Cilt: 07 Sayı: 01 Kategori: 02 Sayfa: 18-40 ISSN No: 1303-6319 Volume: 07 Number: 01 Category: 02 Page: 18-40

PRIOR DISTRIBUTION CLASSES WITH COMPREHENSIVE COVERAGE

Oya EKİCİ [*]

ABSTRACT

The Bayes' theorem which is the kernel of today's Bayesian world incorporates prior knowledge in analysis. Regarding its level, form or application restrictions, the challenging part can be seen as "prior" especially for joiners in this world. In various areas, concerning the requirements, there are various prior distributions suggested to be used. However the studies that give a generic look and review on prior distributions classes are not seen in the literature. With this motivation, the paper discusses prior distributions with comprehensive coverage. Thus it's aimed to introduce prior distribution classes and to give a review on them.

Keywords: Classes of prior distributions, Informative priors, Non-informative priors.

1. INTRODUCTION

Bayesian literature is imperiously growing either due to pragmatic or conservative reasons. But whatever the reason is, the developments imply that the method works, and is found useful, and gathers most of the efforts on this field.

Certainly, science has various questions. In statistical methods as a tool of scientific investigations, executing "objective" estimation and analysis process is probably one of the most critical one of these questions. Bayesian methods by permitting prior beliefs to be involved with a proper way rather than ad hoc manners-inevitably as being in almost all empirical studies-accomplish the objectivity goal indeed. However it should be said that Bayesian principle is criticized also just for this reason, that is "not being objective". If these critics are moved through philosophical basis, another aspect can be proposed here. In Bayesian statistics, probability is induction probability and the purpose is to attain to highest probability; namely accuracy. Actually, impossibility of achieving "perfect information" justifies adopting the induction method within Bayesian philosophy. So, it is reasonable to update the probabilities by following justification process in each step of it. Regarding the beginning of the mentioned process, there is no need for initial assumptions in Bayesian approach. Instead prior information is employed here. The processes followed by Bayesian and Classical approaches can be summarized with the table below.

[*] Res. Ass., Faculty of Economics, Istanbul University, Istanbul. e-mail:oyaekici@istanbul.edu.tr

Table 1. Philosophically, probability processes for Bayesian and Classical approaches

Bayesian approach	*Classical approach*
Without assumptions	With assumptions
⇩	⇩
Trials	Trials
⇩	⇩
Justification	Falsification
⇩	⇩
$\frac{1}{2} << p(t,g) < 1$	$p(t,g) = 0$

where "*t*" is a theory and "*g*" is the relevant observations. As Popper (2003) also asserts, Classical approach follows deduction method as seen from the table above. But the point here is that Classical approach begins with initial assumptions, and a part of subjectivity penetrates to analysis from there in a way. Hence, on the contrary, it can be claimed that these assumptions in fact distort the objectivity goal. It is out of purpose to go through a further discussion on the distinctions between them from philosophical aspect. As being in Bayesian literature, independently prior distributions themselves can be discussed; whether they lead an objective analysis or not. However, using prior information for the analysis incontrovertibly constitutes Bayesian approach's original part. Due to its demanding nature in describing and fitting to data, "prior information" notion evolves as a fundamental issue within the Bayesian context and this key role leads the studies to effort on this area of the literature. A useful study on prior distribution, Kass and Wasserman (1996), presents a search on selecting prior distribution.

The most challenging part can be seen as "prior" especially for joiners. Owing to all motives mentioned above, to see the whole picture, an extensive study of prior distributions that summarizes the literature has been aimed in this paper. And the subject is generically presented with classification rather than a whole discussion. There is no granted end in discussion of which prior is accurate to employ for most of modeling issues. Hence instead guiding which one to be used, a summary for almost all class of priors is tried to be given.

In this paper, the classes are given under two general titles; non-informative and informative prior distributions. The basis of this categorization and the other proposed classifications are briefly discussed in Section 2. The rest of it is organized accordingly. Section 3 is about non-informative prior classes. Section 4 is on informative prior distribution classes. In the final section some concluding remarks are provided.

2. ON THE CLASSIFICATIONS OF PRIOR DISTRIBUTIONS

As Bayes' theorem refers, the prior distribution of a parameter is combined with the probability distribution of current related data to obtain the posterior distribution which is then used for making inferences about the parameter. In the analysis with this principle, prior distributions are the tools to reflect the information that the researcher has. While determination of the tool, namely while choosing or constituting prior distribution, we meet a series of types and classifications that will shed light on these choices. Diaconis and Ylvisaker (1985) divide Bayesians by considering prior distributions that is preferred to use. So the prior distribution classification done accordingly is as in the following.

Classical Bayesians approve choosing non-informative priors as flat or uniform prior. Modern Parametric Bayesians choose conjugate priors which have designed properties. Subjective Bayesians use elicited priors those especially gathered from an expert opinion or impressions from a similar area. (Gill, J., 2002, p.114). Indeed, this classification done by Diaconis and Ylvisaker (1985) is not clear in practice both for the priors and for the approaches. Another classification is given by Gelman (2002); non-informative, highly informative and moderately informative prior distributions, even his study's main focus is other than making classification.

Actually, as a loosening factor for any grouping like mentioned above, it would be needed to express the degree of belief in the prior information of each parameter separately. For this reason, there are methods which make possible to assign some parameters as informative and some others as non-informative in the same prior distribution. For categorical data, Demirhan and Hamurkaroğlu (2008) propose a method that one can represent one's degree of belief in prior information on each of log odds ratios separately and also show that representation of very weak belief in prior information on the relevant parameters using this approach is successful.

Some characteristic of prior distributions can be considered as a class. For instance realistic prior distribution, the name "realistic" here reflects the characteristic of prior distributions rather than being a class of them. If a prior is realistic, it successfully represents the uncertainty on prediction in a realistic way. For this reason, realistic prior improves forecast accuracy. Similarly, improper prior distribution also can be considered as a class, (mentioned below under non-informative priors) it is again a characteristic of prior distributions, essentially concluding probability of the relevant parameter is infinite. In the paper, prior distribution classes are referred considering these principles. And a conventional classification done for prior distributions is adopted; non-informative and informative prior distributions. For this reason, the categorization reflects probable priors' attitudes with general insight helps to capture classes' diversities and to clarify the transition between them. However, their behaviors can change for some cases, these situations -when non-informative prior turns out to informative or informative prior turns out to non-informative- are enlightened in the related part of the paper.

3. NON-INFORMATIVE PRIOR DISTRIBUTIONS

They are poor to explain the parameter, but these kinds of priors eliminate the subjectivity criticism of Bayesian view. Though, in general it should be said that none of the non-informative priors reflects the ignorance. And in some circumstances, its non-informative trait becomes informative.

With a different aspect Berger (1985, p. 406-409) noted, there seem alternatively two viewpoints to handle the problem; non-informative prior approach and invariance approach. He remarked that the studies on invariance suggest reasonable choice for the non-informative prior, namely the right invariant Haar density, the right Haar measure, gives the best invariant decision rule. Besides many relevant advantages of invariance, non-informative prior approach is preferred most of the time.

The priors that might be located under non-informative are flat (or uniform) priors, Jeffreys prior, reference prior, diffuse (vague, weak or locally uniform) prior, maximum entropy prior, intrinsic prior and integral prior distributions.

3.1 Uniform Prior (Flat Priors) Distributions

In case of using uniform prior, the same probability values are assigned to the parameter for a determined interval. While $p(\theta) = c = 1/m$ and $0 \le \theta \le m$, at every point of determined interval, the probability density of the parameter is equal to "*c*".

This is attributed as Bernoulli named "principle of insufficient reason". If there is no reason to believe that any one of these is more likely to be true than another, then we should assign the same probability to all (Sivia, 1996, p.106-7, 120).

Generally when the parameter lies within a specific interval, when it is able to be limited, and when it's a proportion as its nature, this class of prior can be employed. It can not be said that it ensures the situation of "Ignorance". For instance, if we say the parameter belongs to the interval [0 - *m*], as *m* goes to infinity, prior distribution becomes less informative. However in this situation, the probability of parameter $p(\theta)$ converges to zero. No value of θ will be increasingly probable.

On extended real line $[-\infty, \infty]$, for all values of θ, while $p(\theta) = c$, the uniform priors are improper. That means when the probability of pdf (probability density function) is integrated, the result is infinite and violates the axiom of "probability sum equals to 1", whereas it equals to "1" in proper distributions. Improper priors have computational difficulties. But it should be remarked here that the posterior distribution which is derived from this kind of priors doesn't have to be improper.

Other drawback of uniform prior is not possessing invariance property. When the parameter is transformed, the new prior distribution derived from it might not be uniform. It might lose non-informative characteristic and might violate its equal probabilities feature. Uniform priors have strong sides as well, as the sample size increases the effect of settling on uniform distributions becomes slight. Nuisance parameters are easily reduced from the posterior distribution after it's integrated out. Additionally, Gill (2002, p.121-3) pointed out some conjugate prior distributions become the same with uniform priors in the limit.

3.2 Jeffreys Prior Distribution

Jeffreys's philosophical point of view leans on the concepts of "necessarianism" or objectivism. For his objectivist view he believes in the state of ignorance and based his view on "principle of insufficient reason". He thinks that there should be an explanation for an event being more or less probable if the probabilities are not assigned equal. It can be said that the idea of basing on a definite reason is the extension of necessarianism. Besides according to him, it's not essential to represent the ignorance with merely one prior distribution (Kass and Wasserman, 1996).

With the motivation mentioned above, Jeffreys makes prior distribution equal to a constant, whether parameter is restricted with a specific interval ($[-\infty,\infty]$ or $[0,\infty]$) or not, which means the defined uniform prior distribution has the form improper. In his studies' further step, he uses Fisher information matrix, supposing normal distribution;

$$I(\theta) = -E_\theta\left[\frac{\partial^2 \log f(y \setminus \theta)}{\partial\theta^2}\right] \quad (1)$$

$$f(y \setminus \theta) = \sigma^{-1} f\left(\frac{(y-\theta)}{\sigma}\right), \quad (\theta \in R \quad and \quad \sigma \rangle 0)$$

Here to cover all possible results, with most general form, a location-shape density function that involves location and shape parameters are determined as likelihood function. Information matrix derived from this kind of likelihood function is as in the following (Berger, 1985, p.88);

$$I(\theta) = -E_\theta\begin{bmatrix} \frac{\partial^2}{\partial\theta^2}\left(-\log\sigma - \frac{(y-\theta)^2}{2\sigma^2}\right) & \frac{\partial^2}{\partial\theta\,\partial\sigma}\left(-\log\sigma - \frac{(y-\theta)^2}{2\sigma^2}\right) \\ \frac{\partial^2}{\partial\theta\,\partial\sigma}\left(-\log\sigma - \frac{(y-\theta)^2}{2\sigma^2}\right) & \frac{\partial^2}{\partial\sigma^2}\left(-\log\sigma - \frac{(y-\theta)^2}{2\sigma^2}\right) \end{bmatrix}$$

$$= -E_\theta\begin{bmatrix} -\frac{1}{\sigma^2} & \frac{2(\theta-y)}{\sigma^3} \\ \frac{2(\theta-y)}{\sigma^3} & \frac{1}{\sigma^2} - \frac{3(y-\theta)^2}{\sigma^4} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix} \quad (2)$$

Prior distribution is the square root of the determinant of this derived information matrix: $\pi(\theta) = det(I(\theta))^{1/2}$. If the result stated above is expanded with some operations;

$$\pi(\theta) = \left(\frac{1}{\sigma^2}\cdot\frac{2}{\sigma^2}\right)^{1/2} \text{ and } \pi(\theta) \propto \frac{1}{\sigma^2} \quad (3)$$

prior distribution reaches finally to the form above. When the likelihood function is a pdf in which just has location parameter as (θ), the computed information matrix becomes equal to a constant. Hence the prior distribution becomes equal to a constant, too.

The main argument of Jeffreys in preferring this kind of prior distribution is possessing invariance property against power transformation of the parameter, as Hartigan (1964) proposed;

$$\pi_\gamma(\gamma)\cdot\left(\frac{d\gamma}{d\theta}\right)=\pi_\theta(\theta), \qquad \gamma=h(\theta) \tag{4}$$

While $\gamma=\sigma^n$, the following equations and proportion are attained; $d\gamma=n.\sigma^{n-1}d\sigma$, $\frac{d\gamma}{\gamma}=\frac{n.\sigma^{n-1}d\sigma}{\sigma^n}$, $\frac{d\gamma}{\gamma}\propto\frac{d\sigma}{\sigma}$. So, the invariance property for prior distribution is proved by considering variable transformation formula. By the way a model can be parameterized in terms of standard deviation, variance and precision parameters.

In Jeffreys prior which is produced from Fisher information matrix, logarithm of posterior distribution yields the exponent -that involves the parameters- of the distribution. The multi-differential of it gives the marginal of parameters (and leave parameter alone), then the expectation of this attained parameter's function or value is taken. In geometrical aspect, taking the determinant of Fisher information matrix that involves all parameter information, gives the region (volume or hyper volume) that vectors of parameter span in the parameter space. By the way, regarding the existing parameter information, in parameter space, the probability region of parameter is determined.

3.3 Reference Prior Distribution

Bernardo, J. M. (1979) nominated a prior distribution, "reference prior". The term reference prior distribution mostly has been used in a narrow sense. For Box and Tiao (1992, p. 22-23) and most of others reference prior is convenient to use as a standard and it has the form dominated by the likelihood function. Besides if there is dominant likelihood, the attained posterior distribution is still a convenient distribution. Actually the form in which the prior distribution dominates is also possible. So in case of dominant prior, it can be said that the yielding form is informative. In this context, a reference prior doesn't have to be a non-informative prior distribution.

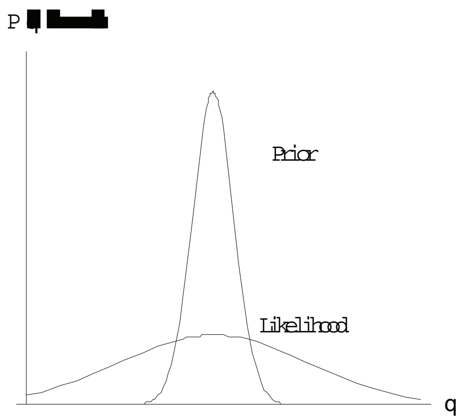

Figure 1. Dominant Prior

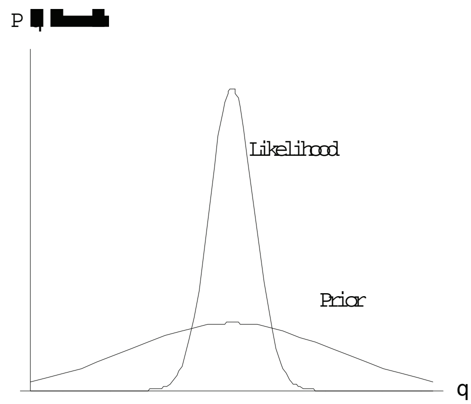

Figure 2. Dominant Likelihood

Berger (1985) and Bernardo (1979) have used "reference prior distribution" in general sense. With Bernardo's words (1996), the reference prior is that which maximizes the missing information. According to the method that they developed reference prior for relevant parameter can be found by maximizing the Kullback-Leibler distance between prior and posterior distribution.

The i.i.d. random variables $Y_1^n = (Y_1, ..., Y_n)$ are seen in the form of Kullback-Leibler distance $K_n\left(\pi\left(\theta \setminus y_1^n\right), \pi(\theta)\right)$. Where $\pi\left(\theta \setminus y_1^n\right)$ is posterior, $\pi(\theta)$ is prior distribution. Accordingly, Kass and Wasserman (1996) provided,

$$K_n\left(\pi\left(\theta \setminus y_1^n\right), \pi(\theta)\right) = \int log\left[\frac{\pi\left(\theta \setminus y_1^n\right)}{\pi(\theta)}\right]\pi\left(\theta \setminus y_1^n\right)d\theta \qquad (5)$$

Here the logarithm of the ratio of these two distributions is taken and multiplied with posterior distribution. $K_n^\pi = E\left(K_n\left(\pi\left(\theta \setminus y_1^n\right), \pi(\theta)\right)\right)$ is the expected value of Kullback-Leibler distance. The purpose here is to maximize the expression; $K_\infty^\pi = lim_{n\to\infty} K_n^\pi$. But in general K_∞^π is infinite and to handle this problem firstly the prior distributions (π_n) that will maximize a finite distance K_n^π are attained. Then limit value of the posterior distributions calculated with this prior distributions are attained. As a result, the prior distribution attained in this manner becomes a distribution that yields a posterior in limit with Bayes theorem. After some simplifications, this prior distribution (for a continuous parameter space) turns into Jeffreys prior (Kass and Wasserman, 1996). Moreover, as Bernardo (1996) also noted, in one-parameter problems, the reference prior reduces to Jaynes (1968) maximum entropy prior if the parameter space has a finite number of points, and it reduces to Jeffreys' prior in the continuous regular case.

Reference prior is successful in handling nuisance parameter. More specifically, while w is the parameter of interest and λ is nuisance parameter, if it is differentiated as $\theta = (w, \lambda)$, reference prior is well enough in making inference about θ. When there is not nuisance parameter and some specific conditions hold, reference prior becomes the same with Jeffreys prior. That means if there is "parameters of interest" and "nuisance parameters" distinction, this method gives different results than Jeffreys' (Kass and Wasserman, 1996).

As to computation of reference priors in practice, Bernardo (1996) expresses that reference priors only depend on the model through its asymptotic behavior; essentially, if the asymptotic of a model is known, than its associated reference priors may be easily found. Under regularity conditions for asymptotic normality, any reference prior may be obtained from a relatively simple algorithm in terms of Fisher's matrix (Berger and Bernardo, 1992). Though, for non-regular or complex models, the derivation of reference priors may be a difficult mathematical problem.

3.4 Diffuse Prior Distribution

Some other denominations seen in the literature of this kind of prior are "vague prior", "weak prior" or "locally uniform prior". As is known, depending on a parameter's (thinking as a random variable within Bayesian principle) distribution, the parameter that determines the scale of this distribution changes. For instance, for location and shape parameter respectively normal distribution and gamma distribution can be assumed. Based on the idea of large variances means uncertainty, such a large value assigned for the shape parameter of this distribution that extending on considerably wide interval, makes the distribution almost flat as uniform prior and non-informative prior constitutes (Raiffa and Schlaifer, 1968, p.63).

$p(\theta) \sim N\left(\overline{\mu}, \overline{\sigma}^2\right)$, large values are assigned to $\overline{\sigma}^2$.
$p(\sigma) \sim G(\alpha, \beta)$, very small values are assigned to α and β (as 0,001).

The strongest argument of diffuse prior distribution attained as above is being a proper distribution. Assuming diffuse prior distribution instead of uniform prior is assuring the advantage of easy calculation.

Diffuse prior distributions have also weakness. Let's assume likelihood function with a broad peak, the situation of the posterior distribution is improper and number of observations is small. Even diffuse prior distribution used with a normal distribution - has a large variance- it doesn't produce good solutions since the obtained posterior distribution becomes sensitive to the new prior distribution (Kass and Wasserman, 1996). Actually the situation of limited sample size is the mostly encountered problem in applications.

3.5 Maximum Entropy Prior Distribution

Entropy is a term that belongs to physics discipline. It exists in statistics, since it has direct relationship to information theory and in a sense measures the amount of uncertainty inherent in the probability distribution (Berger, 1985, p.91). In other words it quantifies the uncertainty of observations.

Entropy prior introduced by E.T. Jaynes (1968) identifies relative level of uncertainty about prior distribution parameters' distributions. Similarly uncertainty or precision that is supplied by different prior distribution are modeled and so it may be said that it is flexible. But transformation result of parameter doesn't have invariant property. So its application is limited (Gill, J., 2002, p.135-6).

For a discrete parameter $\theta = (\theta_1, \theta_2, ..., \theta_n)$, when entropy is $H(\theta)$;

$$H(\theta) = -\sum P(\theta_i).log\, P(\theta_i) \tag{6}$$

Here by taking logarithm of probability, with the intuition of information exponentially increases, makes the information monotonically increase. Negative sign represents "ignorance" as a reverse of information. Two extreme situations can be identified with the equation above. One signifies entropy's minimum (the situation of full information), other signifies entropy's maximum (the situation of ignorance).

In the first case, while $P(\theta_k) = 1$ that is when the probability of the parameter having a particular value is "1", and $i \neq k$, it's stated as $P(\theta_i) = 0$, means it is assigned "0" for the probability of this parameter having any other value. By the way full information is given about the parameter and so ignorance being "0",

$$H(\theta) = -\sum_{i=1}^{n} P(\theta_i) log\, P(\theta_i) = 0 \tag{7}$$

In the second case, if the prior distribution is specified as uniform prior, say $P(\theta_i) = \frac{1}{n}$ is assigned; the entropy of this prior reaches its maximum.

$$H(\theta) = -\sum \frac{1}{n} log\left(\frac{1}{n}\right) = log\, n \tag{8}$$

Entropy given above is identical with the non-informative prior distribution. As the observed value of parameter increases ($n \to \infty$), the value of $P(\theta_i) = \frac{1}{n}$ decreases (Actually, ever more it turns to a situation that no value of parameter will be possible, namely in limit $P(\theta_i) = 0$. So it is a contradiction to reach an informative form. However, defining θ as finite can make one avoid this contradiction). Also regarding these two forms, entropy constitutes the restriction $H(\theta) \leq log\, n$ for prior distribution (Berger, J.O., 1985, p.91).

3.6 Intrinsic Prior – Integral Prior

In Bayesian analysis, inference is based on posterior distribution. Here in model selection case, it's not satisfactory to compare models through their individual posterior distributions. Because different models have different unobservable parts. Instead, the performance of model is evaluated by leaning on directly probability of data. Model selection as a common concern brings further considerations in assigning prior distribution. When needed to use non-informative structure, since the standard non-informative priors are improper, prior evaluations give a ratio of constants, that is, again a constant. Within Bayesian principle, the Bayes Factor is multiplied with this arbitrary constant. For the models M_1 and M_2, the posterior probabilities ratio;

$$\frac{P(M_2 \setminus y)}{P(M_1 \setminus y)} = \frac{P(y \setminus M_2)}{P(y \setminus M_1)} \cdot \frac{P(M_2)}{P(M_1)}$$

$$\frac{P(M_2 \setminus y)}{P(M_1 \setminus y)} = B_{21}(y) \cdot \frac{P(M_2)}{P(M_1)} \quad (9)$$

Bayes factor B_{21} is described as below, and where θ_1 and θ_2 are the parameters of interest for the data y,

$$B_{21} = \frac{m_2(y)}{m_1(y)} = \frac{\int f_2(y \setminus \theta_2)\pi_2(\theta_2)d\theta_2}{\int f_1(y \setminus \theta_1)\pi_1(\theta_1)d\theta_1} \quad (10)$$

and there say $m_2(y)$ is the marginal density of y for M_2. Here B_{21} equals to an arbitrary constant in case an improper prior assignment. To deal with the problems Berger and Pericchi (1996) introduced "intrinsic prior". It gives a Bayes factor free of arbitrary constants and tends to correspond to actual Bayes factor. There is no need to compute training samples[√], on the contrary, the process with intrinsic priors' implement imaginary training samples in a sense. $\pi_1^I(\theta_1)$ and $\pi_2^I(\theta_2)$ denote intrinsic priors and are handled by solving functional equations below (Cano et al, 2004, p.446-7);

$$E_{\theta_1}^{M_1} B_{12}^N(y(\ell)) = \frac{\pi_2^I(\psi_2(\theta_1))}{\pi_2^N(\psi_2(\theta_1))} \frac{\pi_1^N(\theta_1)}{\pi_1^I(\theta_1)} \text{ and}$$

$$E_{\theta_2}^{M_2} B_{12}^N(y(\ell)) = \frac{\pi_2^I(\theta_2)}{\pi_2^N(\theta_2)} \frac{\pi_1^N(\psi_1(\theta_2))}{\pi_1^I(\psi_1(\theta_2))} \quad (11)$$

where $y(\ell) = (y_1, ..., y_\ell)$ is a minimal training sample, a random vector of minimal size ℓ such that $0 < m_i^N(y(\ell)) < \infty$, $i = 1,2$, $\psi_2(\theta_1)$ denotes the limit of the MLE $\hat{\theta}_2(y)$ under M_1 at point θ_1, and $\psi_1(\theta_2)$ the limit of $\hat{\theta}_1(y)$ under model M_2 at point θ_2. Here the expectations above are taken with respect to $f_1(y(\ell) \setminus \theta_1)$ and $f_2(y(\ell) \setminus \theta_2)$. To sum up, there is possibility to get solutions which is not unique from these equations in nested models, however, as Cano et al (2004) and Cano et al (2007) demonstrates, in non-nested models solutions are exactly not unique. Another problem in non-nested models is that intrinsic prior cannot deal with improper solutions while intrinsic priors are well established for nested models. Prior construction for non-nested models are developed by Cano et al (2007) named integral priors. The integral equation systems (Cano et al, 2007, p.60) offered to be solved is as below;

[√] a sample constituted just by using a part of the data.

$$\pi_1(\theta_1) = \int_y \pi_1^N(\theta_1 \setminus y) m_2(y) dy \text{ and}$$
$$\pi_2(\theta_2) = \int_y \pi_2^N(\theta_2 \setminus y) m_1(y) dy \qquad (12)$$

where $\pi_1^N(\theta_1 \setminus y)$ and $\pi_2^N(\theta_2 \setminus y)$ are posterior distributions for the data y. They found that under some assumptions, integral priors are unique. Further discussions on the subject are beyond the purpose of this paper.

4. INFORMATIVE PRIOR DISTRIBUTIONS

In some cases, it is inevitable to use information, for instance in the political science, information needs to come in, whether as regression predictors or regularization (that is, prior distributions) on parameters (Gelman, 2009). Here the problem is how to form this prior information. Zellner (1971) described the nature of prior information with a classification; data-based and non-data-based. When the past data is available for the study, it is termed as data-based prior. Even if it is not possible to employ past data, researcher's personal observations or theories can be regarded as alternative sources. This kind of prior information is termed as non-data-based prior. In case of having specific parameter prior knowledge or a restriction for it, the analysis becomes easier. For this reason, the dangers seen during the modification of prior beliefs is vanished. Conversely, in some other cases the researcher can be under the situation that should be eliciting the prior (distribution, parameter). All these efforts are intentionally to impose the prior information to the analysis.

For some cases non-informative structure turns out to be informative. In time series modeling prior distribution appears as a subject to be cautious in this sense. Ekici and Yorulmaz (2008) briefly discussed the issue; uniform prior distribution of which parameters are assigned the same probability values for a given interval, Jeffreys prior obtained from information matrix and diffuse prior don't give dissimilar results. Besides each of the three distributions are non-informative. They also produce similar results with Classical approach. However, this is not so for non-stationary time series. In non-stationary time series, even a non-informative prior employed, Bayesian and Classical approaches produce different results. Sims (1988) and Zellner (1971, p.186) employed uniform prior for a non-informative prior. But Phillips (1991) notifies that uniform prior turns out to have informative form and suggests Jeffreys prior or diffuse prior in case of non-informative prior distribution is preferred. For this reason, sample about autoregressive parameter in different intervals provides more information. Thus, choosing uniform prior implies making all values of this parameter equally likely and reflects ignorance actually despite of having slight weights but large values of the parameter (Maddala and Kim, 2002, p.266). So in time series context prior distribution preference is primary issue as well.

Under informative prior distribution section, conjugate prior, subjective prior, Minnesota prior, power prior, g-prior, first-difference prior and second-difference prior distributions are referred.

4.1 Conjugate Prior Distributions

Conjugate prior distributions can be determined on the bases of distribution structure of the likelihood function. The likelihood function is divided into its multiplications in terms of sufficient statistics, and this process named as “Neyman’s Factorization Theorem” (Lindley, 1965, p.47, 50). Resulting sufficient statistics’ distributions become a base to conjugate prior distribution that will be constituted.

Some of the reasons about determining a conjugate prior as a prior distribution are given by Raiffa and Schlaifer (1968, p.44) as follows; it should be easy to obtain the posterior distribution from the determined prior distribution and likelihood function. Conjugate prior distribution should be rich. What’s meant by rich is existence of a distribution element that is useful in expressing researcher’s prior information and beliefs.

Beyond these reasons, from the perspective of proposed utility, it is always possible to get new sample information from the same space by preferring to use conjugate prior. By the way the prior information about allied parameter can be explored and obtained more consistent results (Yardımcı, 1992). Prior distribution that is chosen on the bases of sample distribution family has algebraic convenience to join with likelihood. Particularly, it’s easy to compute for the exponential distribution family since probability distributions that belong to an exponential family have natural conjugate prior distributions (Gelman et, 1995, p.38). Furthermore conjugate prior distribution is a proper pdf.

However, conjugate prior should be handled with care. Since this prior distribution is evidence for very specific parametric prior knowledge (Gill, J., 2002, p.120).

So it is critical which conjugate prior will be used with different likelihood functions (Raiffa and Schlaifer, 1968, p.53-4). The generated posterior distribution is required to have a known functional form. This can be exemplified in Table 2 below.

Table 2. Conjugate prior distributions for likelihood functions and yielding posteriors

Likelihood	Prior	Posterior
Binomial	Beta	Beta
Negative Binomial	Beta	Beta
Normal	Normal	Normal
Poisson	Gamma	Gamma
Exponential	Gamma	Gamma
Gamma	Gamma	Gamma

At this point there appear two definitions about conjugate priors; “natural conjugate prior” and “conjugate prior”. When conjugate prior combines with likelihood, yielded posterior again has the same distribution class with conjugate prior. Alternatively, when natural conjugate prior combines with likelihood, yielded posterior has the same distribution with natural conjugate prior distribution, and also has the same distribution with likelihood (Raiffa and Schlaifer, 1968, p.48-49).

As an example for the factorization procedure of conjugate prior of θ, suppose $P(y \backslash \theta)$ is likelihood function in form of binomial distribution;

$$P(y \backslash \theta) \propto \theta^{y} (1-\theta)^{n-y} \qquad \text{for } 0 \le \theta \le 1,$$

Say, if *y=a* and *n-y=b*, sufficient statistics would be here *a* and *b*. Through the factorization, conjugate prior distribution for $p(\theta)$ is (in Beta form);

$$P(\theta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1} \qquad \text{for } 0 \le \theta \le 1, \tag{13}$$

with Bayesian principle,

$$P(\theta \backslash y) \propto \theta^{y} (1-\theta)^{n-y} . \theta^{\alpha-1} (1-\theta)^{\beta-1} \tag{14}$$

yielding posterior density $P(\theta \backslash y)$ again reaches Beta distribution;

$$P(\theta \backslash y) \propto \theta^{y+\alpha-1} (1-\theta)^{n-y+\beta-1} \tag{15}$$

Naturally most of the distributions can be appointed as a prior; however they all do not emerge as one of prior classes. Since having a special weight among the others, the two of distributions, Wishart and Dirichlet distributions, are regarded here as a class of prior distributions under conjugate priors.

4.1.1 Wishart Prior

Under the assumption of the unknown variance case, Wishart distribution can be assigned as prior. That is Wishart prior is the conjugate prior distribution for the inverse covariance matrix under a multivariate normal distribution assumption for the parameter say in a linear model. Similarly, Inverse-Wishart prior is the conjugate prior distribution at this time for the multivariate normal covariance matrix.

4.1.2 Dirichlet Prior

Dirichlet distribution is a multivariate generalization of the beta distribution. So the prior established with this distribution, called as Dirichlet prior, is the conjugate prior distribution for the parameters of the multinomial distribution.

4.2 Subjective Prior Distributions

In Bayesian approach, objectivity-subjectivity debate begins with assessing of probability and is earlier than prior distribution specification. De Finetti who supports subjectivity defines probability from different aspects. One is built on odds ratio, one another is score rule based on punishment (Galavotti, 2001, p.161-3). Quantifying of uncertainty is complicated. So as in determination of probability, while subjectively determining prior distribution, also numerous methods that the study entails can be employed.

Most of the time, since the process of data analysis typically involves a host of subjective choices, there exist problems about being objective during these statistical analysis (Berger, 2006). The distinction between objective and subjective analysis in Bayesian approach then appears with prior distribution specification for parameters of statistical model. So this signs its key role again. If the phrase “objective Bayesian analysis” corresponds too many investigators in this field, as a general agreement on objectivity, Jeffrey’s prior, reference prior and maximum entropy prior are said to be objective priors. In fact there are views standing out against objective Bayesian analysis. One critique to objective Bayesian school done by Wasserman (2006) is that he interpreted “objective Bayesian inference” to mean “Bayesian methods that have good frequency properties” and refer to these methods as “Frequentist-Bayes”. But going through the views on this debate here would take the purpose of the paper far afield.

Depending on the nature of subject, eliciting prior would be needed. Within the context of statistical analysis, expert opinion provides the structure that Bayesian analysis needed where the prior parameters cannot be based on previous studies and data or where they don’t exist. The difficulties during elicitation process are quantifying the uncertainty and reflecting personal judgment to a parameter, an interval or a distribution. It’s obvious that the process bias and various methods can be employed in elicitation process for several empirical studies. But it is said to be the elicited prior conceptually and practically is the closest prior to definition of subjective prior than the other informative prior classes. Here in this section some of the subjective prior distributions are brought up.

4.2.1 Histogram Approach

When θ represents an interval on real line, histogram method can be used for prior specification. In this method, θ is divided into intervals and subjective probabilities are assigned to each interval. Then histogram plot that falls upon these probabilities is done. This histogram provides prior probability density, $p(\theta)$. There is no rule for the number or width of interval. Moreover prior distribution can be in a challenging form for the study. These are the weaknesses of the method. Another weakness of it is the probability of not having a tail of this constituted prior distribution. This means extreme values may not fall upon any quantity in prior distribution (Berger, 1985, p.77).

4.2.2 Relative Likelihood Approach

When θ is a subset of a real line, this method is used. Plotting is done on the basis of intuitively how many times one parameter value probable to another. Then prior distribution can be specified from this draft.

More parameter values can be involved in more sensitive analysis during comparison. Produced prior distribution may not be a proper distribution. Consistency of the statement such as "parameter θ has equal probability in having separate values θ_1 and θ_3"(in the figure above as well) should be checked (Berger, 1985, p.78).

4.2.3 Matching a Given Functional Form

A functional form is defined depending on researcher's opinion. Then the distribution that is the best fit to this functional form is chosen. For instance, a functional form as in Figure 3 brings the idea of Gamma distribution for prior distribution representation, and the researcher makes the assumption of Gamma distribution $G(\alpha, \beta)$. In this step, α and β parameters of distribution of Gamma prior are subjectively specified. It has several ways. One way is calculating from the estimated prior moments. That is, mean and variance are computed from specified functions. Then, α and β values are achieved from the equalities; $Mean = \alpha.\beta$ and $Variance = \alpha.\beta^2$. Another way is dividing the prior distribution into its fractals (Berger, 1985, p.80-1). Researcher assigns the values to these divided fractiles leaning on his/her subjective beliefs. The probability areas are computed that fall upon these fractiles.

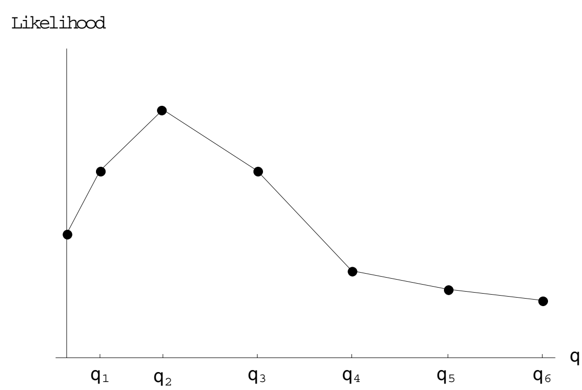

Figure 3. A Draft Plot for Relative Likelihood

4.3 Hierarchical Prior Distributions

One way to handle the uncertainty on prior parameters' values is to assign an additional prior distribution to these parameters. This additional prior distribution's parameters are known as hyper parameters. For this reason, hierarchical prior distribution is called as hyper prior distribution in some relevant texts. Formation of hierarchical prior distribution can be described here with an example (Gill, J., 2002, p.354-5).

While y_i are observations, it's assumed that these observations are Poisson distributed. In the distribution, the parameter is λ. When there is a need to specify a prior for λ, Gamma distribution can be selected since the Gamma distribution is the conjugate of Poisson and has a flexible parameter form as well. Gamma prior distribution is also defined by the two parameters α and β. Gamma parameters α and β (like in λ) are restricted with positive real numbers. For the parameters α and β assigning Gamma distribution is also reasonable. Hyper parameter values (*A, B, C, D*) for Gamma distribution is set to begin the process. Reviewing of the process;

$$y_i \sim Poi(\lambda_i) \quad \Rightarrow \quad \lambda_i \sim G(\alpha, \beta) \quad \Rightarrow \quad \alpha \sim G(A,B) \quad \text{and} \quad \beta \sim G(C,D)$$

where y_i are conditionally independent, α and β are assumed independent, too. Actually as Berger (1985, p.107-8) noted, this is just a gradually comprising process that conveniently represents the prior.

If structural information and subjective prior information is available at the same time, it will be easy to set them in stages like above. While the researcher's assigning Gamma distribution to reflect the structural information, he/she also adds the subjective prior information by imposing hyper parameter values to them.

Beyond this, it is possible to assemble the joint distribution multiplicatively using Bayes' law and the definition of conditional probability. Then joint posterior distribution is obtained (Gill, J., 2002, p.355). It can be defined as;

$$p(y,\lambda,\alpha,\beta) = \prod_{i=1}^{n} p(y_i \setminus \lambda_i).p(\lambda_i \setminus \alpha,\beta).p(\alpha \setminus A,B).p(\beta \setminus C,D) \qquad (16)$$

If distributions are defined for *A*, *B*, *C* and *D* despite setting values for them, a stage would be added to hierarchy. As Gill pointed out, there is no constraint about number of stages in defining hierarchical prior distribution. But in practice, more than two stages are rarely useful.

It should be noted here that modeling with hierarchical prior distribution is started to be employed extensively by the improvement of MCMC technique.

4.4 Minnesota Prior Distributions

"Minnesota prior" was introduced by Litterman (1986). This class of prior is one of the beneficial and commonly used informative prior in a Bayesian Vector autoregression (VAR) analysis. Minnesota prior assumes random walk process as in many of time series and has a symmetric form. To begin a brief description for Minnesota prior, a VAR model considered;

$$x_t = A_0 + A_1 x_{t-1} + ... + A_p x_{t-p} + u_t \qquad (17)$$

u_t is white noise disturbances vector, A_0 is vector of intercept terms, $A_1,..., A_p$ are autoregressive matrices, x_t is vector of n variables included in VAR. Here x_t in which the prior mean incorporated with is in the following;

$$x_t = A_0 + x_{t-1} + u_t \qquad (18)$$

In setting diagonal elements of A_1 as “1”, “0” to rest of coefficients, prior belief is imposed and settled on which variables desired in the model. Mean and variance coefficient for prior distribution are represented as below;

$$E(A_1) = 1 \qquad Var(A_1)_{ij} = \lambda \gamma^2 \frac{\sigma_i^2}{\sigma_j^2} \qquad (19)$$

Hereafter two hyper parameters γ and λ become to have key role in specification of Minnesota prior. γ is the tightness parameter of the prior distribution, and gives the measurement of the confidence in prior. That means smaller γ value is higher confidence level. Conversely a high value of this parameter makes a VAR analysis Non-Bayesian. λ is the cross lag term and takes the values between 0 and 1. When λ equals to “1”, this implies the lags of other variables have the same importance with its own lags. Then again as λ approaches to “0”, the model reaches to univariate autoregression form.

4.5 Power Prior Distributions

The power prior is introduced by Ibrahim and Chen (2000). It is a useful general class of priors that can be used for arbitrary classes of regression models, including generalized linear models, generalized linear mixed models semi-parametric survival models with censored data, frailty models, multivariate models, and non-linear models (Ibrahim et al, 2003). But, this class is especially convenient to model selection issue. In constructing power prior, previously observed historical data would be needed. A general discussion of the power prior and process developed for it are sited in examining a regression model for simplicity. For the purpose $D=(n,y,X)$ is the data collected for current analysis. Here n is the number of observation, y indicates the n x 1 vector of dependent variable and X indicates the n x p matrix of independent variables of the regression model. $L=(\theta \backslash D)$ is likelihood function of this model. Let $D_0=(n_0,y_0,X_0)$ be the historical data from previous study and $\pi_0=(\theta \backslash \cdot)$ be prior for θ before historical data been observed. For a given a_0, power prior distribution of θ;

$$\pi(\theta \backslash D_0,a_0) \propto L(\theta \backslash D_0)^{a_0} \,.\, \pi_0(\theta \backslash c_0) \qquad (20)$$

c_0 is the hyper parameter for the initial prior. a_0 helps to signify the importance of the historical data, so for this intention a_0 is defined in the interval $0 \le a_0 \le 1$. When a_0 equals to "1", historical data become to have weight to the extent that it has on likelihood. In contrast when a_0 equals to "0" it makes the analysis free from the historical data and just done with a priori specification. For a full Bayesian analysis a prior specification is done for a_0 as well and with this addition joint power prior distribution gets to the form;

$$\pi(\theta,a_0 \backslash D_0) \propto L(\theta \backslash D_0)^{a_0} \,.\, \pi_0(\theta \backslash c_0) .\, \pi(a_0 \backslash \gamma_0) \qquad (21)$$

γ_0 is hyper parameter vector. As Ibrahim and Chen (2000) suggest, the prior $\pi(a_0 \backslash \gamma_0)$ can be in form of Beta, truncated Gamma or Normal distribution. With this form of it, the power prior has numerous advantageous. Briefly it can be said that it's flexible in weighting historical data, reflects the impact of historical data combining with the prior set up, and constitutes a proper prior. Once the power prior is constructed, then it is easy to generalize to the multiple forms.

$$\pi(\beta,a_0 \backslash D_0) \propto \left(\prod_{k=1}^{L_0} [L(\beta \backslash D_{0k})]^{a_{0k}} \,.\, \pi(a_{0k} \backslash \gamma_0) \right) .\, \pi_0(\beta \backslash c_0) \qquad (22)$$

where $k=1,2,...,L_0$ and L_0 shows historical datasets, a_{0k} denotes powers for these historical datasets.

As a recent corroborative study done by Ibrahim et al (2003) provides a strong motivation for using the power prior in Bayesian inferences, since it indicates that it is an optimal class of informative priors in the sense that it successfully minimizes a convex sum of Kullback-Leibler (KL) divergences between two specific posterior densities, in which one density is based on no incorporation of historical data and the other density is based on pooling the historical and current data.

4.6 g – Prior Distributions

Zellner (1986) proposed a prior that reflects unit-information, named as g-prior. This class of prior provides prior covariance matrix for regression parameters. Here with g

chosen correspondingly the prior variance is proportional to inverse of a design matrix for a data set and specified according to prior information can be imposed to analysis is defined by this *g* constant. For β and σ (that is, for the coefficients of a regression model), a conjugate prior distribution is specified and resulting posterior distribution mean is the weighted mean with *g*. As explicit explanation for the distribution class, the prior is assumed as;

$$p(\beta,\sigma)=p(\sigma).p(\beta\setminus\sigma) \quad (23)$$

and the prior covariance matrix is assumed as,

$$Var(\beta\setminus\sigma)\propto(D'D)^{-1} \quad (24)$$

where D is design matrix for the given data set. As regards conjugate normal prior represented,

$$\beta\sim N\left(0,\,g.\sigma(D'D)^{-1}\right) \quad (25)$$

4.7 First-Difference Prior and Second-Difference Prior Distributions

This class of prior distribution designed for GLMMs (Generalized Linear Mixed Models). In GLMMs two parts occur to reflect fixed and random effects of the linear model. In Bayesian analysis of GLMMs there is no need to define two separate matrices, instead, one matrix is defined as in GLM and these two effects are reflected by means of precision matrix in the prior. In other words, effects are modeled in the prior model. Here as an extension of this subject, there may be possible first-difference and second-difference prior. For first-difference prior which is also called “random walk prior”, if the categories are in a natural order, one can model the prior for parameters just as affected by previous category and a random term;

$$(\beta_k\setminus\beta_1,\beta_2,\ldots,\beta_{k-1})=\beta_{k-1}+e_k \quad (26)$$

If the categories are from 1 to i, $k=2,\ldots,i$ and where $e_k\sim N\left(0,\frac{1}{\theta}\right)$.

Hence if there is an order, the previous category is the most probable one, and this previous category is made to be added to the prior model. No distribution is assigned for β_1, merely matrix structure is specified.

Another prior model structure here is second-difference prior also called as “stochastic-trend prior”. It is designed as parameter affected by two previous categories and a random term;

$$(\beta_k\setminus\beta_1,\beta_2,\ldots,\beta_{k-1})=2\beta_{k-1}-\beta_{k-2}+e_k\text{ , where }k=3,\ldots,i \quad (27)$$

5. CONCLUDING REMARKS

The choice of the prior distribution for a model's parameters appears as one of the essential problem in implementing Bayesian estimation and it is distinguishing property of Bayesian approach. There are many priors have been proposed in the literature and while introducing priors, there is a need to provide a base about the methods that shape the researcher's prior knowledge in Bayesian context. This paper renders a reassessment of these prior distribution classes under the classification non-informative and informative priors. Its content is composed of a comprehensive coverage so that most of the classes are included. However it should be noted that some key discussions on prior distributions as prior sensitivity, objectivity-subjectivity debate are left out of the paper. So there are many questions not answered yet; to what extent the prior distribution used in the analysis sensitive to other prior distribution choices? To be objective does one should use non-informative priors? Beyond all these discussions, the paper is mainly a review of the prior distribution classes, a generic representation of them.

6. REFERENCES

Berger, J. O., 1985. Statistical decision theory and Bayesian analysis. 2.ed., New York, Springer-Verlag Inc., 617.

Berger, J., 2006. The case for objective Bayesian analysis. Bayesian Analysis, 1(3):385-402.

Berger, J. O., Bernardo, J. M., 1992. On the development of reference priors. Bayesian Statistics 4 (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford: University Press, 35–60 (with discussion).

Berger, J. O., Pericchi, L. R., 1996. "The intrinsic Bayes factor for model selection and prediction." J. Amer. Statist. Assoc., 91: 109-122.

Bernardo, J. M., 1979. Reference posterior distributions for Bayesian inference. Journal of the Royal Statististical Society, B(41): 113-147.

Bernardo, J. M., 1996. Noninformative priors do not exist: A discussion with Jose M. Bernardo. Journal of Statistical Planning and Inference, 5 December 1996.

Box, G. E. P, Tiao, G. C., 1992. Bayesian inference in statistical analysis. Wiley Classics Library Edition, New York, John Wiley & Sons, 588.

Cano, J. A., Kessler, M., Salmeron, D., 2007. Integral priors for the one way random effects model. Bayesian Analysis, 2(1): 59-68.

Cano, J. A., Kessler, M., Moreno, E., 2004. On intrinsic priors for nonnested models. Test, 13(2): 445-463.

Cano, J. A., Salmeron, D., Robert, C. P., 2007. Integral equation solutions as prior distributions for Bayesian model selection. Test, Published online March 2007, DOI name 10.1007/s11749-006-0040-8.

Demirhan, H., Hamurkaroğlu, C., 2008. Bayesian estimation of log odds ratios from $R \times C$ and $2 \times 2 \times K$ contingency tables. Statistica Neerlandica, 62(4): 405–424.

Diaconis, P., Ylvisaker, D., 1985. Quantifying prior opinion, in Bayesian statistics 2, J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (eds.), Amsterdam: North Holland Press.

Ekici, O., Yorulmaz, Ö., 2008. The relationship of aberrant observation and structural break point: Determination with Bayesian autoregressive process. Doğuş University Journal, 9(2):146-157.

Galavotti, M. C., 2001. Subjectivism, Objectivism and Objectivity in Bruno de Finetti's Bayesianism. Foundations of Bayesianism, Ed. David Corfield and Jon Williamson, Kluwer Academic Publishers, 413.

Gelman, A., Carlin, J. B., Stern, H. S., Rubin, D. B., 1995. Bayesian data analysis, London, Chapman & Hall, 526.

Gelman, A., 2002. Prior distribution. Encyclopedia of Environmetrics, John Wiley & Sons Ltd., Chichester, 3: 1634–1637

Gelman, A., 2009. Prior distributions for Bayesian data analysis in political science. Frontier of Statistical Decision Making and Bayesian Analysis, in honor of James O. Berger.

Gill, J., 2002. Bayesian methods, New York, Chapman & Hall, 2002, 459.

Hartigan, J., 1964. Invariant prior distributions. The Annals of Mathematical Statistics, 35: 836-845.

Ibrahim J. G., Chen, M. H., 2000. Power prior distributions for regression models. Statistical Science, 15(1): 46-60.

Ibrahim J. G., Chen M.-H., Sinha D., 2003. On optimality properties of the power prior. Journal of the American Statistical Association, 98: 204-213.

Jaynes, E. T., 1968. Prior probabilities. IEEE Transactions on Systems Science and Cybernetics, SSC-4, 227-241, (Reprinted in Roger D. Rosenkrantz, Compiler. (1983 . E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics. Dordrecht, Holland: Reidel Publishing Company, 116-130).

Kass, R. E., Wasserman, L., 1996. The selection of prior distributions by formal rules. Journal of the American Statistical Association, 91(435): 1343-1362.

Lindley, D. V., 1965. Introduction to probability and statistics from a Bayesian viewpoint, Part 2, London, Cambridge University Press, 292.

Litterman, R. B., 1986. Forecasting with Bayesian vector autoregressions-five years of experience. Journal of Business & Economic Statistics, 4: 25-38.

Maddala, G. S., Kim, M., 2002. Unit roots, cointegration and structural change. Cambridge University Press, 505.

Phillips, P. C., 1991. To criticize the critics: An objective Bayesian analysis of stochastic trends. Journal of Applied Econometrics, 6(4):333–364.

Popper, K.R., 2003. Bilimsel araştırmanın mantığı, Çev. İlknur Aka, İbrahim Turan, 2.B, İstanbul, Yapı Kredi Yayınları, Kazım Taşkent Klasik Yapıtlar Dizisi, 2003, 596.

Raiffa, H., Schlaifer, R., 1968. Applied statistical decision theory. America, MIT Press, 356.

Sims, C. A., 1988. Bayesian skepticism on unit root econometrics. Journal of Economic Dynamics and Control, 12: 463-474.

Sivia, D. S., 1996. Data analysis, a Bayesian tutorial. New York, Oxford Unv. Press, 189.

Wasserman, L., 2006. Frequentist Bayes is objective (Comment on Articles by Berger and by Goldstein). Bayesian Analysis, 1(3):451-456.

Yardımcı, A., 1992. Çoklu bağlantılı çoklu doğrusal regresyonda Bayes yaklaşımı. Y.L., Fen Bilimleri Enstitüsü, Hacettepe Ünv., 72.

Zellner, A., 1971. An introduction to Bayesian inference in econometrics. New York, John Wiley & Sons, 431.

Zellner, A., 1986. On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In Goel, P. and Zellner, A. (eds.), Bayesian Inference and Decision Techniques, 233-243. Amsterdam: Elsevier Science Publishers B.V.

Acknowledgements

This paper based on part of the author's Thesis (2005) written at the University of Istanbul.

KAPSAMLI BİR İÇERİKLE ÖN DAĞILIM TÜRLERİ

ÖZET

Bayesyen analizlerin özünü oluşturan Bayes Teoremi, analizlere ön bilgiyi dahil ederek istatistiksel süreci gerçekleştirmektedir. Ön bilginin düzeyi, yapısı ve uygulama sınırları gözönüne alınınca, özellikle bu alanda yeni çalışan araştırmacılar için en zorlayıcı kısmı "ön dağılım" olarak görülebilir. Farklı alanlarda ihtiyaç doğrultusunda önerilen çeşitli ön dağılım türleri vardır. Öte yandan ön dağılımlar üzerine jenerik bir bakışı yansıtan ve gözden geçirme niteliğinde çalışma literatürde mevcut değildir. Bu motivasyonla, çalışma ön dağılım türlerini kapsamlı bir içerikle ele almaktadır. Böylelikle araştırmacılara ön dağılım türlerinin tanıtımı ve bunlarla ilgili genel bir bakış kazandırmak amaçlanmaktadır.

Anahtar Kelimeler: Ön dağılım türleri, Bilgi veren ön dağılımlar, Bilgi vermeyen ön dağılımlar.