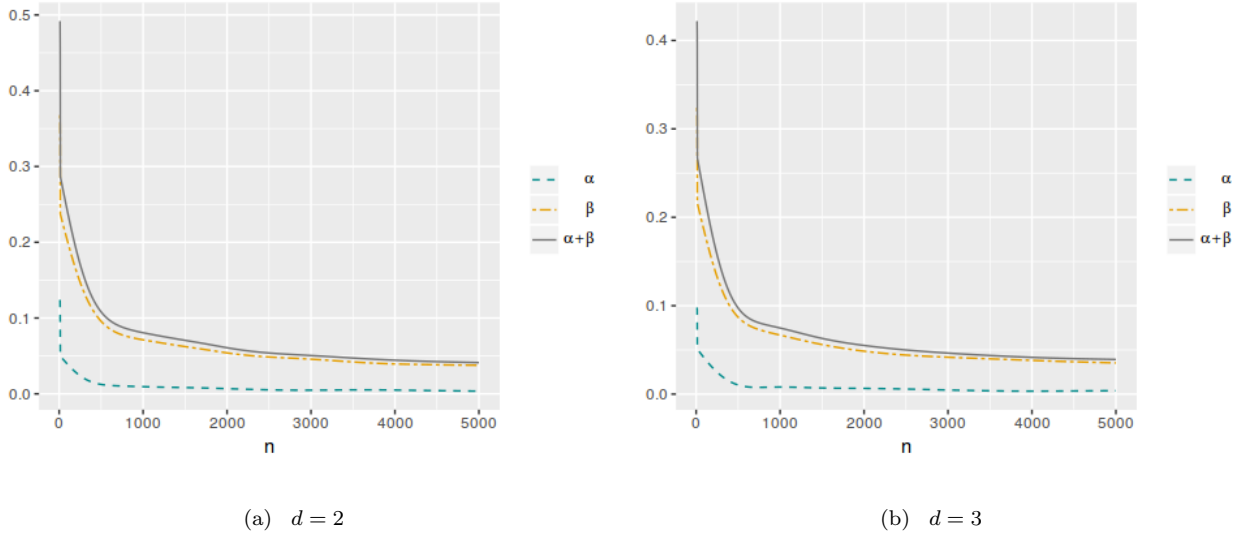


Figure 4.9: *Unknown-variance model optimal averaged error probabilities ($\alpha_{\varphi_P^*}$, $\beta_{\varphi_P^*}$ and $\alpha_{\varphi_P^*} + \beta_{\varphi_P^*}$) as functions of n .*

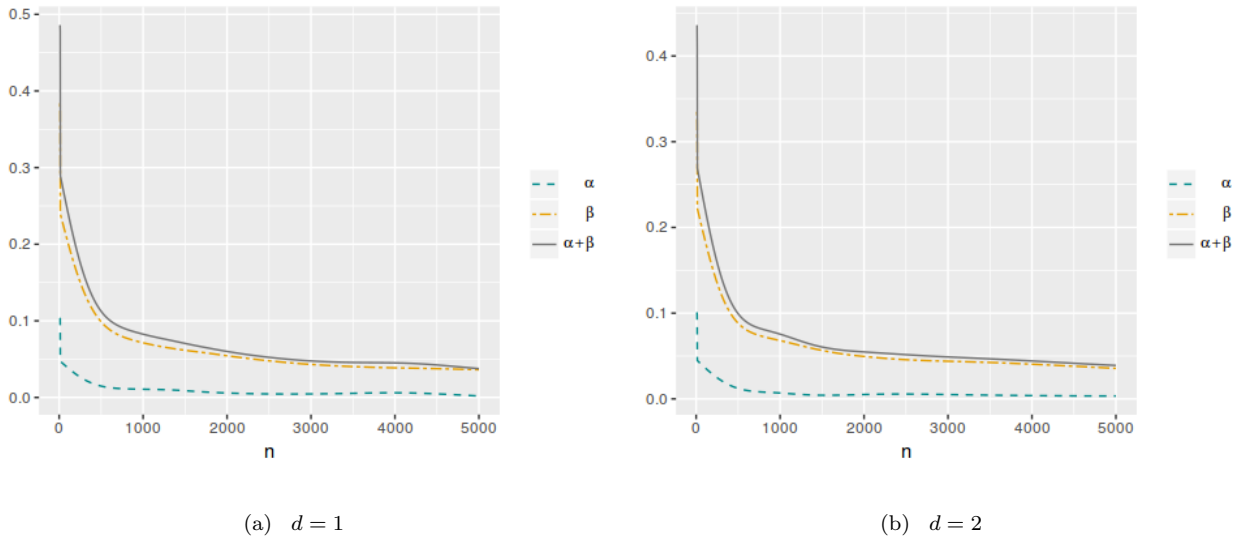


Figure 4.10: *Known-variance model optimal averaged error probabilities ($\alpha_{\varphi_P^*}$, $\beta_{\varphi_P^*}$ and $\alpha_{\varphi_P^*} + \beta_{\varphi_P^*}$) as functions of n .*

We choose a single random sample \mathbf{y}_0 to calculate the e -value and P -value for all models. Tables 4.5 and 4.6 display the results, the cases when \mathbf{H} is rejected being represented by the cells in **boldface**. It can be observed that the decision remains the same regardless of the index used. Besides, it is important to note that the null hypothesis is less easily rejected as the dimensionality of the parameter space increases and even more in the case of the known-variance formulation that we assign.

n	$d = 2$				$d = 3$			
	k^*	ev	$\alpha_{\varphi_P}^*$	Pv	k^*	ev	$\alpha_{\varphi_P}^*$	Pv
10	0.3253	0.9838	0.1240	0.7510	0.5122	0.9696	0.0920	0.4850
50	0.1253	0.0820	0.0451	0.0190	0.2244	0.9261	0.0433	0.3570
100	0.1171	0.0000	0.0390	0.0000	0.2108	0.4176	0.0377	0.0650
150	0.1089	0.0973	0.0333	0.0200	0.1974	0.2965	0.0325	0.0510
200	0.1009	0.0036	0.0282	0.0000	0.1842	0.0466	0.0277	0.0040
250	0.0932	0.0001	0.0238	0.0000	0.1713	0.0620	0.0234	0.0050
300	0.0859	0.0000	0.0202	0.0000	0.1589	0.0119	0.0196	0.0010
350	0.0789	0.0000	0.0173	0.0000	0.1471	0.0282	0.0164	0.0030
400	0.0724	0.0000	0.0151	0.0000	0.1360	0.0347	0.0138	0.0020
450	0.0664	0.0000	0.0135	0.0000	0.1256	0.0628	0.0116	0.0040
500	0.0609	0.0000	0.0124	0.0000	0.1161	0.0181	0.0100	0.0010
1000	0.0303	0.0000	0.0094	0.0000	0.0669	0.0000	0.0068	0.0010
1500	0.0222	0.0000	0.0083	0.0000	0.0709	0.0000	0.0040	0.0010
2000	0.0189	0.0000	0.0068	0.0000	0.0717	0.0000	0.0052	0.0010

Table 4.5: Unknown-variance model cutoff values k^* , $ev(\mathbf{H}; \mathbf{y}_0)$ and P -value(\mathbf{y}_0) as function of n , with $d = 2$ and $d = 3$.

n	$d = 1$				$d = 2$			
	k^*	ev	$\alpha_{\varphi_P}^*$	Pv	k^*	ev	$\alpha_{\varphi_P}^*$	Pv
10	0.1004	0.6785	0.1020	0.6740	0.3526	0.9647	0.1010	0.7710
50	0.0517	0.4918	0.0414	0.4900	0.1126	0.8561	0.0414	0.5850
100	0.0445	0.0749	0.0377	0.0630	0.1047	0.3659	0.0368	0.1530
150	0.0378	0.0254	0.0341	0.0190	0.0970	0.1955	0.0325	0.0840
200	0.0318	0.0048	0.0307	0.0040	0.0895	0.1987	0.0284	0.0820
250	0.0267	0.0002	0.0275	0.0000	0.0823	0.1684	0.0247	0.0690
300	0.0224	0.0000	0.0247	0.0000	0.0754	0.0476	0.0214	0.0120
350	0.0191	0.0000	0.0221	0.0000	0.0690	0.0088	0.0184	0.0030
400	0.0164	0.0000	0.0199	0.0000	0.0631	0.0032	0.0160	0.0010
450	0.0143	0.0000	0.0180	0.0000	0.0577	0.0005	0.0139	0.0010
500	0.0126	0.0000	0.0164	0.0000	0.0527	0.0003	0.0122	0.0010
1000	0.0065	0.0000	0.0101	0.0000	0.0282	0.0000	0.0068	0.0010
1500	0.0062	0.0000	0.0084	0.0000	0.0295	0.0000	0.0042	0.0010
2000	0.0061	0.0000	0.0061	0.0000	0.0300	0.0000	0.0051	0.0010

Table 4.6: Known-variance model cutoff values k^* , $ev(\mathbf{H}; \mathbf{y}_0)$ and P -value(\mathbf{y}_0) as function of n , with $d = 1$ and $d = 2$.

As a second stage in our simulation study, we set two sample sizes $n = 60$ and $n = 120$ to perform the tests for model (4.46), increasing the dimensionality of the parameter space. In that scenario, the vector of coefficients was such that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \theta_2)^\top$ and the hypotheses to be tested were

$$\mathbf{H} : \theta_2 = 0$$

$$\mathbf{A} : \theta_2 \neq 0.$$

So by varying the dimension of vector $\boldsymbol{\theta}_1$ the different models considered for each test

were obtained. Tables 4.7-4.8 and Figures 4.11-4.12 show the k^* and $\alpha_{\varphi_P^*}$ values as functions of d under the two formulations of conjugate prior distributions. For $d = 2$ and $d = 1$ in the unknown-variance and known-variance analysis respectively, the values correspond to model (4.45). We can say that, for a fixed hypothesis, the larger the size of the parameter space, the greater the value of k^* . In the case of the $\alpha_{\varphi_P^*}$ value, it does not change significantly when the dimensionality of the parameter space increases, except when the number of parameters is very large in relation to the sample size.

d	k^*	
	$n = 60$	$n = 120$
2	0.18500	0.08560
3	0.20420	0.19480
4	0.31510	0.39630
5	0.47790	0.49500
6	0.57670	0.53040
7	0.79970	0.67400
8	0.82970	0.70490
9	0.91250	0.80310
10	0.94540	0.92770
11	0.97300	0.92940
21	0.99990	0.99960
31	0.99990	0.99970
41	0.99990	0.99990
51	0.99990	0.99990

Table 4.7: Unknown-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of d , with $n = 60$ and $n = 120$.

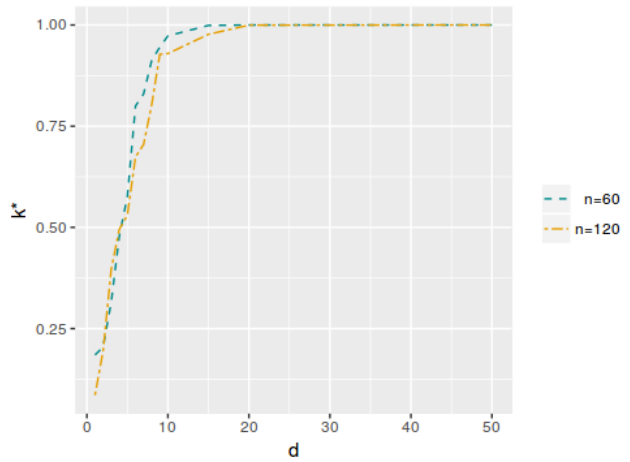


Figure 4.11: Unknown-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of d , with $n = 60$ and $n = 120$.

d	k^*	
	$n = 60$	$n = 120$
1	0.04590	0.01870
2	0.13180	0.12830
3	0.29400	0.18730
4	0.34320	0.14560
5	0.37160	0.44770
6	0.58160	0.53420
7	0.65620	0.62000
8	0.78070	0.65780
9	0.84350	0.80460
10	0.92490	0.77400
20	0.99940	0.99900
30	0.99990	0.99990
40	0.99990	0.99990
50	0.99980	0.99990

Table 4.8: Known-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of d , with $n = 60$ and $n = 120$.

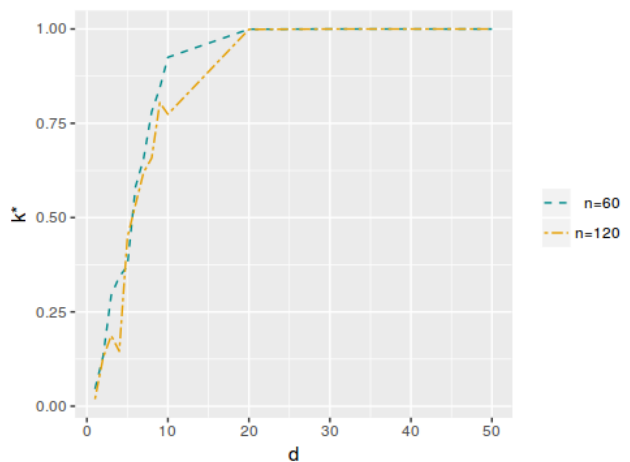


Figure 4.12: Known-variance model cutoff values k^* for $ev(\mathbf{H}; \mathbf{y})$ as a function of d , with $n = 60$ and $n = 120$.

d	$\alpha_{\varphi_P^*}$	
	$n = 60$	$n = 120$
2	0.03700	0.02100
3	0.03300	0.03800
4	0.03700	0.03600
5	0.04100	0.03800
6	0.04800	0.03300
7	0.04400	0.03500
8	0.04600	0.03100
9	0.05000	0.03600
10	0.04500	0.03900
11	0.04600	0.04000
21	0.05100	0.03700
31	0.05300	0.03700
41	0.07200	0.03600
51	0.12600	0.04100

Table 4.9: Unknown-variance model optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as a function of d , with $n = 60$ and $n = 120$.

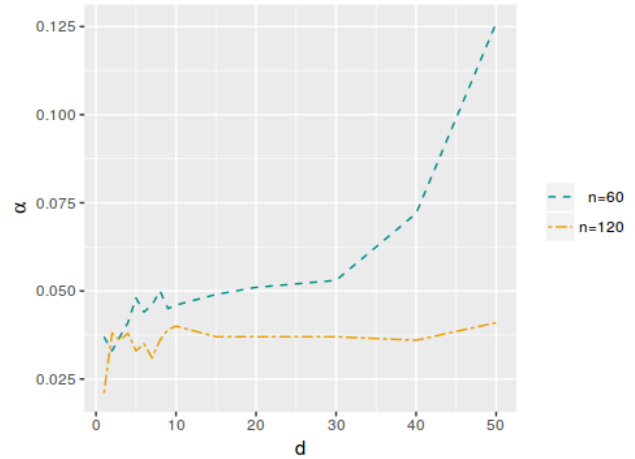


Figure 4.13: Unknown-variance model optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as a function of d , with $n = 60$ and $n = 120$.

d	$\alpha_{\varphi_P^*}$	
	$n = 60$	$n = 120$
1	0.04800	0.03000
2	0.04200	0.03900
3	0.04200	0.03800
4	0.04200	0.03800
5	0.04100	0.03700
6	0.04100	0.03800
7	0.03700	0.03900
8	0.04100	0.03900
9	0.04100	0.04000
10	0.03800	0.04000
20	0.04500	0.04600
30	0.04900	0.04400
40	0.08000	0.03900
50	0.12600	0.04500

Table 4.10: Known-variance model optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as a function of d , with $n = 60$ and $n = 120$.

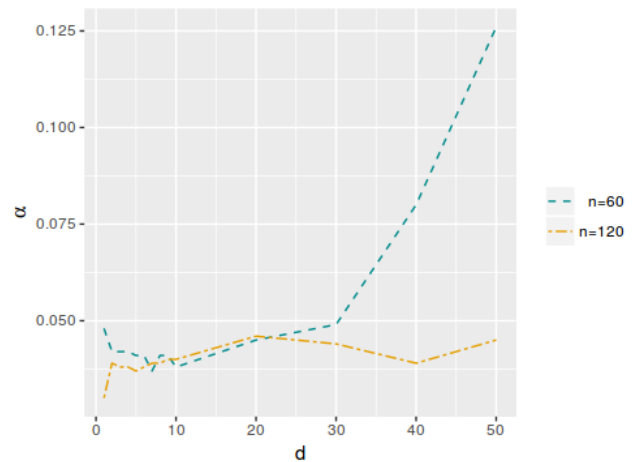


Figure 4.14: Known-variance model optimal averaged type-I error probability ($\alpha_{\varphi_P^*}$) as a function of d , with $n = 60$ and $n = 120$.

Figures 4.15 and 4.16 show the e -value versus P -value under \mathbf{H} for a single sample size $n = 60$ and for three different dimensionalities of the parameter space. Note that as d increases, the e -value grows faster, however, the P -value has a very little variation. Figures 4.17 and 4.18 display the histograms for the e -values and P -values under the same conditions. It can be seen that the P -value appears to have a uniform distribution in all scenarios, instead the e -value seems to approach 1 when d increases.

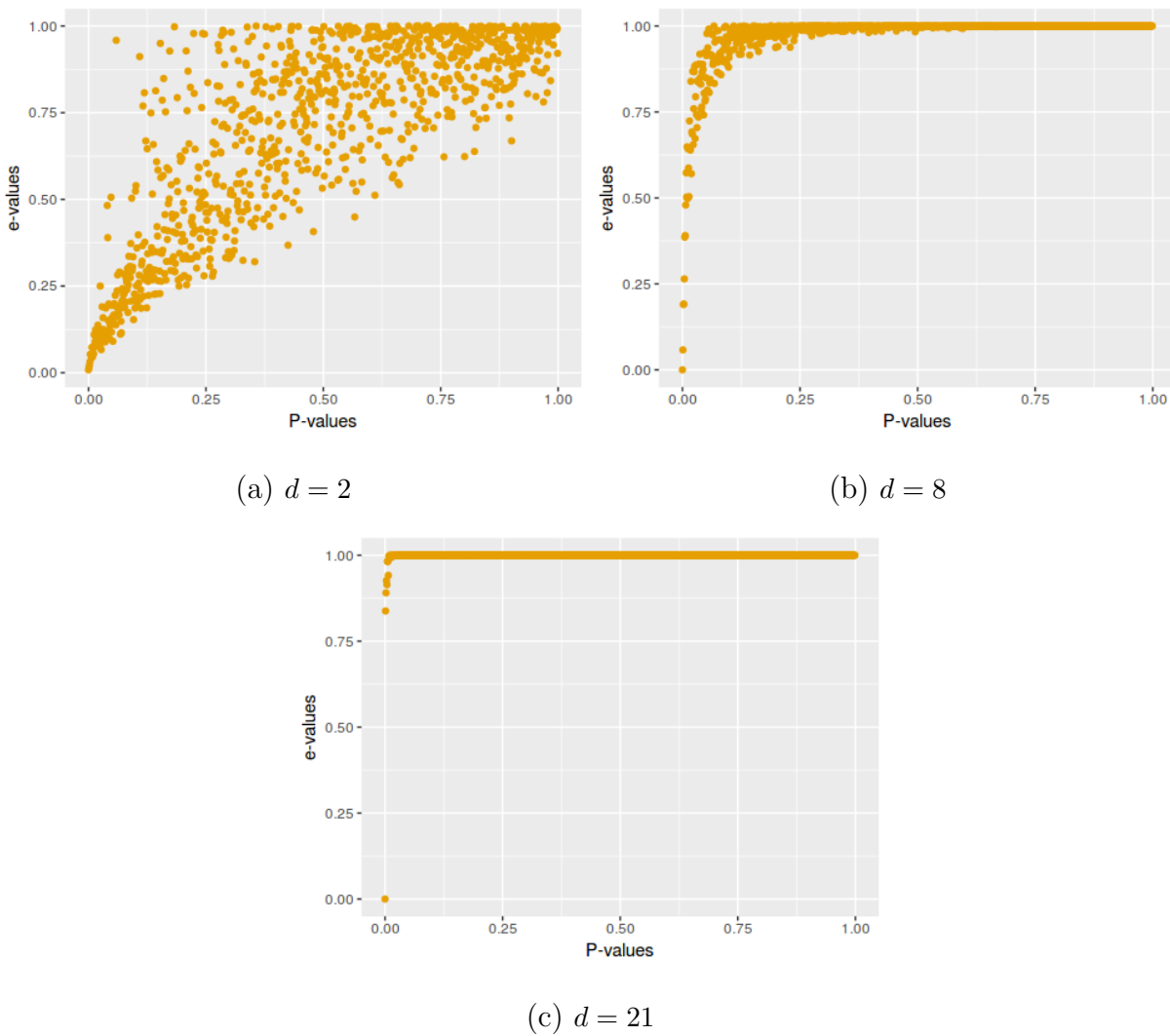


Figure 4.15: Unknown-variance model relationship between e -value and P -value. Sample size $n = 60$.

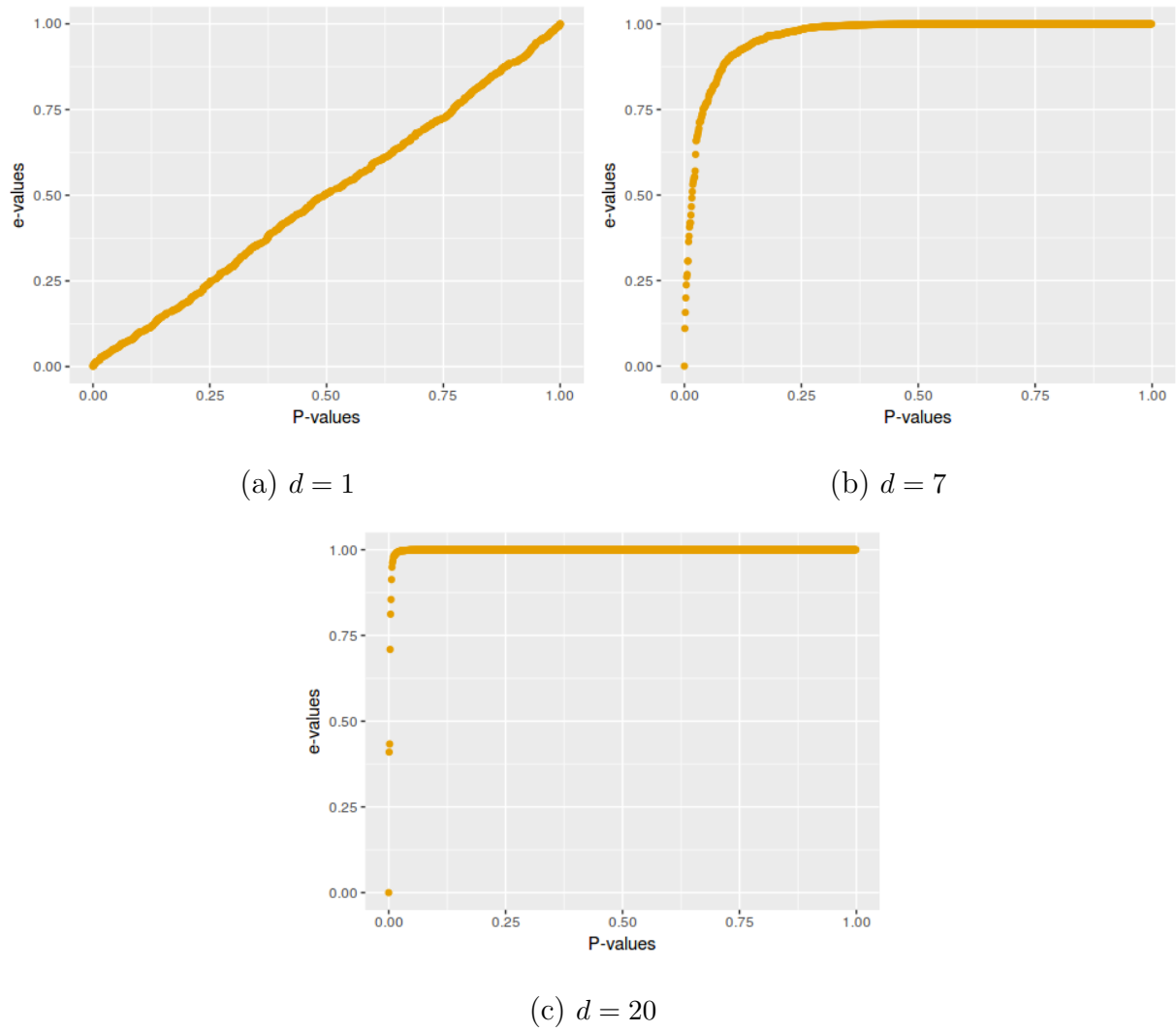
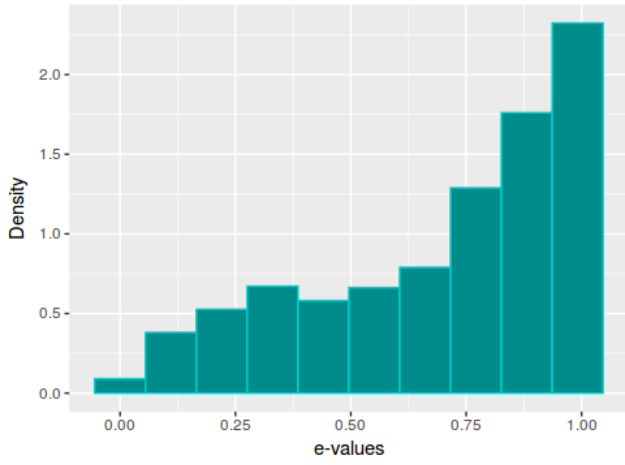
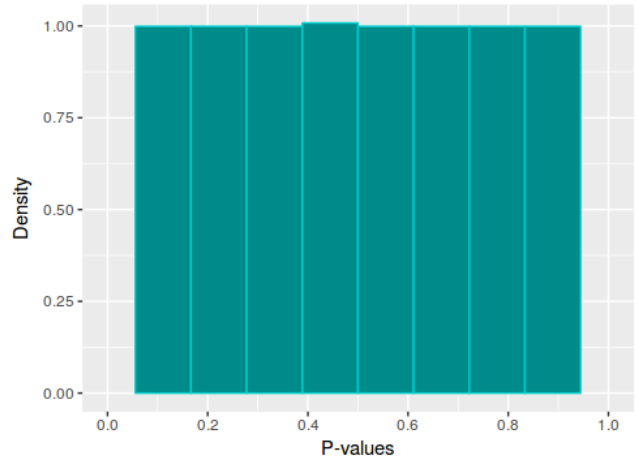


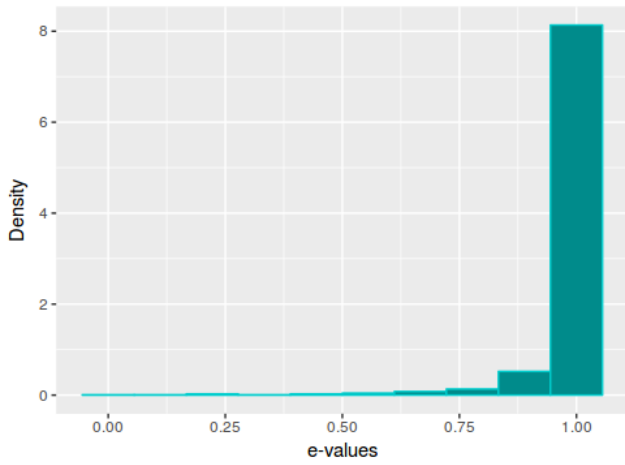
Figure 4.16: *Known-variance model relationship between e-value and P-value. Sample size $n = 60$.*



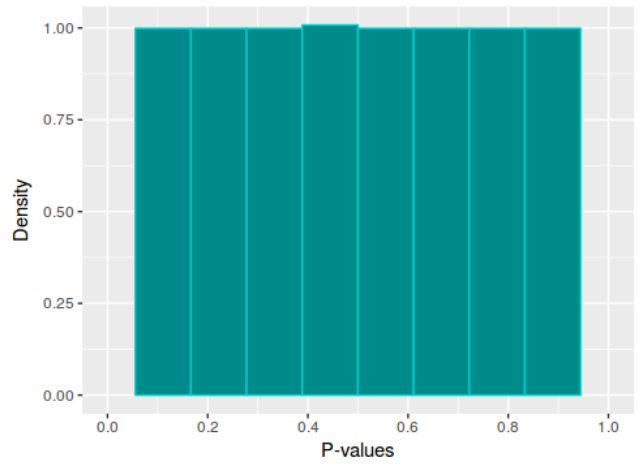
(a) $d = 2$



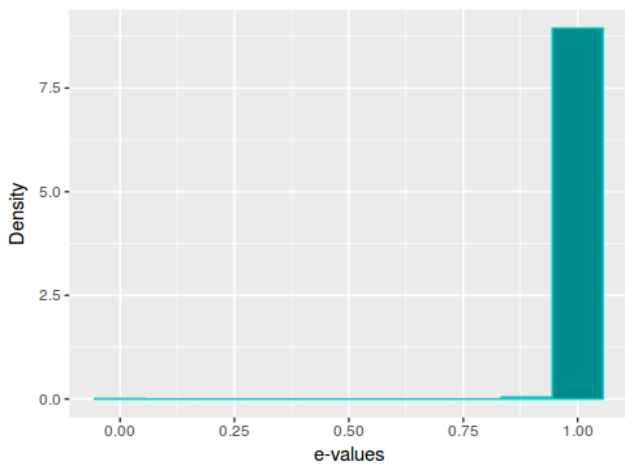
(b) $d = 2$



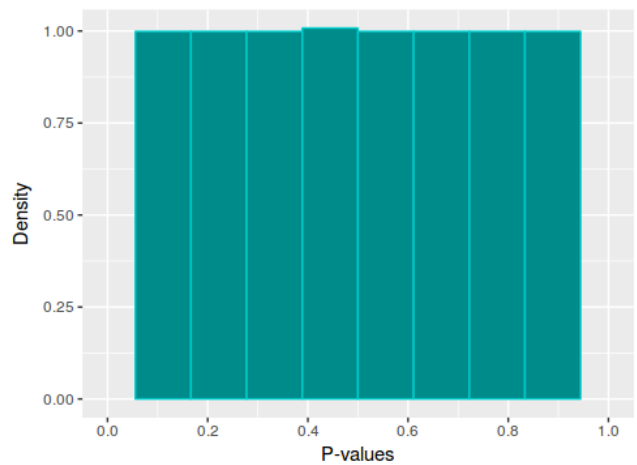
(c) $d = 8$



(d) $d = 8$



(e) $d = 21$



(f) $d = 21$

Figure 4.17: Unknown-variance model *e*-value and *P*-value densities under H . Sample size $n = 60$.

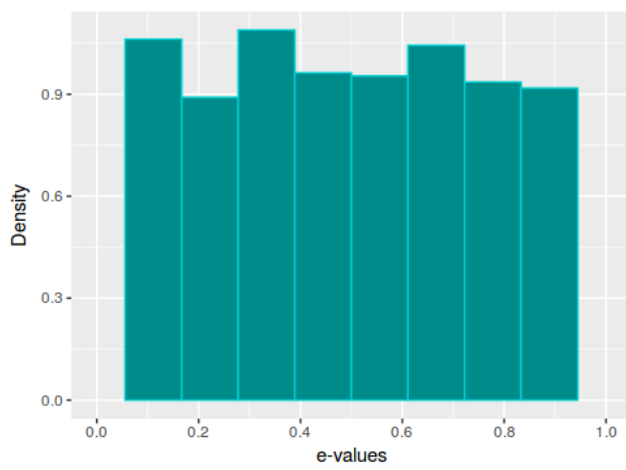
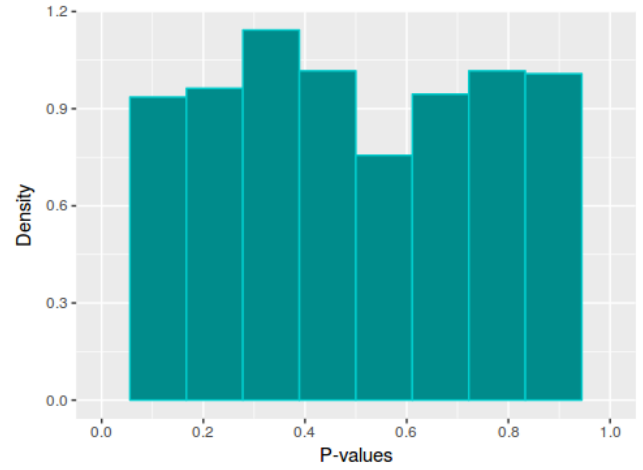
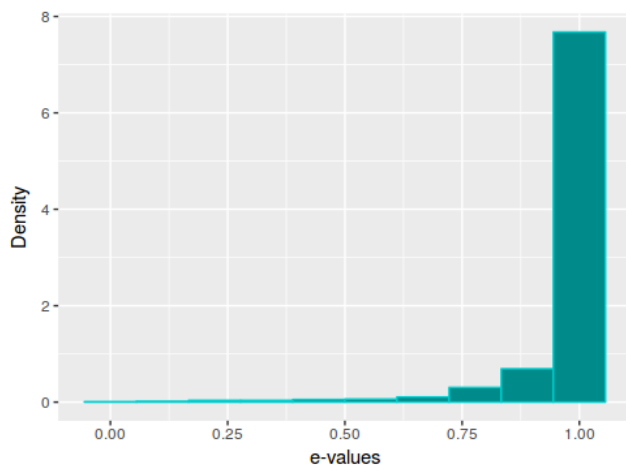
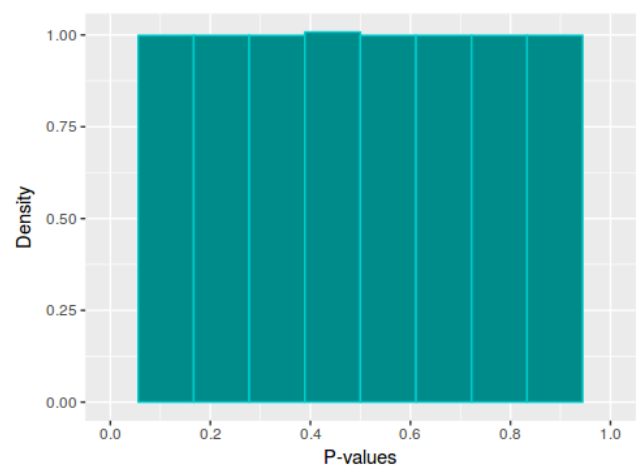
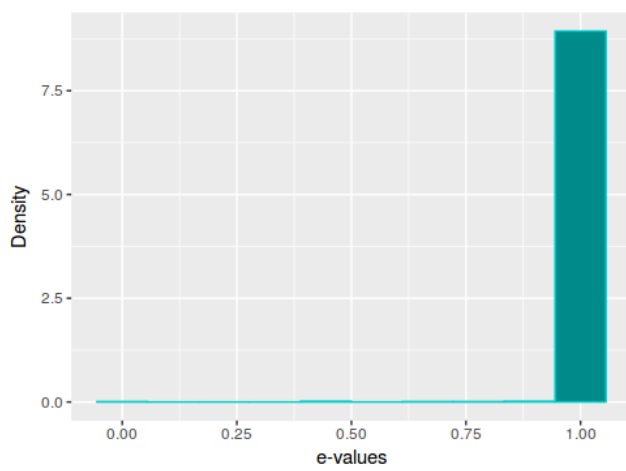
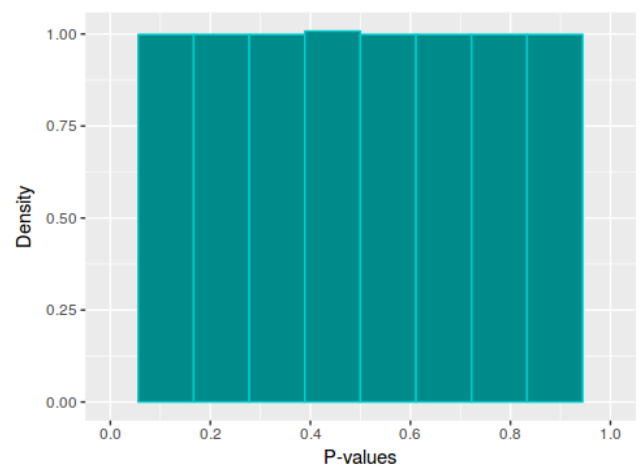
(a) $d = 1$ (b) $d = 1$ (c) $d = 7$ (d) $d = 7$ (e) $d = 20$ (f) $d = 20$

Figure 4.18: *Known-variance model e-value and P-value densities under H . Sample size $n = 60$.*

Tables 4.11 and 4.12 present the e -value and P -value calculated for a single random sample \mathbf{y}_0 in all models. Here, with the e -value the null hypothesis is less easily rejected. This may be related to two things: it may be due to approximation error as a result of the simulation process or due to the fact that the evidence apparently converges to 1 as the dimensionality of the parameter space increases, in which case a more detailed study is required.

$n = 60$					$n = 120$			
d	k^*	ev	$\alpha_{\varphi_P^*}$	Pv	k^*	ev	$\alpha_{\varphi_P^*}$	Pv
2	0.1850	0.6865	0.0370	0.3660	0.0856	0.0082	0.0210	0.0010
3	0.2042	0.5849	0.0330	0.1360	0.1948	0.7199	0.0380	0.1760
4	0.3151	0.8119	0.0370	0.1820	0.3963	0.9230	0.0360	0.2470
5	0.4779	0.0000	0.0410	0.0000	0.4950	0.0000	0.0380	0.0010
6	0.5767	0.5672	0.0480	0.0290	0.5304	0.7002	0.0330	0.0360
7	0.7997	0.8854	0.0440	0.0820	0.6740	0.9992	0.0350	0.2860
8	0.8297	0.3267	0.0460	0.0050	0.7049	0.7858	0.0310	0.0260
9	0.9125	0.1919	0.0500	0.0020	0.8031	0.0009	0.0360	0.0010
10	0.9454	0.0006	0.0450	0.0010	0.9277	0.0001	0.0390	0.0010
11	0.9730	0.0000	0.0460	0.0000	0.9294	0.0000	0.0400	0.0000
21	0.9999	0.0000	0.0510	0.0000	0.9996	0.0000	0.0370	0.0000
31	0.9999	1.0000	0.0530	0.0240	0.9997	0.0495	0.0370	0.0010
41	0.9999	0.9998	0.0720	0.0010	0.9999	0.0004	0.0360	0.0010
51	0.9999	1.0000	0.1260	0.0000	0.9999	0.0000	0.0410	0.0000

Table 4.11: Unknown-variance model cutoff values k^* , $ev(\mathbf{H}; \mathbf{y}_0)$ and P -value (\mathbf{y}_0) as functions of d , with $n = 60$ and $n = 120$.

$n = 60$					$n = 120$			
d	k^*	ev	$\alpha_{\varphi_P^*}$	Pv	k^*	ev	$\alpha_{\varphi_P^*}$	Pv
1	0.0459	0.1801	0.0480	0.1910	0.0187	0.0045	0.0300	0.0020
2	0.1318	0.4944	0.0420	0.2280	0.1283	0.4404	0.0390	0.2360
3	0.2940	0.7751	0.0420	0.2780	0.1873	0.7762	0.0380	0.3140
4	0.3432	0.0000	0.0420	0.0000	0.1456	0.0000	0.0380	0.0000
5	0.3716	0.6141	0.0410	0.0570	0.4477	0.4429	0.0370	0.0380
6	0.5816	0.9377	0.0410	0.1820	0.5342	0.5969	0.0380	0.4310
7	0.6562	0.5500	0.0370	0.0160	0.6200	0.0000	0.0390	0.0270
8	0.7807	0.2510	0.0410	0.0030	0.6578	0.0000	0.0390	0.0010
9	0.8435	0.0037	0.0410	0.0010	0.8046	0.0000	0.0400	0.0010
10	0.9249	0.0000	0.0380	0.0000	0.7740	0.0000	0.0400	0.0000
20	0.9994	0.0000	0.0450	0.0000	0.9990	0.0000	0.0460	0.0000
30	0.9999	1.0000	0.0490	0.1600	0.9999	1.0000	0.0440	0.0070
40	0.9999	1.0000	0.0800	0.0010	0.9999	0.5933	0.0390	0.0010
50	0.9998	1.0000	0.1260	0.0000	0.9999	0.0000	0.0450	0.0000

Table 4.12: Known-variance model cutoff values k^* , $ev(\mathbf{H}; \mathbf{y}_0)$ and P -value (\mathbf{y}_0) as functions of d , with $n = 60$ and $n = 120$.

4.5 Applications

In this section we present two application with real datasets using the unknown-variance model. We choose $a_0 = 3$ and $b_0 = 2$ as parameters of the Inverse Gamma prior distribution for σ^2 . Additionally, in the Normal prior for $\boldsymbol{\theta}$ given σ^2 , $\mathbf{m}_0 = \mathbf{0}_{p \times 1}$ and $\mathbf{V}_0 = \mathbf{I}_p$ are taken as parameters. The Monte Carlo approximations were made generating samples of size $M = 10000$.

4.5.1 Budget shares of British households dataset

We selected the BudgetUK dataset, available in the `Ecdat` package of R software (Croissant and Graves, 2019), which draws 1519 observations from the 1980-1982 British Family Expenditure Surveys (FES) (Blundell *et al.*, 1998). In our application we consider the following five variables:

- `wfood`: budget share for food expenditure
- `income`: total net household income (rounded to the nearest 10 UK pounds sterling)
- `walc`: budget share for alcohol expenditure
- `wfuel`: budget share for fuel expenditure
- `age`: age of household head.

All the expenditures and income are measured in pounds sterling per week. We took the budget share for food expenditure (`wfood`) as the dependent variable.

Figure 4.20 presents a matrix plot where the scatterplots of each pair of variables are drawn on the left part, Pearson correlations are displayed on the right side and variable distributions are on the diagonal.

Table 4.13 summarizes the results for the hypotheses $\mathbf{H} : \theta_j = 0, j = 1 \dots 5$, by performing the test with the p -value at 0.05 significance level and also the e -value and the P -value with their respective adaptive significance levels. The cases when \mathbf{H} is rejected are represented by the cells in **boldface**. It can be seen that, unlike the p -value, the e -value and the P -value do not reject the hypothesis of nullity of the coefficient associated with the `age` variable.

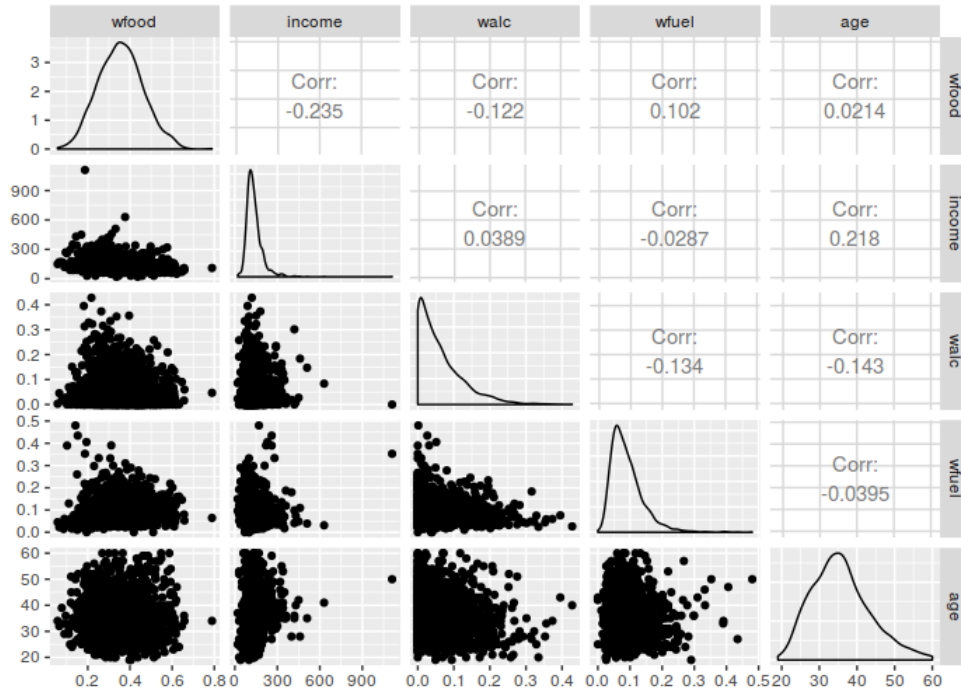


Figure 4.19: Budget shares of British households dataset matrix plot.

Coefficients	$\hat{\theta}_{Freq}$	α	pv	$\hat{\theta}_{Bayes}$	k^*	ev	$\alpha_{\varphi_P^*}$	Pv
Intercept	0.3758	0.0500	0.0000	0.3700	0.7078	0.0000	0.0382	0.0000
income	-0.0004	0.0500	0.0000	-0.0004	0.0113	0.0000	0.0001	0.0000
walc	-0.1533	0.0500	0.0003	-0.1283	0.9410	0.1890	0.1278	0.0172
wfuel	0.1717	0.0500	0.0007	0.1487	0.9520	0.1957	0.1468	0.0143
age	0.0009	0.0500	0.0119	0.0010	0.0764	0.3048	0.0004	0.0666

Table 4.13: Budget shares of British households dataset hypothesis-testing summary.

Table 4.14 exposes the optimal averaged error probabilities using the e -value and the P -value. It can be noted that the values are very similar with both methodologies.

Coefficients	$\alpha_{\varphi_e^*}^*$	$\alpha_{\varphi_P^*}$	$\beta_{\varphi_e^*}^*$	$\beta_{\varphi_P^*}$
Intercept	0.0466	0.0382	0.2157	0.2193
income	0.0000	0.0001	0.0006	0.0006
walc	0.1521	0.1278	0.4146	0.4145
wfuel	0.1508	0.1468	0.4679	0.4410
age	0.0004	0.0004	0.0080	0.0083

Table 4.14: Budget shares of British households dataset optimal averaged error probabilities.

4.5.2 Boston housing dataset

We also took the Boston dataset, which contains information about housing values obtained from census tracts in the Boston Standard Metropolitan Statistical Area (SMSA) in 1970 (Harrison and Rubinfeld, 1978). These data, available in the MASS package (Venables and Ripley, 2002) of R software, has 506 samples and 14 variables. We chose the following 10 variables to fit our model:

- `medv`: median value of owner-occupied homes in 1000s
- `crim`: per capita crime rate by town.
- `zn`: proportion of residential land zoned for lots over 25.000 sq.ft
- `indus`: proportion of non-retail business acres per town
- `rm`: average number of rooms per dwelling
- `age`: proportion of owner-occupied units built prior to 1940
- `dis`: weighted mean of distances to five Boston employment centers
- `tax`: full-value property-tax rate per 10.000
- `ptratio`: pupil-teacher ratio by town
- `black`: $1000(Bk - 0.63)^2$ where Bk is the proportion of blacks by town.

We select the median value of the owner-occupied homes in the census tract (`medv`) as the dependent variable.

Figure 4.20 presents a matrix plot with the scatterplots, the Pearson correlations and the variable distributions as in Figure 4.19.

The results for the hypotheses $\mathbf{H} : \theta_j = 0, j = 1 \dots 10$ by performing the test with the p -value, the e -value and the P -value, are summarized in Table 4.15. This time, with the e -value the null hypotheses are less rejected. The e -value does not reject the hypotheses of nullity of the coefficients associated to the `zn` and `indus` variables, while the p -value does. On the other hand, the P -value, unlike the p -value, does not reject the hypothesis for the `zn` variable but it does for the `Intercept`. As can be observed in Table 4.16, for these data, the optimal averaged error probabilities values are also very close.

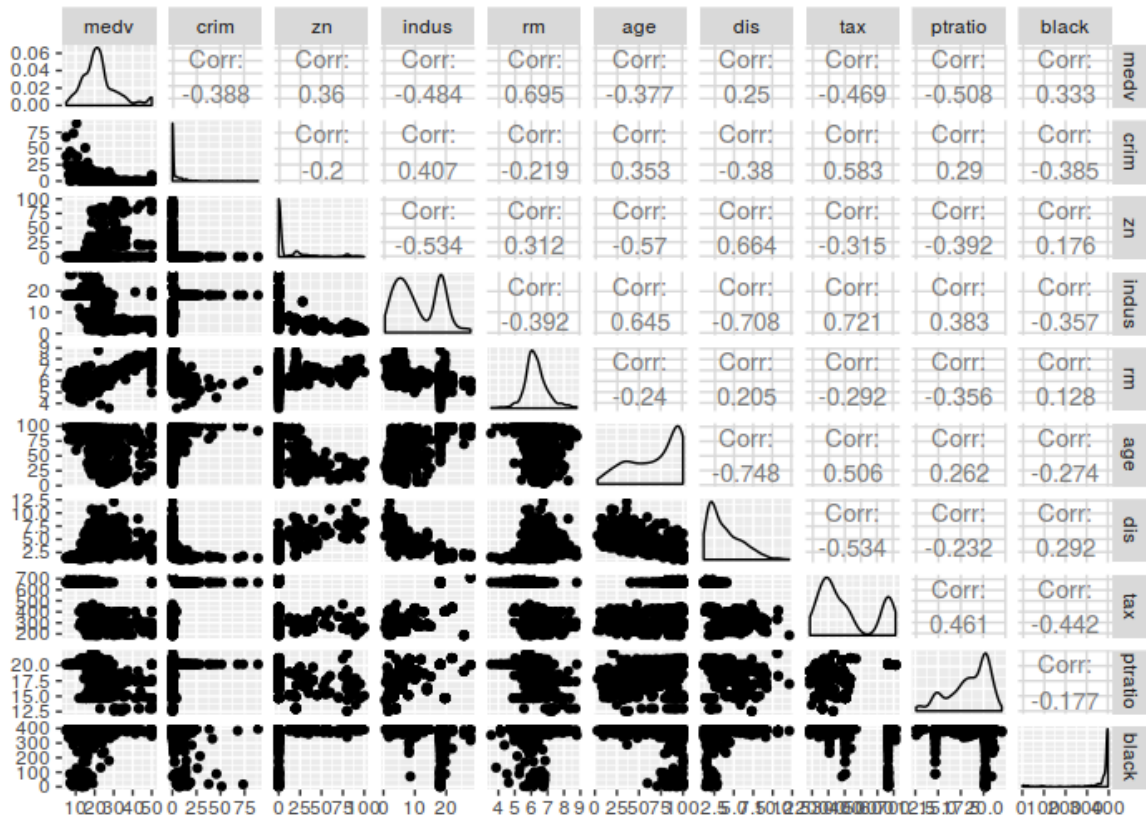


Figure 4.20: Boston housing dataset matrix plot.

Coefficients	$\hat{\theta}_{Freq}$	α	pv	$\hat{\theta}_{Bayes}$	k^*	ev	$\alpha_{\varphi_P^*}$	Pv
Intercept	1.7035	0.0500	0.6958	1.2035	0.9998	1.0000	0.1916	0.0085
crim	-0.1244	0.0500	0.0006	-0.1244	0.5780	0.3365	0.0010	0.0001
zn	0.0359	0.0500	0.0224	0.0362	0.4089	0.9012	0.0004	0.0025
indus	-0.1489	0.0500	0.0235	-0.1473	0.6390	0.9114	0.0025	0.0023
rm	6.7165	0.0500	0.0000	6.7336	0.9296	0.0000	0.0143	0.0000
age	-0.0655	0.0500	0.0000	-0.0648	0.3275	0.0141	0.0001	0.0000
dis	-1.3198	0.0500	0.0000	-1.3091	0.8146	0.0001	0.0095	0.0000
tax	-0.0030	0.0500	0.2324	-0.0030	0.0124	0.9996	0.0002	0.0198
ptratio	-0.7652	0.0500	0.0000	-0.7528	0.8223	0.0003	0.0053	0.0000
black	0.0145	0.0500	0.0000	0.0147	0.0297	0.0113	0.0001	0.0000

Table 4.15: Boston housing dataset hypothesis-testing summary.

Coefficients	$\alpha_{\varphi_e}^*$	$\alpha_{\varphi_P}^*$	$\beta_{\varphi_e}^*$	$\beta_{\varphi_P}^*$
Intercept	0.1321	0.1916	0.6494	0.4946
crim	0.0018	0.0010	0.0165	0.0173
zn	0.0006	0.0004	0.0075	0.0079
indus	0.0030	0.0025	0.0286	0.0292
rm	0.0222	0.0143	0.1123	0.1181
age	0.0000	0.0001	0.0068	0.0068
dis	0.0091	0.0095	0.0825	0.0808
tax	0.0000	0.0002	0.0016	0.0015
prratio	0.0081	0.0053	0.0494	0.0521
black	0.0000	0.0001	0.0019	0.0017

Table 4.16: *Boston housing dataset optimal averaged error probabilities.*

Conclusions

5.1 Final considerations

In this work we present a method to find a cutoff value k^* for the Bayesian evidence in the FBST by minimizing the linear combination of the averaged type-I and type-II error probabilities for a given sample size n and also for a given dimensionality d of the parameter space. In that sense, we provide a solution to the existing problem in the usual approach of hypothesis-testing procedures, where increase the sample size leads to rejection of the null hypothesis. Furthermore, we compare our results with those obtained by using the test proposed by [Pereira *et al.* \(2017\)](#) and [Gannon *et al.* \(2019\)](#).

In chapter 3, two simple illustrative examples are considered. We test a precise hypothesis against a composite alternative, first in a case where the random variable is discrete, and second for a variable following a continuous distribution. We note that the k^* value presents a decreasing trend with increasing n , which shows us the importance of taking the sample size into account when defining a significance level in hypothesis-testing procedures. In both examples, we also consider different priors to evaluate how much their choice affect the cutoff value. When comparing the e -value with the P -value, we observe that the decision is the same, no matter which index is used, however, the null hypothesis is more easily rejected in cases associated with posteriors that have smaller variances, which are generally those cases where the priors are more informative. Another important thing to mention is that the e -value is higher than the P -value for smaller sample sizes when the prior is more informative, but they are very similar when the sample size is larger.

In chapter 4 we present the proposed methodology to determine the adaptive cutoff value k^* for evidence extended to the context of linear regression models under a Bayesian approach. The conjugate model with unknown and known variance as prior formulation are considered. Additionally, all the steps to calculate the adaptive significance levels using P -value are described. For that purpose, we perform a simulation study in two stages. In the first phase, we fix the dimension of the parameter space and we change the sample size. The results show that the optimal averaged error probabilities reach similar values with both methodologies. On the other hand, the cutoff value k^* as well as the optimal averaged type-I error probability $\alpha_{\varphi_P^*}$ have a decreasing trend when the sample size increases. In the k^* case, the values are different when the parameter space changes. Thus, we can conclude that the k^* value depends not only on the sample size but also on the dimensionality of the parameter space. More specifically, k^* is greater when d is higher. However, this do not happen with $\alpha_{\varphi_P^*}$, which maintains almost the same value even if d increases. The results in the unknown-variance and the known-variance analysis do not differ considerably when the models have the same dimensionality, the k^* and $\alpha_{\varphi_P^*}$ values being a little higher in the unknown-variance case. In this context, we also observe that the decision remains the same regardless of the index used, either the e -value or the P -value, but, it is important to say that the null hypothesis is less easily rejected as the dimensionality of the parameter space increases and even more so in the case of the known-variance formulation that we chose.

As a second simulation stage, we set two sample sizes $n = 60$ and $n = 120$ and we increase the dimensionality of the parameter space to perform the tests for a fixed hypothesis. In this scenario, it is possible to notice that, the larger the size of d , the greater the value of k^* . In the $\alpha_{\varphi_P^*}$ value case, it does not change significantly when the dimensionality of the parameter space increases, except when the number of parameters is very large in relation to the sample size. The results for the e -value and the P -value show that, as d increases, the e -value grows faster, while the P -value has a very little variation. Additionally, the P -value display a uniform distribution under \mathbf{H} , while the e -value seems to approach 1 when d increases. Here, with the e -value the null hypothesis is less easily rejected. We believe, this may be related to two things: it may be due to the approximation error as a result of the simulation process, or it may be due to the fact that the evidence apparently converges to 1

as the dimensionality of the parameter space increases, in which case a more detailed study is required.

At the end of this chapter, two applications using real data are presented. The results are quite consistent with those obtained via simulation.

Finally, we can provide a general conclusion derived from the results of the illustrative examples as well as the regression models. With our suggestion of cutoff value for the evidence in the FBST and also with the procedure proposed by [Pereira *et al.* \(2017\)](#) and [Gannon *et al.* \(2019\)](#), increasing the sample size implies that the probabilities of both kind of optimal averaged errors and their linear combination decrease, unlike the most cases, where by setting a single level of significance independent of sample size, only type-II error probability decreases.

5.2 Suggestions for future research

- The methodology we propose to determine the adaptive cutoff value for evidence in the FBST could be extended to more complex models with different prior specifications, which would involve, among other things, using approximate methods to find the prior predictive densities under the null and alternative hypotheses.
- It would be interesting to use a low-level programming language to more efficiently evaluate a greater number of covariates and also to consider larger sample sizes than those presented in this work.

Support Material

A.1 Marginal distribution of a normal sample mean

Let $\bar{X}|\theta \sim \text{Normal}(\theta, \sigma^2/n)$ ($\theta \in \mathbb{R}$ and $\sigma^2 > 0$ known) and $\theta \sim \text{Normal}(m, v^2)$ ($m \in \mathbb{R}$ and $v^2 > 0$). The marginal distribution of \bar{X} is given by

$$\begin{aligned}
 f(\bar{x}) &= \int_{-\infty}^{\infty} f(\bar{x}|\theta) g(\theta) d\theta \\
 &\propto \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \sigma^2/n}} \exp\left\{-\frac{(\bar{x} - \theta)^2}{2 \sigma^2/n}\right\} \times \frac{1}{\sqrt{2\pi v^2}} \exp\left\{-\frac{(\theta - m)^2}{2v^2}\right\} d\theta \\
 &\propto \exp\left\{-\frac{1}{2} \left(\frac{\bar{x}^2}{\sigma^2/n} + \frac{m^2}{v^2}\right)\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \left(\theta^2 \left[\frac{n}{\sigma^2} + \frac{1}{v^2}\right] - 2\theta \left[\frac{n\bar{x}}{\sigma^2} + \frac{m}{v^2}\right]\right)\right\} d\theta \\
 &\propto \exp\left\{-\frac{1}{2} \left(\frac{\bar{x}^2}{\sigma^2/n} + \frac{m^2}{v^2} - \frac{\left[\frac{n\bar{x}}{\sigma^2} + \frac{m}{v^2}\right]^2}{\left[\frac{n}{\sigma^2} + \frac{1}{v^2}\right]}\right)\right\} \times \\
 &\quad \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2} \left[\frac{n}{\sigma^2} + \frac{1}{v^2}\right] \left(\theta^2 - 2\theta \frac{\left[\frac{n\bar{x}}{\sigma^2} + \frac{m}{v^2}\right]}{\left[\frac{n}{\sigma^2} + \frac{1}{v^2}\right]} + \frac{\left[\frac{n\bar{x}}{\sigma^2} + \frac{m}{v^2}\right]^2}{\left[\frac{n}{\sigma^2} + \frac{1}{v^2}\right]}\right)\right\} d\theta \\
 &\propto \exp\left\{-\frac{1}{2} \left(\frac{\bar{x}^2}{\sigma^2/n} + \frac{m^2}{v^2} - \frac{\left[\frac{n\bar{x}}{\sigma^2} + \frac{m}{v^2}\right]^2}{\left[\frac{n}{\sigma^2} + \frac{1}{v^2}\right]}\right)\right\} \times
 \end{aligned}$$

$$\begin{aligned}
& \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2 \left(\frac{\sigma^2 v^2}{\sigma^2 + n v^2} \right)} \left(\theta - \frac{\sigma^2 m + n v^2 \bar{x}}{\sigma^2 + n v^2} \right)^2 \right\} d\theta \\
& \propto \exp \left\{ -\frac{1}{2} \left(\frac{\bar{x}^2}{\sigma^2/n} + \frac{m^2}{v^2} - \left[\frac{\sigma^2 v^2}{\sigma^2 + n v^2} \right] \left[\frac{n^2 \bar{x}^2}{\sigma^4} + \frac{2n \bar{x} m}{\sigma^2 v^2} + \frac{m^2}{v^4} \right] \right) \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left(\bar{x}^2 \left[\frac{n}{\sigma^2} - \frac{n^2 v^2}{\sigma^2(\sigma^2 + n v^2)} \right] - 2\bar{x} \left[\frac{nm}{\sigma^2 + n v^2} \right] \right) \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left(\bar{x}^2 \left[\frac{n}{\sigma^2 + n v^2} \right] - 2\bar{x} \left[\frac{nm}{\sigma^2 + n v^2} \right] \right) \right\} \\
& \propto \exp \left\{ -\frac{1}{2} \left[\frac{n}{\sigma^2 + n v^2} \right] \left(\bar{x}^2 - 2\bar{x} \left[\frac{nm}{\sigma^2 + n v^2} \right] + m^2 \right) \right\} \\
& \propto \exp \left\{ -\frac{1}{2 \left(\frac{\sigma^2}{n} + v^2 \right)} (\bar{x} - m)^2 \right\}.
\end{aligned}$$

Then $\bar{X} \sim \text{Normal} \left(m, \frac{\sigma^2}{n} + v^2 \right)$.

Bibliography

- Azzalini and Genz (2016)** A. Azzalini and A. Genz. *The R package mnormt: The multivariate normal and t distributions (version 1.5-5)*, 2016. URL <http://azzalini.stat.unipd.it/SW/Pkg-mnormt>. From page 80
- Banerjee (2008)** S. Banerjee. *Bayesian Linear Model : Gory Details*, 2008. URL <http://www.biostat.umn.edu/~ph7440/pubh7440/BayesianLinearModelGoryDetails.pdf>. From page 59, 64
- Blundell et al. (1998)** R. Blundell, A. Duncan and K. Pendakur. Semiparametric estimation and consumer demand. *Journal of Applied Econometrics*, 13(5):435–461. From page 94
- Casella and Berger (2002)** G. Casella and R. L. Berger. *Statistical Inference*. Duxbury, Pacific Grove, 2nd ed. From page 8
- Croissant and Graves (2019)** Yves Croissant and Spencer Graves. *Ecdat: Data Sets for Econometrics*, 2019. URL <https://CRAN.R-project.org/package=Ecdat>. R package version 0.3-3. From page 94
- DeGroot (1970)** M. H. DeGroot. *Optimal Statistical Decisions*. McGraw-Hill, New York, 1fs ed. From page 9
- DeGroot (1986)** M. H. DeGroot. *Probability and Statistics*. Addison-Whelesley Publishing Company, Massachusetts, 2nd ed. From page 2, 3, 6, 7, 8, 9, 14
- Diniz et al. (2012)** M. Diniz, C. A. B. Pereira, A. Polpo, J. M. Stern and S. Wechsler. Relationship between bayesian and frequentist significance indices. *International Journal for Uncertainty Quantification*, 2(2):161–172. From page 36
- Fossaluzza (2008)** V. Fossaluzza. Testes de hipóteses em eleições majoritárias. Dissertação de Mestrado, Universidade de São Paulo, Instituto de Matemática e Estatística. Departamento de Estatística, São Paulo. From page 5

- Gannon et al. (2019)** M. A. Gannon, C. A. B. Pereira and A. Polpo. Blending bayesian and classical tools to define optimal sample-size-dependent significance levels. *The American Statistician*, 73(sup1):213–222. URL <https://doi.org/10.1080/00031305.2018.1518268>. From page [iii](#), [v](#), [2](#), [3](#), [12](#), [13](#), [16](#), [17](#), [99](#), [101](#)
- Harrison and Rubinfeld (1978)** D. Harrison and D.L. Rubinfeld. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management*, 5: 81–102. From page [96](#)
- Irony and Pereira (1995)** T. Z. Irony and C. A. B. Pereira. Bayesian hypothesis test: Using surface integrals to distribute prior information among the hypotheses. *Resenhas do Instituto de Matemática e Estatística da Universidade de São Paulo*, 2(1):27–46. URL <http://www.revistas.usp.br/resenhasimeusp/article/view/74716>. From page [2](#)
- Lindley (1957)** D. V. Lindley. A statistical paradox. *Biometrika Statistical Laboratory, University of Cambridge*, 44(1/2):187–192. From page [9](#)
- Madruga and Pereira (2005)** R. Madruga and C. A. B. Pereira. Power of FBST: Standard examples. *Instituto Interamericano de Estadística. Estadística*, 57(1):1–9. From page [10](#)
- Madruga et al. (2001)** R. Madruga, L. G. Esteves and S. Wechsler. On the bayesianity of pereira-stern tests. *TEST*, 10(2):291–299. From page [1](#), [11](#), [12](#)
- Mardia et al. (1979)** K. V. Mardia, J. T. Kent and J. M. Bibby. *Multivariate analysis. Probability and Mathematical Statistics*. Acad. Press, London. From page [63](#), [74](#)
- Montoya-Delgado et al. (2001)** L. E. Montoya-Delgado, T. Z. Irony, C. A. B. Pereira and M. R. Whittle. An unconditional exact test for the hardy-weinberg equilibrium law: sample-space ordering using the bayes factor. *Genetics*, 158 2:875–83. From page [2](#), [13](#)
- O’Hagan and Forster (2004)** A. O’Hagan and J. J. Forster. *Kendall’s Advanced Theory of Statistics, volume 2B: Bayesian Inference*, volume 2B. Arnold, London, 2nd ed. From page [59](#)
- Oliveira (2014)** M. C. Oliveira. Definição do nível de significância em função do tamanho amostral. Dissertação de Mestrado, Universidade de São Paulo, Instituto de Matemática e Estatística. Departamento de Estatística, São Paulo. From page [1](#), [2](#), [3](#), [15](#), [17](#)
- Paulino et al. (2003)** C. D. Paulino, M. A. Turkman and B. Murteira. *Estatística Bayesiana*. Fundação Calouste Gulbenkian, Lisboa. From page [21](#)
- Pereira (1985)** C. A. B. Pereira. *Teste de hipóteses definidas em espaços de diferentes dimensões: visão Bayesiana e interpretação Clássica*. Tese de Livre Docência, Universidade de São Paulo, Instituto de Matemática e Estatística. Departamento de Estatística, São Paulo. From page [1](#), [2](#), [3](#), [15](#), [17](#)

- Pereira and Stern (1999)** C. A. B. Pereira and J. M. Stern. Evidence and credibility: Full bayesian significance test for precise hypotheses. *Entropy*, 1(4):99–110. From page [iii](#), [v](#), [1](#), [9](#), [10](#)
- Pereira and Stern (2001)** C. A. B. Pereira and J. M. Stern. FBST regularization and model selection. <https://www.ime.usp.br/~cpereira/publications/isas2001-1.pdf>, 2001. From page [9](#)
- Pereira and Wechsler (1993)** C. A. B. Pereira and S. Wechsler. On the concept of p-value. *Brazilian Journal of Probability and Statistics*, 7(2):159–177. From page [2](#), [8](#), [12](#)
- Pereira et al. (2008)** C. A. B. Pereira, J. M. Stern and S. Wechsler. Can a significance test be genuinely bayesian? *Bayesian Analysis*, 3(1):79–100. From page [9](#), [10](#), [11](#)
- Pereira et al. (2017)** C. A. B. Pereira, E. Y. Nakano, V. Fossaluza, L. G. Esteves, M. A. Gannon and A. Polpo. Hypothesis tests for bernoulli experiments: Ordering the sample space by bayes factors and using adaptive significance levels for decisions. *Entropy*, 19(12):696. From page [iii](#), [v](#), [2](#), [3](#), [12](#), [14](#), [16](#), [17](#), [99](#), [101](#)
- Pericchi and Pereira (2016)** L. R. Pericchi and C. A. B. Pereira. Adaptative significance levels using optimal decision rules: Balancing by weighting the error probabilities. *Brazilian Journal of Probability and Statistics*, 30(1):70–90. From page [15](#)
- R Core Team (2017)** R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2017. URL <https://www.R-project.org/>. From page [80](#)
- Robert and Casella (2005)** C. P. Robert and G. Casella. *Monte Carlo Statistical Methods (Springer Texts in Statistics)*. Springer-Verlag, Berlin, Heidelberg. From page [44](#)
- Schervish (1995)** M. J. Schervish. *Theory of Statistics*. Springer. From page [8](#)
- Statisticat and LLC. (2018)** Statisticat and LLC. *LaplacesDemon: Complete Environment for Bayesian Inference*, 2018. URL <https://web.archive.org/web/20150206004624/http://www.bayesian-inference.com/software>. R package version 16.1.1. From page [80](#)
- Venables and Ripley (2002)** W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, New York, fourth ed. URL <http://www.stats.ox.ac.uk/pub/MASS4>. ISBN 0-387-95457-0. From page [96](#)