Mangold, Benedikt

Working Paper

Plausible prior estimation

IWQW Discussion Papers, No. 09/2014

Provided in Cooperation with:
Friedrich-Alexander University Erlangen-Nuremberg, Institute for Economics

Suggested Citation: Mangold, Benedikt (2014) : Plausible prior estimation, IWQW Discussion Papers, No. 09/2014, Friedrich-Alexander-Universität Erlangen-Nürnberg, Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung (IWQW), Nürnberg

This Version is available at:
https://hdl.handle.net/10419/101323

IWQW

Institut für Wirtschaftspolitik und Quantitative Wirtschaftsforschung

Diskussionspapier
Discussion Papers

No. 09/2014

Plausible Prior Estimation

Benedikt Mangold
University of Erlangen-Nuremberg

ISSN 1867-6707

Plausible Prior Estimation

Benedikt Mangold*

University of Erlangen-Nuremberg
Lange Gasse 20
D-90403 Nuremberg

August 1, 2014

Abstract

The problem of selecting a prior distribution when it comes to Bayes estimation often constitutes a choice between conjugate or noninformative priors, since in both cases the resulting posterior Bayes estimator (PBE) can be solved analytically and is therefore easy to calculate. Nevertheless, some of the implicit assumptions made by choosing a certain prior can be difficult to justify when a concrete sample of small size has been drawn. For example, when the underlying distribution is assumed to be normal, there is no reason to expect that the true but unknown location parameter is located outside the range of the sample. So why should a distribution with a non-compact domain be used as a prior for the mean? In addition, if there is some skewness in a sample of small size due to outliers when a symmetric distribution is assumed, this finding can be used to correct the PBE when determining the hyperparameters. Both ideas are applied to an empirical Bayes approach called plausible prior estimation (PPE) in the case of estimating the mean of a normal distribution with known variance in the presence of outliers. We propose an approach for choosing a prior and its respective hyperparameters, taking into account the above mentioned considerations. The resulting influence function as a frequentistic measure of robustness is simulated. To conclude, several simulation studies have been carried out to analyze the frequentistic performance of the PPE in comparison to frequentistic and Bayes estimators in certain outlier scenarios.

Keywords: Bayes Statistics, Objective Prior, Robustness.

1 Introduction and summary

There are many ways of selecting an adequate prior distribution, when constructing a Bayes estimator. The type of the prior distribution can be derived

*Benedikt.Mangold@fau.de

from previous experiments or logical considerations. When there is no prior knowledge, noninformative priors can be an appropriate choice. Those distributions do not require any parameters, but they sometimes neglect information about properties of the parameter of interest. Another possibility which is frequently used when investigating standard distributions is to use so called conjugate distributions, for which the Bayes estimators can be derived analytically. Employing a nonconjugate distribution often results in a numerical optimization problem, which brings its own difficulties during calculation.

This paper introduces a new way of choosing the prior distribution and its parameters. Determining the distribution of the underlying sample not only defines the likelihood function, but also implies certain properties as skewness of the distribution. Those properties, in addition to bounds that satisfiy specific objectives that can be derived for most types of parameters, are used for a unique identification of the prior distribution and its parameters.

This paper is organized as follows: Sections 2.1 - 2.3 give a short overview to the Bayes estimation framework, including an introduction to two Bayesian methods in which the plausible prior estimator (PPE) can partially be integrated. Section 2.4 briefly describes the idea behind the PPE, and Section 3 applies the PPE approach to symmetric distributions. In Section 4, asymptotic and finite-sample properties of the PPE are shown in the case of estimating the mean of a Gaussian distribution with known variance. Those properties are compared with the results from the Princton study from Andrews (1972) among others in Section 4.2 - 4.4. Section 5 gives a short outlook on the multivariate case and concluding remarks.

2 Bayes Estimation

2.1 Classical Bayes Estimation

In Bayes estimation, the posterior distribution $\pi(\theta)$ summarizes the information about an unknown distribution parameter θ from a parameter space $\Theta \subseteq \mathbb{R}$, considering the prior knowledge and the information given in the data $\boldsymbol{x}$. By the Bayes formula, one yields for the posterior distribution

$$f(\theta|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int_{\boldsymbol{x}} f(\boldsymbol{x}|\theta)\pi(\theta)\mathrm{d}\theta},$$

where $f(\boldsymbol{x}|\theta)$ denotes the likelihood function. In order to get a point estimation $\hat{\theta}$, a certain value of $f(\theta|\boldsymbol{x})$ has to be chosen. This usually happens through a decision function $d : \mathcal{X} \mapsto \Theta$, where $d \in \mathcal{D}$ from the set of possible decisions. A functional $\mathrm{L}(\theta, \hat{\theta}) : \Theta \times \Theta \mapsto \mathbb{R}^+$ represents some positive 'distance' between the estimated value, given the true θ, where $\mathrm{L}(\theta, \theta) = 0$. There are several loss functions, one of them is the mean squared error (MSE)

$$\mathrm{L}(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2 \tag{1}$$

The risk function

$$\mathrm{R}(\theta, d) := \mathbb{E}_\theta[\mathrm{L}(\theta, d(\boldsymbol{x}))] = \int_{\boldsymbol{x}} \mathrm{L}(\theta, d(\boldsymbol{x})) f(\boldsymbol{x}|\theta) \mathrm{d}\boldsymbol{x},$$

is used to calculate the Bayes risk given the prior distribution $\pi(\theta)$

$$\mathrm{r}_\pi(d) = \mathbb{E}_\pi[\mathrm{R}(\theta, d)] = \int_\Theta \mathrm{R}(\theta, d)\pi(\theta)\mathrm{d}\theta.$$

The decision function d that minimizes the Bayes risk r_π is called Bayes estimator.

In the following, just the squared loss function MSE will be considered – one can show quickly that the resulting Bayes estimator is the mean of the posterior distribution:

$$\theta_{\mathrm{BE}} = \mathbb{E}[\theta|\boldsymbol{x}] = \int_{\boldsymbol{x}} \theta \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{\int_{\boldsymbol{x}} f(\boldsymbol{x}|\theta)\pi(\theta)\mathrm{d}\theta} \mathrm{d}\theta. \tag{2}$$

There are certain strategies in determining the type of the prior distribution when it is unknown. Typically, one first determines the support of the unknown parameter, then one selects the prior distribution from a class of distributions with that very support. For example, for a scale parameter no one would use a distribution with unbounded support since the variance is never less than zero, so $\mathbb{R}$ as support would be unreasonable. In the following, we briefly describe two common approaches to select a prior distribution.

Noninformative Prior When there is no knowledge about the type of prior distribution, one could use a noninformative prior. Implicitly one would give the same probability to each possible value of the parameter of the sample distribution. In certain cases, this results in a prior distribution which is called an improper distribution since its integral does not sum to 1. Nevertheless, in some cases using an improper prior results in a proper posterior distribution.

Example 2.1. *Let X be a univariate random variable which follows $\mathcal{N}(\mu, \sigma^2)$, σ^2 is known and μ is to be estimated. Since there is no additional information about μ, one could use a uniform distribution $\pi(\mu) \propto 1$ on the support of μ as prior, which would be an improper prior distribution. Deriving the posterior distribution given a sample $\boldsymbol{x}$ yields*

$$\mu|\boldsymbol{x} \sim \mathcal{N}\left(\bar{x}, \frac{\sigma^2}{n}\right),$$

which is a proper posterior distribution. For the MSE as risk function, the Bayes estimator is then

$$\mu_{\mathrm{BE}} = \mathbb{E}[\mu|\boldsymbol{x}] = \bar{x}, \tag{3}$$

so the functional form of the maximum-likelihood estimator (MLE) $\bar{x}$ can be seen as a marginal case of the Bayesian framework when we estimate the location parameter of a normal distribution with known variance, the uniform distribution is used as a prior and the MSE as loss function.

Conjugate prior In many cases, although there is no specific justification, one chooses the prior distribution from the class of conjugate priors related to the sample distribution. This allows an analytically closed calculation of the Bayes estimator

Example 2.2. *Let X be a univariate random variable which follows $\mathcal{N}(\mu, \sigma^2)$, σ^2 is known and μ is to be estimated. If one chooses $\mu \sim \mathcal{N}(\tau, \gamma^2)$, one yields the posterior distribution*

$$\mu|\boldsymbol{x} \sim \mathcal{N}\left(\frac{\gamma^2}{\frac{\sigma^2}{n}+\gamma^2}\bar{x} + \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n}+\gamma^2}\tau, \frac{1}{\frac{\sigma^2}{n}+\gamma^2}\right). \tag{4}$$

Using the MSE as error function, the Bayes estimator for the posterior distribution (4) *of μ is a weighted average of the prior mean τ and the MLE $\bar{x}$,*

$$\mu_{\mathrm{BE}} = \frac{\gamma^2}{\frac{\sigma^2}{n}+\gamma^2}\bar{x} + \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n}+\gamma^2}\tau. \tag{5}$$

Those methods suffer from defects: first, both give probability to potential values of μ which are unreasonable since it is compelling to assume that the location parameter lies between the minimum and maximum of the sample, see Section 2.4. Further, no information about the shape of the underlying distribution of the data has been taken into account so far. For example, by assuming a sample to be from a normal distribution one already expects the skewness to be zero.

In addition to that, the procedure of choosing a conjugate prior is subjective and tends to be sensitive to the choice of the prior parameters, especially when the sample size is small. This can be seen in example 2.2, where the choice of the prior parameter τ directly influences the Bayes estimator.

The proposal given in this paper deals with these defects by choosing a prior distribution from a family of distributions that neglect values contradicting reasonable considerations about the range of the parameter. Further, information about skewness in the sample will be transmitted through the prior distribution to the final estimation process when one is expecting a symmetric sample distribution with skewness zero.

One approach that takes care of the mentioned defects can be found in robust Bayes estimation (Section 2.3) combined with elements of empirical Bayes estimation (Section 2.2) and is implemented in Section 2.4. Here those two defects will be considered during the introduced procedure of choosing the prior distribution. This will be called plausible prior estimation (PPE), and will be applied in Section 3 to the estimation of the location parameter of a normal distribution with known variance.

2.2 Empirical Bayes Estimation

The classical Bayes approach distinguishes strictly between the process of specifying a prior and the information in the actual sample. This prevents using

the information given in the data twice. Empirical Bayes means estimating hyperparameters of the prior distribution from the data (cf. Maritz (1970)). This is justified since this process implies an approximation to the hierarchical Bayes approach (cf. Bernardo and Smith (2004)). A possible conflict with this approximation is the missing variability of the hyperparameters (cf. Casella (1985)).

2.3 Robust Bayes Estimation

Instead of one single distribution in classical robust Bayes statistics a wide class Γ of prior distributions is used (see Berger (1990) and references therein). The quantity of interest κ of the posterior distribution is computed for every prior in Γ to establish lower and upper bounds for κ. On the one hand, the PPE approach deploys the class Γ of possible prior distributions, formed by fulfilling some formal objectives – on the other hand the parameters of the final prior distribution are chosen in concordance with the gained information from specifics of the sample, justified from the empirical Bayes approach. According to Berger (1990), when composing the class of prior distributions Γ the following four sometimes exclusionary objectives have to be complied:

1. Calculation of the range of κ has to be as easy as possible;
2. Γ should contain as many 'reasonable' prior distributions as possible to ensure robustness;
3. Γ should not contain unreasonable prior distributions, or robustness may be erroneously judged to be absent;
4. Γ should correspond to easily eliciTable prior information.

The introduced approach combines the idea of empirical Bayes (a plug-in estimator for the range of the sample) with a new approach while taking care of the mentioned objectives:

> deriving information concerning the hyperparameters due to deviation between theoretical properties of the underlying distribution and the properties of the actual sample.

2.4 Plausible Prior Estimation (PPE)

We assume the following holds for the parameter of location μ:

$$x_{(1)} := \min(\boldsymbol{x}) < \mu < \max(\boldsymbol{x}) =: x_{(n)}, \tag{6}$$

which is a minimum requirement for a sample $\boldsymbol{x}$ of size n. The introduced PPE approach consists of three parts: First, distributions with unreasonable parameter values are banned from Γ (Objectives 2 and 3). Since values of μ

outside of the range of the sample are not reasonable according to assumption (6), we yield for Γ:

$$\Gamma := \{\text{Distributions on a compact domain } [a;b] : -\infty < a < b < \infty\} \tag{7}$$

with $a = x_{(1)}$, $b = x_{(n)}$ In concordance with the objectives 1 and 2 we decrease Γ to

$$\Gamma_{p,q} := \{\text{Beta distributions on domain } [a;b] : -\infty < a < b < \infty,\ p,q > 1\}. \tag{8}$$

Note that the last inequality is strict, since we want to assume that the sample $\boldsymbol{x}$ is nontrivial - that is there are at least two different observations - with the implication that we will not expect μ to be positioned at a or b itself. The mode of a distribution from the family (8) therefore is always located in the open interval (a, b). This is equivalent to $p, q > 1$.

In the introduced robust Bayes approach from Section 2.3 one would compute the quantity τ for each distribution from $\Gamma_{p,q}$ to identify the lower and upper bounds. However, it is essential for the PPE approach that the objectives 1, 2 and 3 can only be verified after taking into account the actual sample. Therefore, in the second part of the PPE approach the hyperparameters a and b of the prior distributions in Γ are getting estimated using the information given in the sample (complying to objective 4) – a procedure known from empirical Bayes theory in Section 2.2.

The third part of the PPE approach involves choosing the hyperparameters of shape p and q. Those parameters shall now be used to correct the estimated value of the PBE for some deductions of a discrepancy between empirical and expected theoretical moments.

In the case of a symmetric distribution, one could try to consider a correction of a (highly probable) nonzero skewness in the sample by using the prior distribution as a weighting function of the likelihood. Since a high value of the sample skewness could have been established by asymmetric contamination in the data (c.f. Heymann et al. (2012)), we hope to get good robust properties for finite sample sizes as a side effect of applying this correction.

Again, this drawing of information from information of sample for deriving hyperparameters of the prior distribution is used in the field of empirical Bayes estimation in Section 2.2. Berger stated in Finetti et al. (1986) that among the "sins" which seems to be necessary to practice are the following:

> Delaying at least some of the prior specification until the likelihood function (for the observed data) is available (or alternatively allowing data-based choice of the prior if class of the priors).

In the following Section 3, the PPE is applied to the problem of estimating the location parameter from a symmetrical distribution with known variance.

3 PPE for symmetric distributions

If one specifies a symmetric distribution as the distribution of the population, one implicitly expects the skewness to be zero. It is not surprising that one would not find the sample skewness to be zero in a sample, say of minor size. Objective 2 will now be implemented in the Bayesian approach. A flexible distribution on a bounded support $[a, b]$ is for example the generalized beta distribution with density function

$$f(x; a, b, p, q) = \begin{cases} \frac{1}{\mathcal{B}(a,b,p,q)}(\mu - a)^{p-1}(b - \mu)^{q-1}, & a \leq x \leq b \\ 0, & \text{elsewhere,} \end{cases} \tag{9}$$

where

$$\mathcal{B}(a, b, p, q) = \mathcal{B}(p, q)(b - a)^{p+q-1} = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p + q)}(b - a)^{p+q-1},$$

see Figure 1.

Example 3.1. *Note that assuming the normal distribution for the population, for $p = q = 1$ the generalized beta distribution is the uniform distribution on $[a; b]$, and as n tends to infinity, the length of $[x_{(1)}; x_{(n)}]$ becomes arbitrarily large (see David and Nagaraja (2003)) since*

$$\mathbb{E}\left[[a; b]\right] \leq n\sqrt{\frac{2}{2n-1}\left(1 - \binom{2n-2}{n-1}^{-1}\right)} \sim \sqrt{n}.$$

This implies that $-x_{(1)}, x_{(n)} \to \infty$, as n goes to infinity. Therefore, in this case the noninformative prior introduced in example 2.1 is a marginal case of the PPE.

Given a sample $\boldsymbol{x}$ of size n, one advantage of using a beta distribution as prior is that if one would choose $a := x_{(1)}$ and $b := x_{(n)}$ the objective 2 would automatically be fulfilled as the prior distribution weights values outside of the range of the sample with 0.

Assuming that the sample distribution is normal with known variance σ^2 and using a prior distribution as in (9) for μ with $a := x_{(1)}$ and $b := x_{(n)}$, (2) can be estimated by

$$\frac{\int \mu \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}\right) \mathbb{1}_{[a,b]} \frac{1}{\mathcal{B}(a,b,p,q)}(\mu - a)^{p-1}(b - \mu)^{q-1} \mathrm{d}\mu}{\int \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}}\right) \mathbb{1}_{[a,b]} \frac{1}{\mathcal{B}(a,b,p,q)}(\mu - a)^{p-1}(b - \mu)^{q-1} \mathrm{d}\mu}. \tag{10}$$

Lemma 3.2. *The integral in 10 is bounded by the limits a and b from the beta distribution and can therefore be rewritten as*

$$\theta_{\mathrm{PPE}} = \frac{\mathbb{E}_Y\left[\mu(\mu - a)^{p-1}(b - \mu)^{q-1}\right]}{\mathbb{E}_Y\left[(\mu - a)^{p-1}(b - \mu)^{q-1}\right]}, \tag{11}$$

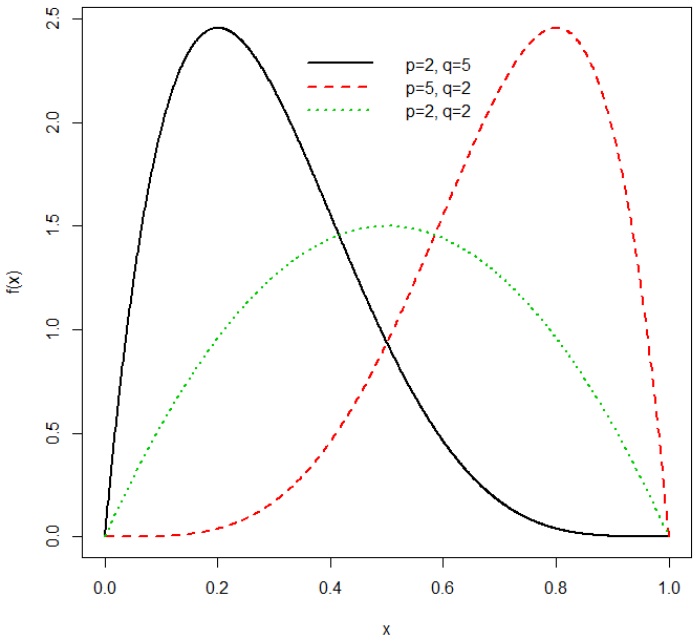

Figure 1: Density function of the beta distribution for certain parameters p and q.

with $Y \sim \mathcal{N}(\bar{x}, \frac{\sigma^2}{n})_b^a$ *– a truncated normal distribution on the interval* $[a, b]$*, where* $\bar{x}$ *denotes the arithmetic mean of the sample* $\boldsymbol{x}$*.*

Proof. cf. appendix 6. □

Given $(p, q) \in \mathbb{N}^2$, the PPE from equation (11) can be derived by evaluating the polynomial moments using the moment generating function of the truncated normal distribution. Since these steps can be done analytically, no elaborated optimization routine has to be applied.

Here is a crucial point since normally the parameters of the prior distribution must be specified before collecting the data – the choice of the parameters can either be made based on former repetitions of the experiment or the approaches introduced in chapter 2. In classical Bayes statistics it is strictly forbidden to analyze the data first and then draw conclusions concerning the parameters of the prior distribution. One interpretation of the PPE approach could be a violation of this interdiction since the minimum and maximum of the sample will be used as bounds for the support of the prior distribution.

On the other hand, as mentioned in the beginning of Section 2, the actual domain of the prior's support is always derived from plausible considerations

due to the nature of some parameter types – e.g. the scale parameter – to lie in predefined regions.

So far, only the assumption (6) played a role in the considerations. Assuming the data to be from a symmetric distribution, one would expect the sample skewness to be near zero. In the absence of contamination, deviations from zero are random and converge to 0 a.s. by the law of large numbers as the sample size n goes to infinity. For finite n in a range of, say, 20 the empirical skewness is almost surely not zero.

In the presence of asymmetric outliers, the deviation of the skewness from zero becomes more and more systematic, and the credibility of an observation that induces skewness is the lower the larger the deviation. Note that the PPE in (11) has two parameters p and q, which have not been specified so far. As can be seen in Figure 1, the beta distribution is able to model skewness, when $p \neq q$. So when the empirical skewness is not zero (e.g. due to outliers in the data), one could try to make a correction by adequately choosing p and q. Again, those two parameters are not specified when the sample is already available, but since the assumption that the data is from a symmetric population has already been made, a correction induced by a nonzero sample skewness is justified since this is the value one would expect under a symmetric distribution.

In the absence of outliers when estimating the mean from a normal population, a correction via the prior distribution is not expected to perform well since in this case the sample mean (cf. example 1, Section 2.1) is an UMVUE (cf. Lehmann and Casella (1998)). On the other hand, if the deviation of the skewness is systematic, i.e. due to outliers, the sample skewness contains information about the direction of the contamination in the data. By introducing a correction factor in the way of choosing the parameters p and q of the prior distribution adequately, one can transmit this information to the posterior distribution.

One sided outliers have an impact on the skewness of the sample, ν_n, as well as on the sample mean in that very direction (cf. Heymann et al. (2012)) –therefore, a nonzero sample skewness can be used as an indicator for the presence of asymmetric outliers when a symmetric distribution is assumed for the population. To counteract the effect of an outlier, one could use the prior distribution to give values on the diametrically opposite side a higher weight by setting the prior parameters adequately.

When using the class of beta distributions as a prior, their skewness can be calculated directly since the skewness ν is a function of p and q, see Johnson et al. (1994):

$$\nu(p,q) = \frac{2(q-p)\sqrt{p+q+1}}{(p+q+2)\sqrt{pq}}. \tag{12}$$

One proposal is to choose (p,q) in a way that the skewness of the prior distribution is the negative of ν_n, the sample skewness, in order to counteract the effect of an outlier on the sample mean. An outlier that pushes the sample skewness systematically further away from zero has two effects on the parameters

of the prior distribution – in this case the support of the prior distribution is expanded and at the same time the combination of (p, q) gives more plausibility to values that are on the diametrically opposite side of the sample mean. Figure 2 illustrates the influence of a single value, increasing from $x = 1.944$ to $x = 4$. The sample skewness gets larger, which results in a shift of the mode of the prior distribution to the opposite side.

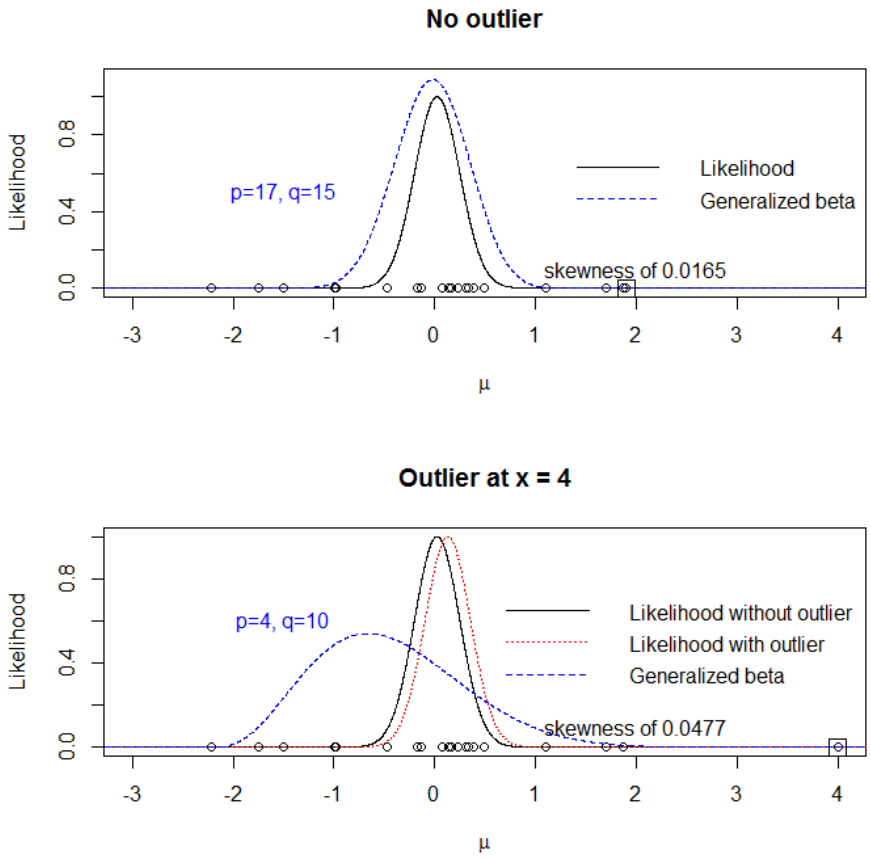

Figure 2: Choice of p and q of the prior distribution according to the sample skewness. The skewness induced by one single outlier can be used to correct the mean of the sample in the opposite direction.

The first approach of the choice of (p, q) was made by solving (12) directly for the values (p^*, q^*), such that

$$\nu_n = \nu(p^*, q^*), \tag{13}$$

but since in general $p^*, q^* \notin \mathbb{N}$, and the fractional derivative $\frac{\partial MGF}{\partial t^\alpha}$ of the moment generating function (MGF) is zero for $\alpha \neq \mathbb{N}$, no further effort has been taken in this direction. Instead, a grid search for $p,\, q \leq 50$, $(p,q) \in \mathbb{N}^2$ has been examined and the tuple (p^*, q^*), for which $|\nu_n - \nu(p^*, q^*)|$ was minimal, has been used to set the parameters for the PPE approach.

The implications of this choice are as follows: If there are no outliers, ν_n randomly diverges from zero and the resulting choice from (13) is $p^* \approx q^*$. The bigger the impact of the outlier, the bigger the induced correction by the prior distribution in the other direction.

4 Properties of the PPE

As shown in O'Hagan et al. (2004), under mild conditions the PBE is consistent and asymptotically normal because the prior distribution is essentially irrelevant as $n \to \infty$ and the PPE is asymptotically equivalent to the resulting maximum likelihood estimator (MLE). In the following we show explicitly that the PPE and the PBE with a conjugate Gaussian prior dirstibution are asymptotically equivalent when estimating a normal mean μ with known variance.

Lemma 4.1. *As $n \to \infty$, the PPE approach chooses $q \to p \to \infty$*

Proof. cf. appendix 6. □

Lemma 4.2. *Let $X \sim \text{Beta}(a, b, p, q)$ as in (9). For $p = q \to \infty$ the following holds:*

$$\frac{X - \frac{p}{p+q}\,(b-a) + a}{(b-a)\,\sqrt{\frac{p\,q}{(p+q)^2\,(p+q+1)}}} \xrightarrow{d} \mathcal{N}(0,1)$$

Proof. cf. appendix 6. □

Resulting from the lemmas 4.1 and 4.2 we can state the following

Theorem 4.3. *For $p \to q \to \infty$, the asymptotic prior distribution of the PPE aproach is*

$$\mu \overset{a}{\sim} \mathcal{N}\left(\tau_{\text{PPE}} = \frac{b+a}{2}, \gamma^2_{\text{PPE}} = \frac{(b-a)^2}{4\,(2\,p+1)}\right),\; p \approx q >> 0, \tag{14}$$

the resulting PBE is (cf. example 2.2)

$$\theta_{PPE} = \frac{\gamma^2_{\text{PPE}}}{\frac{\sigma^2}{n} + \gamma^2_{\text{PPE}}}\bar{x} + \frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} + \gamma^2_{\text{PPE}}}\tau_{\text{PPE}}.$$

From the lemma 4.1 follows that

$$\theta_{\text{PPE}} \overset{a}{=} \bar{x} = \theta_{\text{MLE}}$$

holds true as $n \to \infty$.

In the case of $n < 10$, the influence given by the prior distribution is severe which can be seen in the simulation results in Table 2 and 3, Section 4.2 - 4.4. The simulation results for larger sample sizes in Section 4.2 prove that the PPE estimator seems to have desirable properties in the presence of moderate contamination in the data in the sense of a smaller MSE by comparison with the MLE, and in some settings even with the median of the sample. This fact is supported by Section 4.1 – here the sensitivity curve has been simulated, and in a neighborhood of zero this curve seems to be descending within a symmetric interval around zero, which is a typical indicator for b-robustness in the M-estimator Framework, see Hampel (1986).

Another desirable property of the PPE is the absence of tuning parameters – all necessary parameters are set satisfying the objectives 1 to 4 that have been derived from robust Bayes statistics and take into account certain properties of the assumed distribution of the data. As a consequence, it can be tricky to compare the results from the simulations with e.g. Hampel's three-part estimator cf. Hampel (1986), where three tuning parameters have to be specified. Depending on the choice of these parameters, the result of the estimation can be arbitrarily good or bad.

4.1 Robustness results

One instrument to measure robustness is the influence curve, c.f. Huber (1981). It measures the impact of a change in sample distribution F, modeled as a contaminated version of F,

$$F_{\epsilon,x} := (1-\epsilon)F + \epsilon\delta_x,$$

where δ_x denotes the Dirac measure at a point x. Formally, the influence curve is a Gâteaux derivative of an estimator functional $\theta(\cdot)$ at F in direction x, defined as

$$\mathrm{IF}(x;\theta,F) := \lim_{\epsilon\to 0} \frac{\theta(F_{\epsilon,x}) - \theta(F)}{\epsilon}.$$

If the influence curve is bounded, an arbitrarily large observation can only have a limited influence on the value of the estimator. This property is called b-robustness. One example for a b-robust estimator is the median, one for a non-b-robust estimator the mean (c.f. Figure 3). The influence curve of the PPE for $n = 20$ and $\sigma = 1$ was simulated according to Huber (1981) and is shown in Figure 3. As can be seen, in a neighborhood of zero, the curve drifts away from the one of the ML estimator towards the x-axis, which can be interpreted as down weighting of observations that do not fit the model. This shape of the sensitivity curve can often be seen in redescending M-estimator framework. Outside of this neighborhood, however, the curve slowly bends back to the ML-curve, which implies that if the value of contamination is too huge, it cannot be handled by the PPE model. This understanding is supported in Section 4.2, where moderate contamination in the data can be handled well in the sense of a smaller MSE, whereas the model fails in the case of extreme contamination.

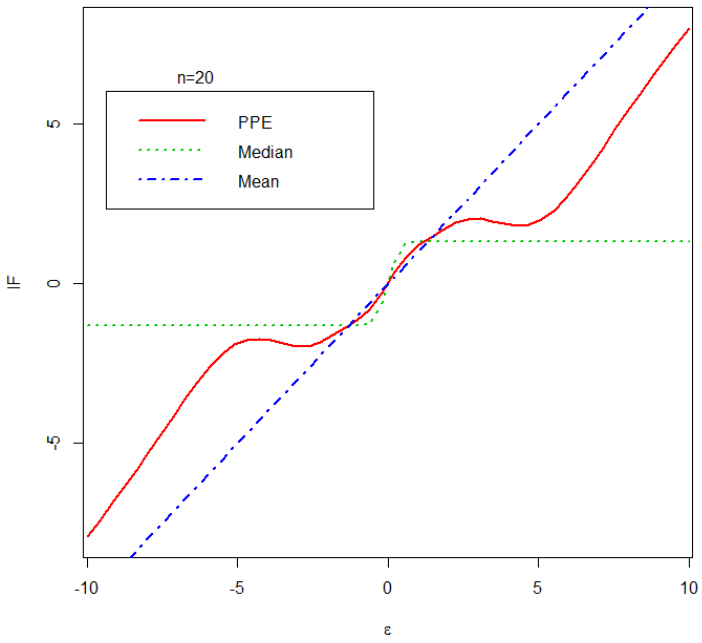

Figure 3: Simulated (N=10,000 repetitions) influence function for the PPE for $n = 20$ and $\sigma = 1$.

For moderate contamination ϵ, the trend of the sensitivity curve is similar to the one of the median, which can be seen in Figure 3. So it is not surprising to find the simulated performance results near those of the median, which can be seen in the following Section.

4.2 Simulation results – PPE vs. estimators from Andrews (1972)

In this Section, the performance of the PPE in comparison to other classes of estimators is investigated. The Princton study in Andrews (1972) was an early work which compared over 50 estimators in several scenarios, mostly estimating the location parameter of a Gaussian distribution, where the sample is from a mixture of Gaussian and heavy-tailed distributions, both symmetric and asymmetric.

Table 1 shows an excerpt of this study – in this case the pollution of the sample was from a Gaussian distribution with deviating mean μ_ϵ but identical variance. There are several estimators that perform better in the sense of a smaller average deviation from the true value, in most cases with a decreased

Table 1: Asymmetric pollution scenario from Andrews (1972), p. 110. Mean of the estimated value (Av) and variances (in relation to the variance of M, Var) of certain estimators, corrected by the sample size $n = 20$. 2 out of 20 observations from $\mathcal{N}(\mu_\epsilon, 1)$, 18 from $\mathcal{N}(0, 1)$. The **highlighted** estimators values were not simulated.

		$\mu_\epsilon = 2$		$\mu_\epsilon = 4$	
	Abbr.	*Av*	*Var*	*Av*	*Var*
Mean	M	0.903	1.000	1.798	1.000
10% symmetric trimmed mean	**10%**	0.752	1.092	0.893	1.153
Median	50%	0.592	1.607	0.626	1.651
Huber proposal 2, $k = 0.7$	**H07**	0.656	1.266	0.683	1.308
Huber proposal 2, $k = 2.0$	**H20**	0.859	1.023	1.258	1.115
1-Step Huber, $k = 0.7$*, start = median*	**D07**	0.663	1.233	0.699	1.273
1-Step Huber, $k = 2.0$*, start = median*	**D20**	0.838	1.042	1.143	1.159
M-estimator, ψ *bends at (1.2,3.5,8.0)*	**12A**	0.635	1.291	0.452	1.417
M-estimator, ψ *bends at (2.5,4.5,9.5)*	**25A**	0.770	1.103	0.785	1.346
Plausible prior estimator	PPE	0.758	1.092	1.288	1.068

average variance of the estimator. However, most of the compared estimators have several tuning parameters which make the performance sensitive to the choice of those parameters. One specification of each – the *Huber 2* estimators **H07**, **H20**, the *1-Step Huber* estimators **D07**, **D20** and the *Three-part M-estimator* **12A**, **25A** – performs worse than or similar to the PPE. The other constellation performs better than the PPE in the sense of a smaller MSE. All in all, the PPE yields comparable performance results to other, established estimators in the setting given by Andrews (1972). The next Section 4.3 will compare the PPE to the mean and a simple robust estimators without any tuning parameters.

4.3 Simulation results – PPE vs. tuning parameter free estimators

The Tables 2 and 3 show the performance of the PPE in relation to comparable, tuning-parameter-free estimators – the mean and the median. Here one can observe that the PPE outperforms the mean and the median in the sense of a smaller MSE when there is moderate asymmetric pollution in the data and the sample size is between $n = 20$ and $n = 35$.

Merely when the sample size is $n \leq 10$, the influence of the prior distribution is too strong which results in numerical problems as the denominator in (11) is getting to small. The bigger n, the smaller is the influence of the prior distribution on the posterior distribution in comparison to the likelihood function.

Table 2: MSE from $N = 10000$ repetitions for $\hat{\mu}$ from sample of size n drawn from $\mathcal{N}(0,1)$ with 10% contamination from $\mathcal{N}(\mu_\epsilon, 1)$.

	$\mu_\epsilon = 2$			$\mu_\epsilon = 5$		
n	Mean	Median	PPE	Mean	Median	PPE
10	**0.1426**	0.1657	·····	0.2645	**0.1721**	0.2121
20	0.0898	0.0953	**0.0816**	0.2087	**0.0996**	0.1339
30	0.0739	0.0726	**0.0664**	0.1947	**0.0762**	0.1382
40	0.0665	**0.0594**	0.0601	0.1877	**0.0630**	0.1432
50	0.0609	**0.0514**	0.0556	0.1814	**0.0547**	0.1455

Table 3: MSE from $N = 10000$ repetitions for $\hat{\mu}$ from sample of size n drawn from $\mathcal{N}(0,1)$ with 10% contamination from $\mathcal{N}(0, \sigma_\epsilon)$.

	$\sigma_\epsilon = \sqrt{3}$			$\sigma_\epsilon = \sqrt{5}$		
n	Mean	Median	PPE	Mean	Median	PPE
10	**0.1215**	0.1519	·····	**0.1422**	0.1567	·····
20	0.0613	0.0806	**0.0609**	0.0716	0.0828	**0.0653**
30	**0.0397**	0.0559	**0.0397**	0.0463	**0.0575**	0.0429
40	0.0304	0.0416	**0.0300**	0.0355	0.0428	**0.0328**
50	0.0245	0.0333	**0.0242**	0.0285	0.0343	**0.0266**

Therefore it is not surprising that the values of the MSE of the mean and the PPE are getting close to each other when n is increasing.

4.4 Simulation results – PPE vs. classical Bayes estimators

In this Section, we compare the PPE with several other classical Bayes estimators when estimating a normal mean with known variance $\sigma^2 = 1$ and several sample sizes n. The following estimators are used:

- Mean, Median (Med) and PPE from Section 4.3,
- Bayes estimators with a conjugate prior: centered normal prior distribution with $\sigma_{\mathrm{CN}} = 1, 3, 5$ (CN1, CN3 and CN5, resp.),
- Bayes estimator with noninformative prior: uniform prior on the interval $[-10; 10]$ (UNNI),
- Bayes estimator with noninformative prior, but taking into account objective 2: uniform prior on the interval $[x_{(1)}; x_{(n)}]$ (UNAS),
- Bayes estimator with heavy-tailed prior: student-t prior with $\nu = 1, 3, 5, 10$ (AT1, AT3, AT5 and AT10 resp.),
- ML estimator: given the bounds [$x_{(1)}$ and $x_{(n)}$], an MLE for a truncated normal distribution has been calculated (MLTR).

Resultsare given in Table 4 for $n = 10$, in Table 5 for $n = 20$ and in Table 6 for $n = 30$.

Generally, as mentioned before, the PPE performs well when the sample size n is below 50 and the contamination is moderate. As can be seen in the Tables 4 - 6, there are scenarios where the MSE is comparable or better than the MSE of the median and any other Bayes estimator. In addition to that, when contamination is absent, the PPE outperforms the median by a lot since the correction induced by the PPE is small.

When the sample size is small ($n = 10$), a single observation has a huge impact on both sample skewness (and thus the PPE-prior) and the likelihood. Therefore, when n is increasing the impact of a single outlier is declining and the correction made by the PPE is getting smaller. This effect can be seen in Table 4 – in three scenarios the PPE is the estimator with the lowest MSE when samples of size $n = 10$ are used and a moderate contamination is in the sample. For no or weak contamination, a Bayes estimator with a Cauchy prior is performing best. For heavy contamination, the median should be used since it has the smallest MSE.

When n is increasing, the following effects can be seen in Table 5 and 6: The PPE is still performing well in comparison to the other Bayes estimators or the MLE with a truncated normal distribution, but it is getting outperformed by the median when the contamination is strong.

Difference in MSE between the mean and the PPE gets smaller with increasing n, which is in agreement with the asymptotic behavior of the PPE as mentioned in Section 4.

Table 4: MSE of $N = 10000$ repetitions of the estimation of μ with known variance $\sigma^2 = 1$. Sample size $n = 10$, $(1-\epsilon)n$ from $\mathcal{N}(0,1)$, ϵn from $\mathcal{N}(\mu_\epsilon, \sigma_\epsilon)$ with degree of contamination $\epsilon = 0.1$.

μ_ϵ	0	2	4	5	0	0	0	2
σ_ϵ	1	1	1	1	$\sqrt{5}$	$\sqrt{9}$	$\sqrt{15}$	$\sqrt{5}$
Mean	0.0993	0.1447	0.2581	0.3473	0.1380	0.1804	0.2417	0.1855
Med	0.1392	0.1712	0.1696	**0.1678**	0.1552	0.1580	**0.1633**	0.1628
PPE	0.1087	0.1408	**0.1635**	0.1820	0.1279	**0.1413**	0.1717	**0.1457**
CN1	0.0821	0.1196	0.2133	0.2870	0.1141	0.1491	0.1997	0.1533
CN3	0.0971	0.1415	0.2524	0.3397	0.1350	0.1765	0.2364	0.1815
CN5	0.0986	0.1436	0.2562	0.3447	0.1371	0.1791	0.2399	0.1842
UNAS	0.0994	0.1447	0.2581	0.3473	0.1381	0.1805	0.2417	0.1856
UNNI	0.0993	0.1447	0.2581	0.3473	0.1380	0.1804	0.2417	0.1855
AT1	**0.0763**	**0.1125**	0.2039	0.2771	**0.1073**	0.1420	0.1933	0.1458
AT3	0.0799	0.1169	0.2099	0.2837	0.1115	0.1466	0.1978	0.1506
AT5	0.0807	0.1179	0.2113	0.2851	0.1125	0.1476	0.1987	0.1517
AT10	0.0814	0.1187	0.2123	0.2861	0.1133	0.1484	0.1993	0.1525
MLTR	0.1890	0.1638	0.1963	0.2663	0.1795	0.1971	0.2432	0.1903

5 Concluding remarks and outline

This paper deals with an approach of determining the parameters of a prior distribution which is exemplified by the problem of estimating the mean of a univariate normal distribution with known variance. Simulation studies show that the introduced method of choosing the necessary prior parameters leads to good statistical properties in the presence of asymmetric pollution in the underlying data. Especially when the sample size is around $n = 20$ and the pollution is moderate, the PPE yields results that are even better in the sense of a smaller MSE than the median, which is often used as a simple robust estimator. In comparison with other estimators, the PPE performs comparably to worse – but in most of the cases, the compared estimators have several parameters to specify before it comes to actual estimation. Depending on the parameter constellation, those estimators performed well or badly. So an advantage of the PPE is the way of parameter determination derived from the field of robust Bayes statistics, which takes away the influence of the arbitrary choice of the prior parameters and type of distribution on the posterior distribution and therefore the PBE.

Table 5: MSE of $N = 10000$ repetitions of the estimation of μ with known variance $\sigma^2 = 1$. Sample size $n = 20$, $(1-\epsilon)n$ from $\mathcal{N}(0,1)$, ϵn from $\mathcal{N}(\mu_\epsilon, \sigma_\epsilon)$ with degree of contamination $\epsilon = 0.1$.

μ_ϵ	0	2	4	5	0	0	0	2
σ_ϵ	1	1	1	1	$\sqrt{5}$	$\sqrt{9}$	$\sqrt{15}$	$\sqrt{5}$
Mean	0.0500	0.0901	0.2109	0.2992	0.0714	0.0900	0.1218	0.1107
Med	0.0740	0.0958	**0.1029**	**0.1002**	0.0845	0.0828	**0.0878**	0.0896
PPE	0.0537	0.0825	0.1359	0.1814	0.0648	**0.0714**	0.0896	**0.0837**
CN1	0.0453	0.0817	0.1913	0.2714	0.0648	0.0816	0.1105	0.1004
CN3	0.0494	0.0892	0.2086	0.2959	0.0707	0.0890	0.1205	0.1095
CN5	0.0498	0.0899	0.2103	0.2984	0.0712	0.0897	0.1215	0.1104
UNAS	0.0500	0.0901	0.2109	0.2992	0.0714	0.0900	0.1218	0.1107
UNNI	0.0500	0.0901	0.2109	0.2992	0.0715	0.0900	0.1218	0.1107
AT1	**0.0428**	**0.0777**	0.1840	0.2627	**0.0615**	0.0778	0.1060	0.0959
AT3	0.0443	0.0802	0.1885	0.2680	0.0635	0.0801	0.1088	0.0987
AT5	0.0447	0.0808	0.1895	0.2693	0.0640	0.0807	0.1094	0.0993
AT10	0.0450	0.0813	0.1904	0.2703	0.0644	0.0811	0.1100	0.0999
MLTR	0.0731	0.0852	0.1671	0.2552	0.0763	0.0841	0.1102	0.0971

Table 6: MSE of $N = 10000$ repetitions of the estimation of μ with known variance $\sigma^2 = 1$. Sample size $n = 30$, $(1-\epsilon)n$ from $\mathcal{N}(0,1)$, ϵn from $\mathcal{N}(\mu_\epsilon, \sigma_\epsilon)$ with degree of contamination $\epsilon = 0.1$.

μ_ϵ	0	2	4	5	0	0	0	2
σ_ϵ	1	1	1	1	$\sqrt{5}$	$\sqrt{9}$	$\sqrt{15}$	$\sqrt{5}$
Mean	0.0334	0.0728	0.1960	0.2825	0.0473	0.0594	0.0793	0.2123
Med	0.0505	0.0701	**0.0758**	**0.0758**	0.0574	0.0575	**0.0574**	**0.0691**
PPE	0.0351	**0.0652**	0.1394	0.1973	0.0436	**0.0489**	0.0606	0.1511
CN1	0.0313	0.0682	0.1836	0.2646	0.0443	0.0556	0.0742	0.1988
CN3	0.0332	0.0723	0.1946	0.2804	0.0470	0.0590	0.0787	0.2107
CN5	0.0333	0.0727	0.1958	0.2822	0.0472	0.0592	0.0791	0.2120
UNAS	0.0334	0.0728	0.1960	0.2825	0.0473	0.0594	0.0793	0.2123
UNNI	0.0334	0.0728	0.1960	0.2825	0.0473	0.0594	0.0793	0.2122
AT1	**0.0299**	0.0655	0.1779	0.2577	**0.0425**	0.0534	0.0716	0.1935
AT3	0.0308	0.0672	0.1814	0.2619	0.0436	0.0548	0.0732	0.1967
AT5	0.0310	0.0676	0.1822	0.2629	0.0439	0.0551	0.0736	0.1975
AT10	0.0311	0.0679	0.1829	0.2637	0.0441	0.0553	0.0739	0.1981
MLTR	0.0433	0.0659	0.1668	0.2540	0.0454	0.0543	0.0716	0.1883

The multivariate case will be the target for further investigation since the calculation of the multivariate median is cumbersome (Oja (1983)) and the derived PPE-framework seems to be extended easily to the multivariate case, by using the generalized Dirichlet distribution Kotz et al. (2000) as an extension of the beta distribution in the multivariate case.

6 Appendix

Proof of lemma 3.2.

$$\mu_{\mathrm{BE}} = \frac{\int \mu \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}} \right) \mathbb{1}_{[a,b]} \frac{1}{\mathcal{B}(a,b,p,q)} (\mu-a)^{p-1}(b-\mu)^{q-1} \mathrm{d}\mu}{\int \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x_i-\mu)^2}{\sigma^2}} \right) \mathbb{1}_{[a,b]} \frac{1}{\mathcal{B}(a,b,p,q)} (\mu-a)^{p-1}(b-\mu)^{q-1} \mathrm{d}\mu} =$$

$$= \frac{\int_a^b e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2\right)} \mu(\mu-a)^{p-1}(b-\mu)^{q-1}\mathrm{d}\mu}{\int_a^b e^{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\bar{x}-\mu)^2\right)} (\mu-a)^{p-1}(b-\mu)^{q-1}\mathrm{d}\mu} =$$

$$= \frac{e^{-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2}} \int_a^b e^{-\frac{1}{2\sigma^2}n(\bar{x}-\mu)^2} \mu(\mu-a)^{p-1}(b-\mu)^{q-1}\mathrm{d}\mu}{e^{-\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{2\sigma^2}} \int_a^b e^{-\frac{1}{2\sigma^2}n(\bar{x}-\mu)^2} (\mu-a)^{p-1}(b-\mu)^{q-1}\mathrm{d}\mu} =$$

$$= \frac{\int_a^b e^{-\frac{1}{2\sigma^2}n(\bar{x}-\mu)^2} \mu(\mu-a)^{p-1}(b-\mu)^{q-1}\mathrm{d}\mu}{\int_a^b e^{-\frac{1}{2\sigma^2}n(\bar{x}-\mu)^2} (\mu-a)^{p-1}(b-\mu)^{q-1}\mathrm{d}\mu},$$

which can be expanded to

$$\frac{\mathbb{E}_Y\left[\mu(\mu-a)^{p-1}(b-\mu)^{q-1}\right]}{\mathbb{E}_Y\left[(\mu-a)^{p-1}(b-\mu)^{q-1}\right]},$$

with $Y \sim \mathcal{N}(\bar{x}, \frac{\sigma^2}{n})_b^a$, the truncated normal distribution on the interval $[a,b]$, where $\bar{x}$ denotes the arithmetic mean of the sample $\boldsymbol{x}$. □

Proof of lemma 4.1. $p, q \in \mathbb{N}$, are chosen in the manner of the PPE approach according to (13).

Since in a sample from a symmetric distribution of finite size n the skewness of the sample ν_n is deviating from the true value zero with probability 1, $p = q$ is not admissible, since this would result in a skewness of the prior distribution of exactly zero. The further p diverges from q, the higher the value of the resulting skewness of the prior distribution will be. Therefore, for small deviations from ν_n of zero (which become more probable as n increases) p and q need to get as close as possible to each other. Since $p = q$ is not admissible, the smallest resulting skewness given p would be obtained by setting $q = p+1$ for a positive and $q = p-1$ for a negative sign of the sample skewness. This results in a skewness of the prior distribution of

$$\nu^+(p) = \frac{2^{\frac{3}{2}}\sqrt{p+1}}{\sqrt{p\,(p+1)}\,(2\,p+3)} \quad \text{and} \quad \nu^-(p) = \frac{2^{\frac{3}{2}}\sqrt{p}}{\sqrt{(p-1)\,p}\,(1-2\,p)} \tag{15}$$

for positive (negative)sample skewness.

As ν_n is tending to zero when n is increasing, the PPE approach has to choose an increasing p in order to minimize the distance between $\nu^{\pm}(p)$ and ν_n. Therefore, for $n \to \infty$, it is necessary that $p \to \infty$ (and implicitly $q \to \infty$), which means that $\nu^{\pm}(p) \to 0$ (according to (15)). □

Proof of Lemma 4.2. Let $f_X(x)$ be a density function of the generalized beta distribution of (9) for $q = p$, $p > 0$. This leads to

$$f_X(x) = \frac{\Gamma(2p)}{\Gamma(p)^2} \frac{((x-a)(b-x))^{p-1}}{(b-a)^{2p-1}}, \qquad a < b,\ a \le x \le b$$

with

$$\mathbb{E}[X] = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{4(2p+1)}.$$

Then the density function of a random variable $Y := T(X)$ with

$$T(x) = \frac{x - \frac{a+b}{2}}{\frac{(b-a)}{2\sqrt{2p+1}}}$$

can be derived via univariate variable transformation:

$$\begin{aligned}
f_Y(y) &= f_X\left(T^{-1}(y)\right) \left|\frac{\mathrm{d}T^{-1}(y)}{\mathrm{d}y}\right| = \\
&= f_X\left(\frac{(b-a)y}{2\sqrt{2p+1}} + \frac{a+b}{2}\right) \frac{(b-a)}{2\sqrt{2p+1}} = \\
&= \frac{\Gamma(2p)}{\Gamma(p)^2} \frac{\left(\left(\frac{(b-a)y}{2\sqrt{2p+1}} + \frac{a+b}{2} - a\right)\left(b - \frac{(b-a)y}{2\sqrt{2p+1}} - \frac{a+b}{2}\right)\right)^{p-1}}{(b-a)^{2p-1}} \frac{(b-a)}{2\sqrt{2p+1}} = \\
&= \frac{(b-a)^{2-2p}}{2\sqrt{2p+1}} \frac{\Gamma(2p)}{\Gamma(p)^2} \left(\left(\frac{(b-a)y}{2\sqrt{2p+1}} + \frac{b-a}{2}\right)\left(-\frac{b-a}{2} + \frac{(b-a)y}{2\sqrt{2p+1}}\right)\right)^{p-1} = \\
&= \frac{1}{2\sqrt{2p+1}} \frac{\Gamma(2p)}{\Gamma(p)^2} \left(\frac{1}{4} - \frac{y^2}{4(2p+1)}\right)^{p-1}, \qquad -\sqrt{2p+1} < y < \sqrt{2p+1}
\end{aligned}$$

Applying (Gradshtein et al., 2007, p. 895, 8.327.1) for $\Gamma(z)$ and (Gradshtein et al., 2007, p. 26, 1.211.4) for the limit in b one yields

$$\begin{aligned}
f_Y(y) &= \frac{1}{2\sqrt{2p+1}} \frac{\sqrt{2\pi}(2p)^{2p-\frac{1}{2}}e^{-2p}}{2\pi(2p)^{2p-1}e^{-2p}} \left(1 + \mathrm{o}(p^{-1})\right) \left(\frac{1}{4} - \frac{y^2}{4(2p+1)}\right)^{p-1} = \\
&= \frac{1}{\sqrt{2\pi}} \left(1 - \frac{y^2}{2p+1}\right)^{p-1} \frac{\sqrt{2p}}{\sqrt{2p+1}} \left(1 + \mathrm{o}(p^{-1})\right) \overset{p\to\infty}{=} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}
\end{aligned}$$

where o() denotes the Landau little-o. Therefore, for all $y \in \mathbb{R}$ $f_Y(y)$ converges pointwise to the standard Gaussian density function. Applying the theorem of Scheffé, Y converges in distribution to a random variable with the standard Gaussian density as density function. □

Proof of Theorem 4.3. (14) is a consequence of Lemma 4.1 and Lemma 4.2 when the continuous mapping theorem is applied and $p = q$. □

References

Andrews, D. F. (1972). *Robust estimates of location: Survey and advances.* Princeton University Press, Princeton (N.J.).

Berger, J. O. (1990). Robust Bayesian analysis: sensitivity to the prior. *Journal of Statistical Planning and Inference*, 25(3):303–328.

Bernardo, J. M. and Smith, A. F. M. (2004). *Bayesian theory.* Wiley series in probability and statistics. Wiley, Chichester and New York.

Casella, G. (1985). An Introduction to Empirical Bayes Data Analysis. *The American Statistician*, 39(2):83–87.

David, H. and Nagaraja, H. (2003). *Order Statistics.* Wiley series in probability and mathematical statistics. Probability and mathematical statistics. Wiley.

Finetti, B. d., Goel, P. K., and Zellner, A. (1986). *Bayesian inference and decision techniques: Essays in honor of Bruno de Finetti*, volume v. 6 of *Studies in Bayesian econometrics and statistics.* North-Holland and Sole distributors for the U.S.A. and Canada, Elsevier Science Pub. Co., Amsterdam and New York and New York and N.Y. and U.S.A.

Gradshtein, I. S., Ryzhik, I. M., Jeffrey, A., and Zwillinger, D. (2007). *Table of integrals, series, and products.* Elsevier and Academic Press, Amsterdam and Boston and Paris [et al.], 7th ed edition.

Hampel, F. R. (1986). *Robust statistics: The approach based on influence functions.* Wiley series in probability and mathematical statistics Probability and mathematical statistics. Wiley, New York and NY.

Heymann, S., Latapy, M., and Magnien, C. (2012). Outskewer: Using Skewness to Spot Outliers in Samples and Time Series. In *Advances in Social Networks Analysis and Mining (ASONAM), 2012 IEEE/ACM International Conference on*, pages 527–534. IEEE.

Huber, P. J. (1981). *Robust statistics.* Wiley series in probability and mathematical statistics. Wiley, New York.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous univariate distributions: Volume I*, volume 1 of *A Wiley-Interscience publication.* Wiley, New York and NY, 2 edition.

Kotz, S., Johnson, N., and Blakrishnan, N. (2000). *Continuous multivariate distributions: Volume 1: Models and applications.* Wiley series in probability and statistics. John Wiley & Sons Inc, New York, 2 edition.

Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation*. Springer texts in statistics. Springer, New York, 2nd ed. edition.

Maritz, J. S. (1970). *Empirical Bayes methods*. Methuen's monographs on applied probability and statistics. Methuen, London.

O'Hagan, A., Forster, J., and Kendall, M. G. (2004). *Bayesian inference*, volume 2B of *Kendall's advanced theory of statistics*. Arnold, London, 2 edition.

Oja, H. (1983). Descriptive statistics for multivariate distributions. *Statistics & Probability Letters*, 1(6):327–332.

Diskussionspapiere 2014
Discussion Papers 2014

01/2014 **Rybizki, Lydia**: Learning cost sensitive binary classification rules accounting for uncertain and unequal misclassification costs

02/2014 **Abbiati, Lorenzo, Antinyan, Armenak and Corazzini, Lucca:** Are Taxes Beautiful? A Survey Experiment on Information, Tax Choice and Perceived Adequacy of the Tax Burden

03/2014 **Feicht, Robert, Grimm, Veronika and Seebauer, Michael**: An Experimental Study of Corporate Social Responsibility through Charitable Giving in Bertrand Markets

04/2014 **Grimm, Veronika, Martin, Alexander, Weibelzahl, Martin and Zoettl, Gregor**: Transmission and Generation Investment in Electricity Markets: The Effects of Market Splitting and Network Fee Regimes

05/2014 **Cygan-Rehm, Kamila and Riphahn, Regina**: Teenage Pregnancies and Births in Germany: Patterns and Developments

06/2014 **Martin, Alexander and Weibelzahl, Martin:** Where and when to Pray? - Optimal Mass Planning and Efficient Resource Allocation in the Church

07/2014 **Abraham, M., Lorek, K., Richter, F. and Wrede, M.**: Strictness of Tax Compliance Norms: A Factorial Survey on the Acceptance of Inheritance Tax Evasion in Germany

08/2014 **Hirsch, B., Oberfichtner, M. and Schnabel C.**: The levelling effect of product market competition on gender wage discrimination

Diskussionspapiere 2013
Discussion Papers 2013

01/2013 **Wrede, Matthias**: Rational choice of itemized deductions

02/2013 **Wrede, Matthias**: Fair Inheritance Taxation in the Presence of Tax Planning

03/2013 **Tinkl, Fabian**: Quasi-maximum likelihood estimation in generalized polynomial autoregressive conditional heteroscedasticity models

04/2013 **Cygan-Rehm, Kamila**: Do Immigrants Follow Their Home Country's Fertility Norms?

05/2013 **Ardelean, Vlad and Pleier, Thomas**: Outliers & Predicting Time Series: A comparative study

06/2013 **Fackler, Daniel and Schnabel, Claus:** Survival of spinoffs and other startups: First evidence for the private sector in Germany, 1976-2008

07/2013 **Schild, Christopher-Johannes:** Do Female Mayors Make a Difference? Evidence from Bavaria

08/2013 **Brenzel, Hanna, Gartner, Hermann and Schnabel Claus**: Wage posting or wage bargaining? Evidence from the employers' side

09/2013 **Lechmann, D. S. and Schnabel C.:** Absence from work of the self-employed: A comparison with paid employees

10/2013 **Bünnings, Ch. and Tauchmann, H.:** Who Opts Out of the Statutory Health Insurance? A Discrete Time Hazard Model for Germany

Diskussionspapiere 2012
Discussion Papers 2012

01/2012 **Wrede, Matthias**: Wages, Rents, Unemployment, and the Quality of Life

02/2012 **Schild, Christopher-Johannes:** Trust and Innovation Activity in European Regions - A Geographic Instrumental Variables Approach

03/2012 **Fischer, Matthias:** A skew and leptokurtic distribution with polynomial tails and characterizing functions in closed form

04/2012 **Wrede, Matthias**: Heterogeneous Skills and Homogeneous Land: Segmentation and Agglomeration

05/2012 **Ardelean, Vlad**: Detecting Outliers in Time Series

Diskussionspapiere 2011
Discussion Papers 2011

01/2011 **Klein, Ingo, Fischer, Matthias and Pleier, Thomas:** Weighted Power Mean Copulas: Theory and Application

02/2011 **Kiss, David:** The Impact of Peer Ability and Heterogeneity on Student Achievement: Evidence from a Natural Experiment

03/2011 **Zibrowius, Michael:** Convergence or divergence? Immigrant wage assimilation patterns in Germany

04/2011 **Klein, Ingo and Christa, Florian:** Families of Copulas closed under the Construction of Generalized Linear Means

05/2011 **Schnitzlein, Daniel**: How important is the family? Evidence from sibling correlations in permanent earnings in the US, Germany and Denmark

06/2011 **Schnitzlein, Daniel**: How important is cultural background for the level of intergenerational mobility?

07/2011 **Steffen Mueller**: Teacher Experience and the Class Size Effect - Experimental Evidence

08/2011 **Klein, Ingo:** Van Zwet Ordering for Fechner Asymmetry

09/2011 **Tinkl, Fabian and Reichert Katja:** Dynamic copula-based Markov chains at work: Theory, testing and performance in modeling daily stock returns

10/2011 **Hirsch, Boris and Schnabel, Claus:** Let's Take Bargaining Models Seriously: The Decline in Union Power in Germany, 1992 – 2009

11/2011 **Lechmann, Daniel S.J. and Schnabel**, **Claus** : Are the self-employed really jacks-of-all-trades? Testing the assumptions and implications of Lazear's theory of entrepreneurship with German data

12/2011 **Wrede, Matthias**: Unemployment, Commuting, and Search Intensity

13/2011 **Klein, Ingo**: Van Zwet Ordering and the Ferreira-Steel Family of Skewed Distributions

Diskussionspapiere 2010
Discussion Papers 2010

01/2010 **Mosthaf, Alexander, Schnabel, Claus and Stephani, Jens:** Low-wage careers: Are there dead-end firms and dead-end jobs?

02/2010 **Schlüter, Stephan and Matt Davison:** Pricing an European Gas Storage Facility using a Continuous-Time Spot Price Model with GARCH Diffusion

03/2010 **Fischer, Matthias, Gao, Yang and Herrmann, Klaus:** Volatility Models with Innovations from New Maximum Entropy Densities at Work

04/2010 **Schlüter, Stephan and Deuschle, Carola:** Using Wavelets for Time Series Forecasting – Does it Pay Off?

05/2010 **Feicht, Robert and Stummer, Wolfgang:** Complete closed-form solution to a stochastic growth model and corresponding speed of economic recovery.

06/2010 **Hirsch, Boris and Schnabel, Claus:** Women Move Differently: Job Separations and Gender.

07/2010 **Gartner, Hermann, Schank, Thorsten and Schnabel, Claus:** Wage cyclicality under different regimes of industrial relations.

08/2010 **Tinkl, Fabian**: A note on Hadamard differentiability and differentiability in quadratic mean.

Diskussionspapiere 2009
Discussion Papers 2009

01/2009 **Addison, John T. and Claus Schnabel:** Worker Directors: A German Product that Didn't Export?

02/2009 **Uhde, André and Ulrich Heimeshoff:** Consolidation in banking and financial stability in Europe: Empirical evidence

03/2009 **Gu, Yiquan and Tobias Wenzel:** Product Variety, Price Elasticity of Demand and Fixed Cost in Spatial Models

04/2009 **Schlüter, Stephan:** A Two-Factor Model for Electricity Prices with Dynamic Volatility

05/2009 **Schlüter, Stephan and Fischer, Matthias**: A Tail Quantile Approximation Formula for the Student t and the Symmetric Generalized Hyperbolic Distribution

06/2009 **Ardelean, Vlad:** The impacts of outliers on different estimators for GARCH processes: an empirical study

07/2009 **Herrmann, Klaus:** Non-Extensitivity versus Informative Moments for Financial Models - A Unifying Framework and Empirical Results

08/2009 **Herr, Annika:** Product differentiation and welfare in a mixed duopoly with regulated prices: The case of a public and a private hospital

09/2009 **Dewenter, Ralf, Haucap, Justus and Wenzel, Tobias**: Indirect Network Effects with Two Salop Circles: The Example of the Music Industry

10/2009 **Stuehmeier, Torben and Wenzel, Tobias**: Getting Beer During Commercials: Adverse Effects of Ad-Avoidance

11/2009 **Klein, Ingo, Köck, Christian and Tinkl, Fabian**: Spatial-serial dependency in multivariate GARCH models and dynamic copulas: A simulation study

12/2009 **Schlüter, Stephan:** Constructing a Quasilinear Moving Average Using the Scaling Function

13/2009 **Blien, Uwe, Dauth, Wolfgang, Schank, Thorsten and Schnabel, Claus:** The institutional context of an "empirical law": The wage curve under different regimes of collective bargaining

14/2009 **Mosthaf, Alexander, Schank, Thorsten and Schnabel, Claus:** Low-wage employment versus unemployment: Which one provides better prospects for women?

Diskussionspapiere 2008
Discussion Papers 2008

01/2008 **Grimm, Veronika and Gregor Zoettl:** Strategic Capacity Choice under Uncertainty: The Impact of Market Structure on Investment and Welfare

02/2008 **Grimm, Veronika and Gregor Zoettl:** Production under Uncertainty: A Characterization of Welfare Enhancing and Optimal Price Caps

03/2008 **Engelmann, Dirk and Veronika Grimm:** Mechanisms for Efficient Voting with Private Information about Preferences

04/2008 **Schnabel, Claus and Joachim Wagner:** The Aging of the Unions in West Germany, 1980-2006

05/2008 **Wenzel, Tobias:** On the Incentives to Form Strategic Coalitions in ATM Markets

06/2008 **Herrmann, Klaus:** Models for Time-varying Moments Using Maximum Entropy Applied to a Generalized Measure of Volatility

07/2008 **Klein, Ingo and Michael Grottke:** On J.M. Keynes' "The Principal Averages and the Laws of Error which Lead to Them" - Refinement and Generalisation