Modern Methods of Statistical Learning sf2935 Auxiliary material: Exponential Family of Distributions Timo Koski

TK

Second Quarter 2016

KTH Matematik

The family of distributions with densities (w.r.t. to a σ-finite measure μ) on $\mathcal{X}$ defined by

$$f(x \mid \theta) = C(\theta)h(x)e^{R(\theta)\cdot T(x)}$$

is called an exponential family, where

- $C(\theta)$ and $h(x)$ are functions from Θ and $\mathcal{X}$ to R_+,
- $R(\theta)$ and $T(x)$ are functions from Θ and $\mathcal{X}$ to R^k,
- $R(\theta) \cdot T(x)$ is a scalar product in R^k, i.e.,

$$R(\theta) \cdot T(x) = \sum_{i=1}^{k} R_i(\theta) \cdot T_i(x)$$

Exponential Families: Role of μ

The σ-finite measure μ appears as follows:

$$1 = C(\theta) \int_{\mathcal{X}} h(x) e^{R(\theta) \cdot T(x)} d\mu(x).$$

$$C(\theta) = \frac{1}{\int_{\mathcal{X}} h(x) e^{R(\theta) \cdot T(x)} d\mu(x)}.$$

$$\mathcal{N} := \left\{ \theta | \int_{\mathcal{X}} h(x) e^{R(\theta) \cdot T(x)} d\mu(x) < \infty \right\}$$

$\mathcal{N}$ is called the natural parameter space.

KTH Matematik

EXAMPLES OF EXPONENTIAL FAMILIES: $Be(\theta)$

$$f(x|\theta) = \theta^x \cdot (1-\theta)^{1-x}. \quad x = 0, 1$$

We write

$$f(x|\theta) = C(\theta) e^{R(\theta) \cdot x},$$

where

$$C(\theta) = e^{\log 1 - \theta}, T(x) = x, R(\theta) = \log \frac{\theta}{1-\theta}, h(x) = 1.$$

KTH VETENSKAP OCH KONST
KTH Matematik

EXAMPLES OF EXPONENTIAL FAMILIES: $N(\mu, \sigma^2)$

$x^{(n)} = (x_1, x_2, \ldots, x_n)$, x_i I.I.D. $\sim N(\mu, \sigma^2)$.

$$f(x^{(n)}|\mu, \sigma^2) = \frac{1}{\sigma^n (2\pi)^{n/2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

$$= \frac{1}{(2\pi)^{n/2}} \sigma^{-n} e^{-\frac{n\mu^2}{2\sigma^2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} n\overline{x}}.$$

EXAMPLES OF EXPONENTIAL FAMILIES: $N(\mu, \sigma^2)$

$$f\left(x^{(n)} | \mu, \sigma^2\right) = \frac{1}{(2\pi)^{n/2}} \sigma^{-n} e^{-\frac{n\mu^2}{2\sigma^2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2} n\bar{x}}.$$

$$C(\theta) = \sigma^{-n} e^{-\frac{n\mu^2}{2\sigma^2}}, h(x) = \frac{1}{(2\pi)^{n/2}}.$$

$$R(\theta) \cdot T\left(x^{(n)}\right) = R_1(\theta) T_1\left(x^{(n)}\right) + R_2(\theta) T_2\left(x^{(n)}\right)$$

$$T_1\left(x^{(n)}\right) = \sum_{i=1}^{n} x_i^2, T_2\left(x^{(n)}\right) = n\bar{x}$$

$$R_1(\theta) = -\frac{1}{2\sigma^2}, R_2(\theta) = \frac{\mu}{\sigma^2}$$

KTH VETENSKAP OCH KONST
KTH Matematik

Exponential Families & Sufficient Statistics (1)

In an exponential family there exists a *sufficient statistic* of constant dimension (i.e., not depending on n) for any I.I.D. sample

$$x_1, x_2, \ldots, x_n \sim f\left(x|\theta\right).$$

This means that

$$f\left(x_1|\theta\right) \cdot f\left(x_2|\theta\right) \cdot \ldots \cdot f\left(x_n|\theta\right)$$

$$= C(\theta)^n \prod_{i=1}^{n} h(x_i) e^{R(\theta) \cdot \sum_{i=1}^{n} T(x_i)}$$

KTH Matematik

Exponential Families & Sufficient Statistics (2)

$$C(\theta)^n \cdot \prod_{i=1}^{n} h(x_i) \cdot e^{R(\theta)\cdot\sum_{i=1}^{n} T(x_i)}$$

where

$$\sum_{i=1}^{n} T(x_i) \in \mathcal{X}$$

is a sufficient statistic (explained next $\Rightarrow$).

Sufficient Statistic: A General Definition

For $x \sim f(x|\theta)$, a function T of x is called a sufficient statistic (for θ), if the distribution of x conditional on $T(x)$ does not depend on θ.

$$f\left(x|T,\theta\right) = f\left(x|T\right)$$

Bayesian definition:

$$\pi\left(\theta|x\right) = f\left(T,\theta\right)$$

Sufficient Statistic: Example

$$x^{(n)} = (x_1, x_2, \ldots, x_n),\ x_i \text{ I.I.D.} \sim \mathrm{Be}(\theta).$$

$$f\left(x^{(n)}|\theta\right) = C(\theta)^n \prod_{i=1}^{n} e^{R(\theta)\cdot x_i} = C(\theta)^n e^{R(\theta)\sum_{i=1}^{n} t(x_i)}$$

$$= \theta^t (1-\theta)^{n-t}$$

We know that

$$\sum_{i=1}^{n} t\left(x_i\right) \sim \mathrm{Bin}\left(\theta, n\right)$$

Sufficient Statistic: Example (contn'd)

Since $t = t(x_i)$ is determined by $x^{(n)}$,

$$f(x^{(n)}|\theta, t) = \frac{f(x^{(n)}, t|\theta)}{f(t \mid \theta)}$$

$$= \frac{\theta^t (1-\theta)^{n-t}}{\binom{n}{t} \theta^t (1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}$$

which does not depend on θ, $\sum_{i=1}^{n} t(x_i)$ is sufficient.

If

$$f(x \mid \theta) = C(\theta)h(x)e^{\theta \cdot x}$$

where $\Theta \subseteq R^k$ and $\mathcal{X} \subseteq R^k$, the family is said to be *natural* . Here $\theta \cdot x$ is inner product on R^k.

Natural Exponential Families (2)

$$f(x \mid \theta) = h(x)e^{\theta \cdot x - \psi(\theta)}$$

where

$$\psi(\theta) = -\log C(\theta).$$

Natural Exponential Families: Poisson Distribution

$$f(x \mid \lambda) = e^{-\lambda}\frac{\lambda^x}{x!}, x = 0, 1, 2, \ldots,$$

$$= \frac{1}{x!}e^{\theta x - e^{\theta}}$$

$$\psi(\theta) = e^{\theta}, \theta = \log\lambda, h(x) = \frac{1}{x!}$$

KTH Matematik

Mean in a Natural Exponential Family

If $E_\theta[x]$ denotes the mean (vector) of $x \sim f(x|\theta)$ in a natural family, then

$$E_\theta[x] = \int_{\mathcal{X}} x f(x \mid \theta)\, dx = \nabla_\theta \psi(\theta).$$

where $\Theta \in \mathrm{int}(\mathcal{N})$ and $\mathcal{X} \subseteq R^k$.

Proof:

$$\int_{\mathcal{X}} x f(x \mid \theta)\, dx = e^{-\psi(\theta)} \int_{\mathcal{X}} h(x) x e^{\theta \cdot x} dx.$$

Mean in a Natural Exponential Family

$$e^{-\psi(\theta)} \int_{\mathcal{X}} h(x) x e^{\theta \cdot x} dx = e^{-\psi(\theta)} \int_{\mathcal{X}} h(x) \nabla_\theta e^{\theta \cdot x} dx$$

$$= e^{-\psi(\theta)} \nabla_\theta \int_{\mathcal{X}} h(x) e^{\theta \cdot x} dx = e^{-\psi(\theta)} \nabla_\theta \frac{1}{C(\theta)} =$$

$$= e^{-\psi(\theta)} \frac{(-\nabla_\theta C(\theta))}{C(\theta)^2}$$

$$= C(\theta) \frac{(-\nabla_\theta C(\theta))}{C(\theta)^2} = \frac{(-\nabla_\theta C(\theta))}{C(\theta)}$$

$$= \nabla_\theta (-\log C(\theta)) = \nabla_\theta \psi(\theta).$$

□

Mean in a Natural Exponential Family : Poisson Distribution

$$f(x \mid \lambda) = \frac{1}{x!} e^{\theta x - e^{\theta}}$$

$$\psi(\theta) = e^{\theta}$$

$$E_{\theta}[x] = \frac{d}{d\theta}\psi(\theta) = e^{\theta} = \lambda.$$

Conjugate Priors for Exponential Families: An Intuitive Example

$x^{(n)} = (x_1, x_2, \ldots, x_n)$. $x_i \sim \mathrm{Po}(\lambda)$, I.I.D.,

$$f\left(x^{(n)} \mid \lambda\right) = e^{-n\lambda} \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

The likelihood is

$$l\left(\lambda | x^{(n)}\right) \propto e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}$$

This suggests the conjugate density as the density of the Gamma distribution, which is of the form

$$\pi\left(\lambda\right) \propto e^{-\beta\lambda} \lambda^{\alpha-1}$$

and hence

$$\pi\left(\lambda | x^{(n)}\right) \propto e^{-\lambda(\beta+n)} \lambda^{\sum_{i=1}^{n} x_i + \alpha - 1}$$

Conjugate Family of Priors for Exponential Families

Consider the natural exponential family

$$f(x \mid \theta) = h(x)e^{\theta \cdot x - \psi(\theta)}.$$

Then the conjugate family is given by

$$\pi(\theta) = \psi(\theta|\mu, \lambda) = K(\mu, \lambda)\, e^{\theta \cdot \mu - \lambda \psi(\theta)}$$

and the posterior is

$$\psi(\theta|\mu + x, \lambda + 1)$$

KTH Matematik

Proof: By Bayes' rule

$$\pi(\theta|x) = \frac{f(x \mid \theta)\, \pi(\theta)}{m(x)}$$

We have

$$f(x \mid \theta)\, \pi(\theta) = h(x) e^{\theta \cdot x - \psi(\theta)} \psi(\theta|\mu, \lambda)$$

$$= h(x) K(\mu, \lambda)\, e^{\theta \cdot (x+\mu) - (1+\lambda)\psi(\theta)}$$

$$m(x) = \int_{\Theta} f(x \mid \theta)\, \pi(\theta)\, d\theta =$$

$$= h(x) K(\mu, \lambda) \int_{\Theta} e^{\theta \cdot (x+\mu) - (1+\lambda)\psi(\theta)} d\theta$$

$$= h(x) K(\mu, \lambda)\, K(x+\mu, \lambda+1)^{-1}$$

as ψ is a density on Θ.

Conjugate Priors for Exponential Families: Proof

$$\pi(\theta|x) = \frac{h(x)K(\mu,\lambda)\, e^{\theta\cdot(x+\mu)-(1+\lambda)\psi(\theta)}}{h(x)K(\mu,\lambda)\, K(x+\mu,\lambda+1)^{-1}}$$

$$= K(x+\mu,\lambda+1)\, e^{\theta\cdot(x+\mu)-(1+\lambda)\psi(\theta)},$$

which shows that the posterior belongs to the same family as the prior and that

$$\pi(\theta|x) = \psi(\theta|\mu+x,\lambda+1)$$

as claimed. □

The proof requires that

$$\pi(\theta) = \psi(\theta|\mu, \lambda) = K(\mu, \lambda)\, e^{\theta\cdot\mu - \lambda\psi(\theta)}$$

is a probability density on Θ. The conditions for this are given in exercise 3.35.

A Mean for Exponential Families

We have the following properties:

- if $\pi(\theta) = K(x_o, \lambda)\, e^{\theta \cdot x_o - \lambda \psi(\theta)}$ then

$$\xi(\theta) = \int_{\Theta} E_\theta[x]\, \pi(\theta)\, d\theta = \frac{x_o}{\lambda}$$

This has been proved by Diaconis and Ylvisaker (1979). The proof is not summarized here.

Posterior Means with Conjugate Priors for Exponential Families

- if $\pi(\theta) = K(\mu, \lambda) e^{\theta \cdot \mu - \lambda \psi(\theta)}$ then

$$\int_{\Theta} E_{\theta}[x] \, \pi\left(\theta | x^{(n)}\right) d\theta = \frac{\mu + n\overline{x}}{\lambda + n}$$

This follows from the preceding, as shown by Diaconis and Ylvisaker (1979).

Mean of a Predictive Distribution

$$\int_{\Theta} E_{\theta}[x]\, \pi\left(\theta|x^{(n)}\right) d\theta = \int_{\Theta}\int_{\mathcal{X}} x f(x|\theta)\, dx \pi\left(\theta|x^{(n)}\right) d\theta$$

(by Fubini's theorem)

$$= \int_{\mathcal{X}} x \int_{\Theta} f(x|\theta)\, \pi\left(\theta|x^{(n)}\right) d\theta dx$$

(by definition in lecture 9 of sf3935)

$$= \int_{\mathcal{X}} x g(x|x^{(n)}) dx$$

the mean of the predictive distribution.

KTH Matematik

Mean of a Predictive Distribution

Hence if conjugate priors for exponential families are used, then

$$\int_{\mathcal{X}} x g(x|x^{(n)}) dx = \frac{\mu + n\overline{x}}{\lambda + n}$$

is the mean of the corresponding predictive distribution. This suggests μ and λ as 'virtual observations'.

Laplace's Prior

P.S. Laplace[1] formulated the principle of insufficient reason to choose a prior as a uniform prior. There are drawbacks in this. Consider Laplace's prior for $\theta \in [0,1]$

$$\pi(\theta) = \begin{cases} 1 & 0 \leq \theta \leq 1 \\ 0 & \text{elsewhere,} \end{cases}$$

Then consider

$$\phi = \theta^2.$$

[1] http://www-groups.dcs.st-and.ac.uk/~history/Mathematicians/Laplace.html

KTH Matematik

Laplace's Prior

We find the density of $\phi = \theta^2$. Take $0 < v < 1$.

$$F_\phi(v) = P(\phi \le v) = P\left(\theta \le \sqrt{v}\right) = \int_0^{\sqrt{v}} \pi(\theta)\, d\theta$$

$$= \sqrt{v}.$$

$$f_\phi(v) = \frac{d}{dv} F_\phi(v) = \frac{d}{dv}\sqrt{v} = \frac{1}{2}\frac{1}{\sqrt{v}}$$

which is no longer uniform. But how come we should have non-uniform prior density for θ^2 when there is full ignorance about θ ?

Jeffreys' Prior: the idea

We want to use a method (M) for choosing a prior density with the following property:
If $\psi = g(\theta)$, g a monotone map, then the density of ψ given by the method (M) is

$$\pi_{\Psi}(\psi) = \pi\left(g^{-1}(\psi)\right) \cdot \left| \frac{d}{d\psi} g^{-1}(\psi) \right|$$

which is the standard probability calculus rule for change of variable in a probability density.

We shall now describe one method (M), i.e., Jeffreys' prior.
In order to introduce Jeffreys' prior we need first to define Fisher information, which will be needed even for purposes other than choice of prior.

Fisher Information of x

A parametric model $x \sim f(x|\theta)$, where $f(x|\theta)$ is differentiable w.r.t to $\theta \in R$, we define $I(\theta)$, *Fisher information* of x, as

$$I(\theta) = \int_{\mathcal{X}} \left(\frac{\partial \log f(x|\theta)}{\partial \theta} \right)^2 f(x|\theta)\, d\mu(x)$$

Conditions for existence of $I(\theta)$ are given in Schervish (1995), p. 111.

Fisher Information of x: An Example

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

Example:

$$f(x|\theta) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\theta)^2/2\sigma^2},$$

σ is known.

$$\frac{\partial \log f(x|\theta)}{\partial \theta} = \frac{(x-\theta)}{\sigma^2}$$

$$I(\theta) = E\left[\frac{(x-\theta)^2}{\sigma^4}\right] = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

Fisher Information of x, $\theta \in R^k$

$x \sim f(x|\theta)$, where $f(x|\theta)$ is differentiable w.r.t to $\theta \in R^k$, we define $I(\theta)$, *Fisher information* of x, as the matrix

$$I(\theta) = \left(I_{ij}(\theta)\right)_{i,j=1}^{k,k}$$

$$I_{ij}(\theta) = \mathrm{Cov}_\theta\left(\frac{\partial \log f(x|\theta)}{\partial \theta_i}, \frac{\partial \log f(x|\theta)}{\partial \theta_j}\right)$$

KTH Matematik

Fisher Information of $x^{(n)}$

Same parametric model $x_i \sim f(x|\theta)$, I.I.D., $x^{(n)} = (x_1, x_2, \ldots, x_n)$.

$$f\left(x^{(n)}|\theta\right) = f(x_1|\theta) \cdot f(x_2|\theta) \cdot \ldots \cdot f(x_n|\theta)$$

Fisher information of $x^{(n)}$ is

$$I_{x^{(n)}}(\theta) = \int_{\mathcal{X}} \left(\frac{\partial \log f\left(x^{(n)}|\theta\right)}{\partial \theta} \right)^2 f\left(x^{(n)}|\theta\right) d\mu\left(x^{(n)}\right)$$

$$= n \cdot I(\theta).$$

KTH Matematik

Fisher Information of x: another form

A parametric model $x \sim f(x|\theta)$, where $f(x|\theta)$ is twice differentiable w.r.t to $\theta \in R$. If we can write

$$\frac{d}{d\theta} \int_{\mathcal{X}} \left(\frac{\partial \log f(x|\theta)}{\partial \theta} \right) f(x|\theta)\, d\mu(x) =$$

$$= \int_{\mathcal{X}} \frac{\partial}{\partial \theta} \left(\frac{\partial \log f(x|\theta)}{\partial \theta} \right) f(x|\theta)\, d\mu(x),$$

then

$$I(\theta) = - \int_{\mathcal{X}} \left(\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right) f(x|\theta)\, d\mu(x)$$

Fisher Information of x, $\theta \in R^k$

$x \sim f(x|\theta)$, where $f(x|\theta)$ is differentiable w.r.t to $\theta \in R^k$, then under some conditions

$$I(\theta) = \left[\left(-E_\theta \left(\frac{\partial^2 \log f(x|\theta)}{\partial\theta_i \partial\theta_j} \right) \right)_{ij} \right]_{i,j=1}^{k,k}$$

For a natural exponential family

$$f(x \mid \theta) = h(x)e^{\theta \cdot x - \psi(\theta)}$$

$$\frac{\partial^2 \log f(x|\theta)}{\partial\theta_i \partial\theta_j} = -\frac{\partial^2 \psi(\theta)}{\partial\theta_i \partial\theta_j}$$

so no expectation needs to be computed to obtain $I(\theta)$.

Jeffreys' Prior defined

$$\pi(\theta) := \frac{\sqrt{I(\theta)}}{\int_{\Theta} \sqrt{I(\theta)} d\theta}$$

assuming that the standardizing integral in the denominator exists. Otherwise the prior is improper.

Jeffreys' Prior is a method (M)

Let $\psi = g(\theta)$, g a monotone map. The prior $\pi(\theta)$ is Jeffreys' prior. Let us compute the prior density $\pi_\Psi(\psi)$ for ψ:

$$\pi_\Psi(\psi) = \pi\left(g^{-1}(\psi)\right) \cdot \mid \frac{d}{d\psi} g^{-1}(\psi) \mid$$

$$\propto \sqrt{E_\theta\left[\left(\frac{\partial \log f(X|\theta)}{\partial\theta}\right)^2\right]} \mid \frac{d}{d\psi} g^{-1}(\psi) \mid$$

$$= \sqrt{E_{g^{-1}((\psi)}\left[\left(\frac{\partial \log f\left(X|g^{-1}(\psi)\right)}{\partial\theta} \frac{d}{d\psi} g^{-1}(\psi)\right)^2\right]}$$

$$= \sqrt{E_{g^{-1}(\psi)}\left[\left(\frac{\partial \log f\left(X|g^{-1}(\psi)\right)}{\partial\psi}\right)^2\right]} = I(\psi)$$

Hence the prior for ψ is the Jeffreys' prior.

KTH Matematik

Maximum Entropy Priors

We are going to discuss maximum entropy prior densities. We need a new definition: Kullback's Information Measure.

Kullback's Information Measure

Let $f(x)$ and $g(x)$ be two densities. Kullback's information measure $I(f;g)$ is defined as

$$I(f;g) := \int_{\mathcal{X}} f(x) \log \frac{f(x)}{g(x)} d\mu(x).$$

We intertpret $\log \frac{f(x)}{0} = \infty$, $0 \log 0 = 0$. It can be shown that $I(f;g) \geq 0$.
Kullback's Information Measure does not require the same kind of conditions for existence as the Fisher information.

Kullback's Information Measure: Two Normal Distributions

Let $f(x)$ and $g(x)$ be densities for $N(\theta_1; \sigma^2)$, $N(\theta_2; \sigma^2)$, respectively.
Then

$$\log \frac{f(x)}{g(x)} = \frac{1}{2\sigma^2}\left[(x-\theta_2)^2 - (x-\theta_1)^2\right]$$

$$I(f;g) = \frac{1}{2\sigma^2} E_{\theta_1}\left[(x-\theta_2)^2 - (x-\theta_1)^2\right]$$

$$= \frac{1}{2\sigma^2}\left[E_{\theta_1}(x-\theta_2)^2 - \sigma^2\right].$$

KTH Matematik

Kullback's Information Measure: Two Normal Distributions

We have

$$E_{\theta_1}(x-\theta_2)^2 = E_{\theta_1}\left(x^2\right) - 2\theta_2 E_{\theta_1}(x) + \theta_2^2$$

$$= \sigma^2 + \theta_1^2 - 2\theta_2\theta_1 + \theta_2^2 = \sigma^2 + (\theta_1-\theta_2)^2 .$$

Then

$$I(f;g) = \frac{1}{2\sigma^2}\left[\sigma^2 + (\theta_1-\theta_2)^2 - \sigma^2\right] =$$

$$= \frac{1}{2\sigma^2}(\theta_1-\theta_2)^2 .$$

$$I(f;g) = \frac{1}{2\sigma^2}(\theta_1-\theta_2)^2$$

KTH Matematik

Kullback's Information Measure: Natural exponential densities

Let $f_i(x) = h(x)e^{\theta_i \cdot x - \psi(\theta_i)}$, $i = 1, 2$. Then

$$I(f_1; f_2) = (\theta_1 - \theta_2) \cdot \nabla_\theta \psi(\theta_1) - (\psi(\theta_1) - \psi(\theta_2))$$

Kullback's Information Measure for Prior Densities

Let $\pi(\theta)$ and $\pi_o(\theta)$ be two densities on Θ

$$I(\pi; \pi_o) = \int_\Theta \pi(\theta) \log \frac{\pi(\theta)}{\pi_o(\theta)} dv(\theta).$$

Here v is another σ-finite measure.

Maximum Entropy Prior (1)

Find $\pi(\theta)$ so that

$$I(\pi; \pi_o) := \int_\Theta \pi(\theta) \log \frac{\pi(\theta)}{\pi_o(\theta)} dv(\theta).$$

is maximized under the constraints (on moments, quantiles e.t.c.)

$$E_\pi [g_k(\theta)] = \omega_k.$$

The method is due to E. Jaynes, see, e.g., his *Probability Theory: The Logic of Science*

KTH Matematik

Maximum Entropy Prior (1)

Find $\pi(\theta)$ so that

$$I(\pi; \pi_o) := \int_{\Theta} \pi(\theta) \log \frac{\pi(\theta)}{\pi_o(\theta)} dv(\theta).$$

is maximized under the constraints (on moments, quantiles e.t.c.)

$$E_{\pi}[g_k(\theta)] = \omega_k.$$

Maybe Winkler's experiments could be redone like this: the assessor gives several ω_k, and maximum entropy π.

Maximum Entropy Prior (2)

This gives, by use of **calculus of variation**,

$$\pi^* (\theta) = \frac{e^{\sum_k \lambda_k g_k(\theta)} \pi_o (\theta)}{\int e^{\sum_k \lambda_k g_k(\eta)} \pi_o (d\eta)},$$

where λ_k are derived from Lagrange multipliers.