Chapter 3. Construction of Priors

♠ **Subjective Determination**

• **The Histogram Approach**

When Θ is an interval of the real line, the most obvious approach to use is the histogram approach.

Divide Θ into intervals, determine the subjective probability of each interval, and then plot a probability histogram. From this histogram, a smooth density $\pi(\theta)$ can be sketched.

Difficulties: There is no clearcut rule which establishes how many intervals, what sizes intervals, etc., should be used in the histogram. In addition, the prior density has no tails (i.e., gives probability one to a bounded set).

- **Relative Likelihood Approach**

Suppose we are interested in only two events E and E^c. If E is twice as likely as E^c, then $\pi(E) = \frac{2}{3}$ and $\pi(E^c) = \frac{1}{3}$. However, this approach is of most use when Θ is a subset of the real line. It consists simply of comparing the intuitive "likelihoods" of various points in Θ, and directly sketching a prior density from these determinations.

- **Example 1**: Assume $\Theta = [0, 1]$. Suppose that the parameter point $\theta = \frac{3}{4}$ is felt to be the most likely, while $\theta = 0$ is the least likely. Also, $\frac{3}{4}$ is estimated to be three times as likely to be the true value of θ as is 0. It is deemed sufficient to determine the relative likelihoods of three other points, $\frac{1}{4}$, $\frac{1}{2}$, and 1. For simplicity, all points are compared with $\theta = 0$. It is decided that $\theta = \frac{1}{2}$ and 1 are twice as likely as $\theta = 0$, while $\frac{1}{4}$ is 1.5 times as likely as $\theta = 0$. Assign the base point $\theta = 0$ the prior density value 1. Figure 3.1 shows the resulting (unnormalized) prior density.

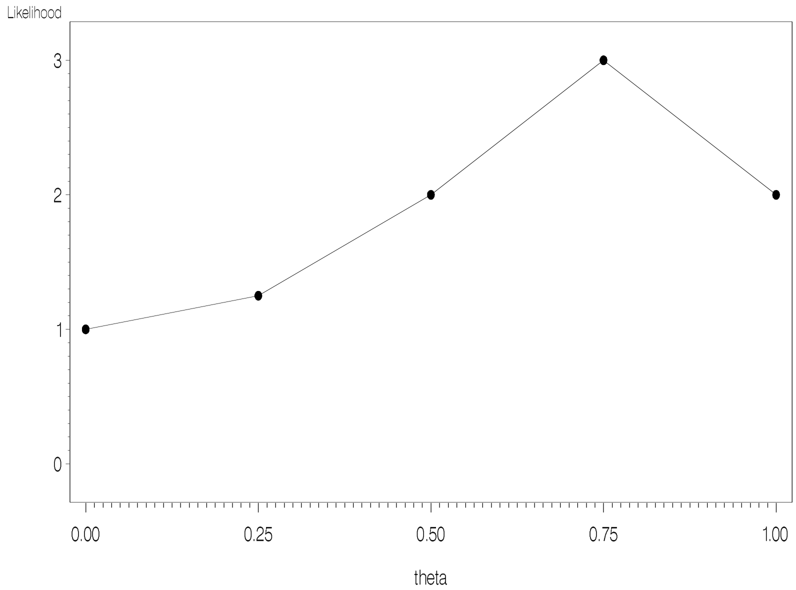

Figure 3.1

The prior density found in Figure 3.1 is not proper, in the sense that it does not integrate to one. A constant c could be found for which $c\pi(\theta)$ is a proper density, but fortunately there is no need to do so. The reason is that the Bayes action is the same whether $\pi(\theta)$ or $c\pi(\theta)$ is used as the prior density. This is clear since

$$\rho(c\pi, a) = \int_{\Theta} L(\theta, a)c\pi(\theta)d\theta = c\rho(\pi, a),$$

so that any a minimizing $\rho(c\pi, a)$ will also minimize $\rho(\pi, a)$.

Difficulties: (a) The relative likelihood determinations can be done only in a finite region. So, one must decided what to do outside the finite region when Θ is unbounded. (b) The density is not normalized.

- **Matching a Given Function Form**

The idea is to simply assume that $\pi(\theta)$ is of a given functional form, and to then choose the density of this given form which most closely matches prior beliefs.

For example, if the prior is assumed to have a $N(\mu, \sigma^2)$ functional form, one need only decide upon a prior mean and a prior variance to specify the density.

If the prior has a $\mathcal{B}e(\alpha, \beta)$ functional form, one can estimate the prior mean, μ, and variance, σ^2, and use the relationships

$$\mu = \frac{\alpha}{\alpha + \beta},$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\mu(1-\mu)}{\alpha+\beta+1}$$

to determine α and β.

• **Example 2**: Assume $\Theta = (-\infty, \infty)$ and the prior is thought to be from the normal family. It is subjectively determined that the median of the prior is 0, and the quartiles (i.e., $\frac{1}{4}$-fractile and $\frac{3}{4}$-fractile) are -1 and 1. Since, for a normal distribution, the mean and median are equal, it is clear that the desired normal mean is $\mu = 0$. Using tables of normal probabilities, the variance of the normal prior must be $\sigma^2 = 2.19$, since

$$P(Z < -1/(2.19)^{1/2}) = \frac{1}{4},$$

when $Z \sim N(0, 1)$. Hence π will be chosen to be a $N(0, 2.19)$ density.

If alternatively, it is assumed that the prior is Cauchy, one finds that $\mathcal{C}(0, 1)$ density is the appropriate choice, since the median is zero, and it can be checked that

$$\int_{-\infty}^{-1} \frac{1}{\pi(1 + \theta^2)} d\theta = \frac{1}{4}.$$

- **CDF Determination**

This approach is through subjective construction of the CDF. This can be done by subjectively determining several α-fractiles, $z(\alpha)$, plotting the points $(z(\alpha), \alpha)$, and sketching a smooth curve joining them.

- **Multivariate Cases**

Use

$$\pi(\boldsymbol{\theta}) = \prod \pi(\theta_i)$$

or

$$\pi(\boldsymbol{\theta}) = \pi(\theta_p|\theta_{p-1}, \ldots, \theta_1) \cdots \pi(\theta_2|\theta_1)\pi(\theta_1).$$

♠ Noninformative Priors

• Noninformative

A noninformative prior is a prior which contains no information about θ (or more crudely which "favors" no possible values of θ over others).

For example, in testing between two hypotheses, the prior which gives probability $\frac{1}{2}$ to each hypothesis is noninformative.

• **Example 4**: Suppose the parameter of interest is a normal mean, θ, so that the parameter space is $\Theta = (-\infty, \infty)$. If a noninformative prior density is desired, it seems reasonable to give equal weight to all possible values of θ. If $\pi(\theta) = c > 0$ is chosen, then π has infinite mass (i.e., $\int_{-\infty}^{\infty} \pi(\theta) d\theta = \infty$ and is not a proper density. The choice of c is unimportant, so that typically the noninformative prior density for this problem is chosen to be $\pi(\theta) = 1$. This is often called the improper *uniform density* on R^1.

• Improper Prior

An improper prior is the one which has infinite mass.

• **Lack of Invariance under Transformation**

• **Example 4 (continued)**: Instead of considering θ, suppose the problem had been parameterized in terms of $\eta = \exp(\theta)$. If $\pi(\theta)$ is the density for θ, then the corresponding density for η is

$$\pi^*(\theta) = \eta^{-1}\pi(\log \eta).$$

Hence, if the noninformative prior for θ is chosen to be constant, we should choose the noninformative prior for η to be proportional to η^{-1} to maintain consistency. Thus, we cannot maintain consistency and choose both the noninformative prior for θ and that for η to be constant.

- **Noninformative Prior for Location Problems**

<u>**Location Parameters**</u>:

Suppose $\mathcal{X}$ and Θ are subsets of R^p, and that the density of $\boldsymbol{X}$ is of the form $f(\boldsymbol{x} - \boldsymbol{\theta})$. The density is then said to be a *location density*, and $\boldsymbol{\theta}$ is called a *location parameter*. The $N(\mu, \sigma^2)$ (σ^2 fixed), $\mathcal{T}(\alpha, \mu, \sigma^2)$ (α and σ^2 fixed), $\mathcal{C}(\alpha, \beta)$ (β fixed), and $N_p(\boldsymbol{\theta}, \Sigma)$ (Σ fixed) densities are all examples of location densities.

Derivation of A Noninformative Prior for $\boldsymbol{\theta}$:

Consider instead of observing $\boldsymbol{X}$, we observe $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{c}$ ($\boldsymbol{c} \in R^p$). Define $\boldsymbol{\eta} = \boldsymbol{\theta} + \boldsymbol{c}$. Then

$$\boldsymbol{Y} \sim f(\boldsymbol{y} - \boldsymbol{\eta}).$$

Since $(\boldsymbol{X}, \boldsymbol{\theta})$ and $(\boldsymbol{Y}, \boldsymbol{\eta})$ are identical in structure, it is reasonable to insist they have the same noninformative prior.

Letting π and π^* denote the noninformative priors in the $(\boldsymbol{X}, \boldsymbol{\theta})$ and $(\boldsymbol{Y}, \boldsymbol{\eta})$ problems, then π and π^* should be equal, i.e., that

$$P^{\pi}(\boldsymbol{\theta} \in A) = P^{\pi^*}(\eta \in A)$$

for any set A in R^p. Since $\boldsymbol{\eta} = \boldsymbol{\theta} + \boldsymbol{c}$, we should have

$$P^{\pi^*}(\eta \in A) = P^{\pi}(\boldsymbol{\theta} + \boldsymbol{c} \in A) = P^{\pi}(\boldsymbol{\theta} \in A - \boldsymbol{c}),$$

where $A - \boldsymbol{c} = \{\boldsymbol{z} - \boldsymbol{c} : \boldsymbol{z} \in A\}$.

Assume that the prior has a density, and we then can write

$$\int_A \pi(\boldsymbol{\theta})d\boldsymbol{\theta} = \int_{A-\boldsymbol{c}} \pi(\boldsymbol{\theta})d\boldsymbol{\theta} = \int_A \pi(\boldsymbol{\theta} - \boldsymbol{c})d\boldsymbol{\theta}.$$

If this is to hold for all sets A, it must be true that

$$\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta} - \boldsymbol{c})$$

for all $\boldsymbol{c}$. Any π satisfying this relationship is said to be a *location invariant prior*. Setting $\boldsymbol{\theta} = \boldsymbol{c}$ gives

$$\pi(\boldsymbol{c}) = \pi(0).$$

Recall that this should hold for all $\boldsymbol{c} \in R^p$. Therefore, π must be a constant function. As discussed earlier, the constant is unimportant. So, the noninformative prior density for a location parameter is

$$\pi(\boldsymbol{\theta}) = 1.$$

- **Relatively Invariant Location Prior**

Suppose
$$\pi^{\pi^*}(A) = h(\boldsymbol{c})P^{\pi}(A - \boldsymbol{c}).$$
We have
$$\int_A \pi(\boldsymbol{\theta})d\boldsymbol{\theta} = h(\boldsymbol{c})\int_{A-\boldsymbol{c}} \pi(\boldsymbol{\theta})d\boldsymbol{\theta} = h(\boldsymbol{c})\int_A \pi(\boldsymbol{\theta} - \boldsymbol{c})d\boldsymbol{\theta},$$
which implies
$$\pi(\boldsymbol{\theta}) = h(\boldsymbol{c})\pi(\boldsymbol{\theta} - \boldsymbol{c}).$$
Setting $btheta = \boldsymbol{c}$ gives $h(\boldsymbol{c}) = \pi(\boldsymbol{c})/\pi(0)$. Thus
$$\pi(\boldsymbol{\theta} - \boldsymbol{c}) = \frac{\pi(0)\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{c})}.$$
There are many improper priors besides the uniform, which satisfy this relationship. An example is $\pi(\boldsymbol{\theta}) = \exp\{\boldsymbol{\theta}'\boldsymbol{z}\}$, where $\boldsymbol{z}$ is any fixed vector. A prior satisfying the above relationship is called *relatively location invariant.*

• **Noninformative Prior for Scale Problems**

Scale Parameters:

A (one-dimensional) *scale density* is a density of the form

$$\sigma^{-1} f\left(\frac{x}{\sigma}\right),$$

where $\sigma > 0$. The parameter σ is called a *scale parameter*. The $N(0, \sigma^2)$, $\mathcal{T}(\alpha, 0, \sigma^2)$ (α fixed), and $\mathcal{G}(\alpha, \beta)$ (α fixed) densities are all examples of scale densities.

Derivation of A Noninformative Prior for σ:

Suppose we observe $Y = cX$ $(c > 0)$ instead of observing X. Define $\eta = c\sigma$. Then $c = \frac{\eta}{\sigma}$ and the density of Y is

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \eta^{-1} f_X(y/\eta).$$

Let π and π^* be the priors in (X, σ) and (Y, η) problems. The equality

$$P^{\pi}(\sigma \in A) = P^{\pi^*}(\eta \in A)$$

should hold for all $A \subset (0, \infty)$. Since $\eta = c\sigma$, it should be true that

$$P^{\pi^*}(\eta \in A) = P^{\pi}(\sigma \in c^{-1}A),$$

where $c^{-1}A = \{c^{-1}z: \ z \in A\}$. Thus,

$$P^{\pi}(\sigma \in A) = P^{\pi}(\sigma \in c^{-1}A)$$

should hold for all $c > 0$, and any distribution π for which this is true is called *scale invariant.*

Assuming that the prior has a density, we can write

$$\int_A \pi(\sigma)d\sigma = \int_{c^{-1}A} \pi(\sigma)d\sigma = \int_A \pi(c^{-1}\sigma)c^{-1}d\sigma$$

for all A, which implies

$$\pi(\sigma) = c^{-1}\pi(c^{-1}\sigma) \ \text{ for all } \sigma.$$

Choosing $\sigma = c$, it follows that

$$\pi(c) = c^{-1}\pi(1) \ \text{ for all } c > 0.$$

Setting $\pi(1) = 1$ for convenience, it follows that a reasonable *noninformative prior for a scale parameter* is

$$\pi(\sigma) = \sigma^{-1}.$$

Observe that this is also an improper prior, since $\int_0^\infty \sigma^{-1}d\sigma = \infty$.

- **Jeffreys's Prior**

The most widely used noninformative prior is that of Jeffreys (1961), which is to choose

$$\pi(\theta) = [I(\theta)]^{1/2},$$

where $I(\theta)$ is the expected Fisher information; under certain regularity conditions, this is given by

$$I(\theta) = -E_\theta\left[\frac{\partial^2 \log f(\boldsymbol{X}|\theta)}{\partial\theta^2}\right].$$

If $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)'$ is a vector, Jeffreys (1961) suggests the use of

$$\pi(\boldsymbol{\theta}) = [\det I(\boldsymbol{\theta})]^{1/2},$$

where $I(\boldsymbol{\theta})$ is the $p \times p$ Fisher information matrix with (i, j) element

$$I_{ij}(\boldsymbol{\theta}) = -E_\theta\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(\boldsymbol{X}|\boldsymbol{\theta})\right].$$

- **Example 7 (Location-Scale Parameters)**:

A *location-scale* density is a density of the form $\sigma^{-1}f((x-\theta)/\sigma)$, where $\theta \in R^1$ and $\sigma > 0$ are the unknown parameters. The $N(\theta, \sigma^2)$ and $\mathcal{T}(\alpha, \theta, \sigma^2)$ (α fixed) densities are the examples of location-scale densities.

Working with the normal distribution, and noting that $\boldsymbol{\theta} = (\theta, \sigma)$, the Fisher information matrix is

$$
\begin{aligned}
& I(\boldsymbol{\theta}) \\
&= -E_{\boldsymbol{\theta}} \begin{pmatrix} \frac{\partial^2}{\partial\theta^2}\left[-\log\sigma - \frac{(X-\theta)^2}{2\sigma^2}\right] & \frac{\partial^2}{\partial\theta\partial\sigma}\left[-\log\sigma - \frac{(X-\theta)^2}{2\sigma^2}\right] \\ \frac{\partial^2}{\partial\theta\partial\sigma}\left[-\log\sigma - \frac{(X-\theta)^2}{2\sigma^2}\right] & \frac{\partial^2}{\partial\sigma^2}\left[-\log\sigma - \frac{(X-\theta)^2}{2\sigma^2}\right] \end{pmatrix} \\
&= -E_{\boldsymbol{\theta}} \begin{pmatrix} -1/\sigma^2 & 2(\theta - X)/\sigma^3 \\ 2(\theta - X)/\sigma^3 & 1/\sigma^2 - 3(X-\theta)^2/\sigma^4 \end{pmatrix} \\
&= \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix}.
\end{aligned}
$$

Hence, the Jeffreys's prior is

$$\pi(\boldsymbol{\theta}) = \left(\frac{1}{\sigma^2} \cdot \frac{2}{\sigma^2}\right)^{1/2} \propto \frac{1}{\sigma^2}.$$

This is improper. So, we can again ignore any multiplicative constants.

A noninformative prior for this situation could, alternatively, be derived by assuming *independence* of θ and σ, and multiplying the noninformative priors obtained earlier. The result is

$$\pi(\theta, \sigma) = \pi(\theta)\pi(\sigma) = \frac{1}{\sigma};$$

which is recommended by Jeffreys (1961).

- **A Proper Jeffreys's Prior**

Assume $x_1, x_2, \ldots, x_n$ are independent observation from a binomial regression model with the logit link such that

$$X_i \sim \mathcal{B}(n_i, p_i),$$

where $p_i = \frac{\exp(\boldsymbol{z}_i \boldsymbol{\beta})}{1+\exp(\boldsymbol{z}_i \boldsymbol{\beta})}$, $\boldsymbol{z}_i = (z_{i1}, \ldots, z_{ip})'$ is a vector of covariates, and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a vector of regression coefficients. We can write

$$\begin{aligned} f(x_i|\boldsymbol{\beta}) &= \binom{n_i}{x_i} p_i^{x_i} (1-p_i)^{n_i - x_i} \\ &= \binom{n_i}{x_i} \exp\{x_i \theta_i - b(\theta_i)\}, \end{aligned}$$

where

$$\theta_i = \log(p_i/(1-p_i)) = \boldsymbol{z}_i \boldsymbol{\beta}$$

and

$$b(\theta_i) = -n_i \log(1 - p_i) = n_i \log(1 + e^{\theta_i}) = n_i \log(1 + e^{\boldsymbol{z}_i \boldsymbol{\beta}}).$$

Then, the Fisher information matrix is

$$I(\boldsymbol{\beta}) = \sum_{i=1}^{n} I_i(\boldsymbol{\beta}),$$

and the (j, k) element of $I_i(\boldsymbol{\beta})$ is

$$\begin{aligned} I_{ijk}(\boldsymbol{\beta}) = & - E_{\boldsymbol{\beta}} \left[\frac{\partial^2}{\partial \beta_j \partial \beta_k} \log f(X_i | \boldsymbol{\beta}) \right] \\ = & n_i z_{ij} z_{ik} \frac{e^{\boldsymbol{z}_i \boldsymbol{\beta}}}{(1 + e^{\boldsymbol{z}_i \boldsymbol{\beta}})^2}. \end{aligned}$$

Let $V(\boldsymbol{\beta})$ denote an $n \times n$ diagonal matrix with the i^{th} diagonal element

$$v_i = v(\boldsymbol{z}' \boldsymbol{\beta}) = \frac{d^2 b(\theta_i)}{d\theta_i^2} = n_i \frac{e^{\boldsymbol{z}_i \boldsymbol{\beta}}}{(1 + e^{\boldsymbol{z}_i \boldsymbol{\beta}})^2}.$$

Then, the Jeffreys's prior for $\boldsymbol{\beta}$ is given by

$$\pi(\boldsymbol{\beta}) = |Z' V(\boldsymbol{\beta}) Z|^{1/2},$$

where Z is the $n \times p$ design matrix with the i^{th} row $\boldsymbol{z}_i'$.

Ibrahim and Laud (1991) show that if Z is of full rank p, then $\pi(\boldsymbol{\beta})$ is proper, i.e., $\int_{R^p} \pi(\boldsymbol{\beta}) d\boldsymbol{\beta} < \infty$.

♠ Maximum Entropy Priors

• Discrete Case

<u>**Definition**</u>: Assume Θ is discrete, and let π be a probability density on Θ. The *entropy* of π, to be denoted $\mathcal{E}n(\pi)$, is defined as

$$\mathcal{E}n(\pi) = -\sum_{\Theta} \pi(\theta_i) \log \pi(\theta_i).$$

(If $\pi(\theta_i) = 0$, the quantity $\pi(\theta_i) \log \pi(\theta_i)$ is defined to be zero.)

• **Example 9**: Assume $\Theta = \{\theta_1, \theta_2, \ldots, \theta_n\}$. If $\pi(\theta_k) = 1$, while $\pi(\theta_i) = 0$ for $i \neq k$, then clearly the probability distribution describes exactly which parameter point will occur. The "uncertainty" is zero. Correspondingly,

$$\mathcal{E}n(\pi) = -\sum_{i=1}^{n} \pi(\theta_i) \log \pi(\theta_i) = 0.$$

At the other extreme, the "most uncertainty" or *maximum entropy* probability distribution is that with $\pi(\theta_i) = \frac{1}{n}$ for all i. For this π,

$$\mathcal{E}n(\pi) = -\sum_{i=1}^{n} \frac{1}{n} \log\left(\frac{1}{n}\right) = \log n.$$

Class Discussion Problem:

Show that $\mathcal{E}n(\pi) \leq \log n$ for all proper π.

Restricted Maximum Entropy Prior:

Assume that

$$E^{\pi}[g_k(\theta)] = \sum_i \pi(\theta_i) g_k(\theta_i) = \mu_k, \quad k = 1, 2, \ldots, m.$$

Then, the maximum entropy prior is

$$\bar{\pi}(\theta_i) = \frac{\exp\{\sum_{k=1}^m \lambda_k g_k(\theta_i)\}}{\sum_i \exp\{\sum_{k=1}^m \lambda_k g_k(\theta_i)\}},$$

where the λ_k are constants to be determined from the above m constraints.

Proof: Using Lagurang multiplicative, consider

$$h(\pi) = \sum_i \pi(\theta_i) \log \pi(\theta_i) - \sum_i \sum_{k=1}^{m} \lambda_k (\pi(\theta_i) g_k(\theta_i) - \mu_k).$$

Setting

$$\frac{\partial h(\pi)}{\partial \pi(\theta_i)} = 0$$

for all i, we obtain

$$\pi(\theta_i) \propto \exp\{\sum_{k=1}^{m} \lambda_k g_k(\theta_i)\}.$$

Since $\sum_i \pi(\theta_i) = 1$, then we have

$$\pi(\theta_i) = \frac{\exp\{\sum_{k=1}^{m} \lambda_k g_k(\theta_i)\}}{\sum_i \exp\{\sum_{k=1}^{m} \lambda_k g_k(\theta_i)\}}.$$

- **Example 10**: Assume $\Theta \subset R^1$, $g_1(\theta) = theta$, and $g_k(\theta) = (\theta - \mu_1)^k$ for $2 \leq k \leq m$. Then the m constraints correspond to the specification of the first m central moments, μ_i, of π.

- **Example 11**: Assume $\Theta \subset R^1$, and

$$g_k(\theta) = I_{(-\infty, z_k]}(\theta).$$

Clearly

$$E^\pi[g_k(\theta)] = P^\pi(\theta \leq z_k),$$

so z_k is the μ_k-fractile of π. In this case, we specify m fractiles of π.

- **Example 12**: Assume $\Theta = \{0, 1, 2, \dots\}$, and $E^\pi[\theta] = 5$. This restriction gives $g_1(\theta) = \theta$ and $\mu_1 = 5$. The restricted maximum entropy prior is

$$\bar{\pi}(\theta) = \frac{e^{\lambda_1 \theta}}{\sum_{\theta=0}^{\infty} e^{\lambda_1 \theta}} = (1 - e^{\lambda_1})(e^{\lambda_1})^\theta.$$

This is a $\mathcal{G}e(e^{\lambda_1})$ density, the mean of which is $(1 - e^{\lambda_1})/e^{\lambda_1}$. Setting this to $\mu_1 = 5$, and solving, gives $e^{\lambda_1} = \frac{1}{6}$. Hence $\bar{\pi}$ is $\mathcal{G}e(1/6)$.

- **Continuous Case**

__**Definition**__: If Θ is continuous, the Θ. The *entropy* of π is defined as

$$\mathcal{E}n(\pi) = -\int_{\Theta} \pi(\theta) \log\left(\frac{\pi(\theta)}{\pi_0(\theta)}\right) d\theta,$$

where $\pi_0(\theta)$ us the natural "invariant" noninformative prior for the problem.

__**Restricted Maximum Entropy Prior**__:

Assuming

$$E^{\pi}[g_k(\theta)] = \int_{\Theta} g_k(\theta)\pi(\theta)d\theta = \mu_k$$

for $k = 1, 2, \ldots, m$, the (proper) prior density which maximizes $\mathcal{E}n(\pi)$ is given by

$$\bar{\pi}(\theta) = \frac{\pi_0(\theta)\exp[\sum_{k=1}^{m} \lambda_k g_k(\theta)]}{\int_{\Theta} \pi_0(\theta)\exp[\sum_{k=1}^{m} \lambda_k g_k(\theta)]d\theta},$$

where the λ_k are constants to be determined from the constraints.

• **Example 13**: Assume $\Theta = R^1$, and that θ is a location parameter. The natural noninformative prior is $\pi_0(\theta) = 1$. Assume the true prior mean is μ and the true prior variance is σ^2. These restrictions give $g_1(\theta) = \theta$, $\mu_1 = \mu$, $g_2(\theta) = (\theta - \mu)^2$, and $\mu_2 = \sigma^2$. The maximum entropy prior is thus

$$\bar{\pi}(\theta) = \frac{\exp[\lambda_1\theta + \lambda_2(\theta - \mu)^2]}{\int_\Theta \exp[\lambda_1\theta + \lambda_2(\theta - \mu)^2]d\theta},$$

where λ_1 and λ_2 are to be chosen so that the two constraints are satisfied. Note that

$$\lambda_1\theta + \lambda_2(\theta - \mu)^2 = \lambda_2\theta^2 + (\lambda_1 - 2\mu\lambda_2)\theta + \lambda_2\mu^2$$

$$= \lambda_2\left[\theta - \left(\mu - \frac{\lambda_1}{2\lambda_2}\right)\right]^2 + \left[\lambda_1\mu - \frac{\lambda_1^2}{4\lambda_2}\right].$$

Hence

$$\bar{\pi}(\theta) \propto \exp\left\{\lambda_2\left[\theta - \left(\mu - \frac{\lambda_1}{2\lambda_2}\right)\right]^2\right\},$$

which is a normal density with mean $\mu - \lambda_1/(2\lambda_2)$ and variance $-1/(2\lambda_2)$. Choosing $\lambda_1 = 0$ and $\lambda_2 = -1/(2\sigma^2)$ satisfies both constraints. Thus, $\bar{\pi}(\theta)$ is a $N(\mu, \sigma^2)$ density.

- **Example 14**: In Example 13, assume that the only restriction given is $E^{\pi}[\theta] = \mu$. The solution $\bar{\pi}$ must then be of the form

$$\bar{\pi}(\theta) = \frac{\exp\{\lambda_1 \theta\}}{\int_{-\infty}^{\infty} \exp\{\lambda_1 \theta\} d\theta}.$$

It is clear that $\bar{\pi}(\theta)$ cannot be a proper density for any λ_1. Hence there is no solution.