# *BAYESIAN INFERENCE*

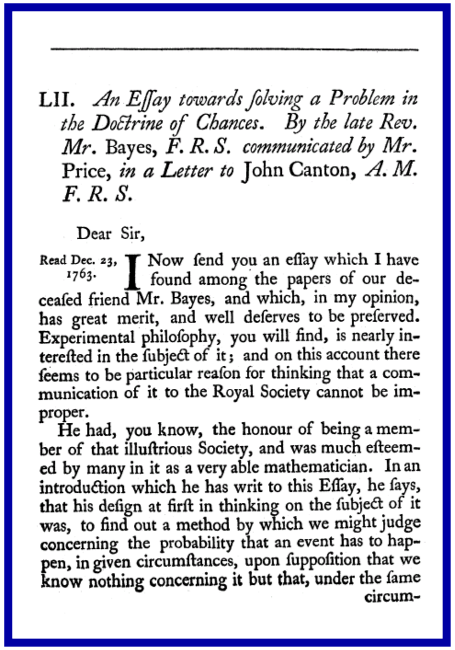
***"The theory of probabilities is basically only common sense reduced to a calculus."***

***P.S. Laplace***

***Carlos Mana***
***Astrofísica y Física de Partículas***
***CIEMAT***

# Some references…

*… The origins…*


LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

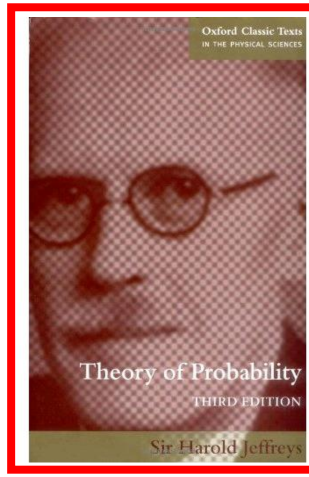
Dear Sir,

Read Dec. 23, 1763. I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion, has great merit, and well deſerves to be preſerved. Experimental philoſophy, you will find, is nearly intereſted in the ſubject of it; and on this account there ſeems to be particular reaſon for thinking that a communication of it to the Royal Society cannot be improper.

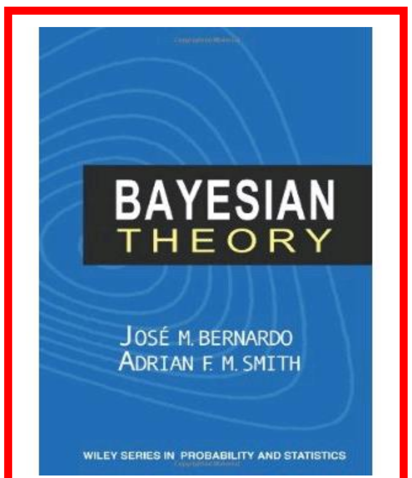
He had, you know, the honour of being a member of that illuſtrious Society, and was much eſteemed by many in it as a very able mathematician. In an introduction which he has writ to this Eſſay, he ſays, that his deſign at firſt in thinking on the ſubject of it was, to find out a method by which we might judge concerning the probability that an event has to happen, in given circumſtances, upon ſuppoſition that we know nothing concerning it but that, under the ſame circum-

*(1763)*

*… to find out a method by which we might judge* **concerning** *the probability that an event has to happen*, **in given circumstances,** *upon supposition that we know nothing concerning it but that, under the same circumstances, it has happened a certain number of times, and failed a certain other number of times*.




*Theory of Probability (1st pub. 1939)*
*Sir Harold Jeffreys*
*Oxford Classic Texts in Physical Sciences*




*Bayesian Theory; J.M. Bernardo, A.F.M. Smith*
*Wiley Series in Probability and Statistics*

*See also:*

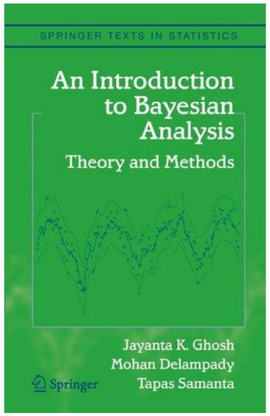
*Bernardo, J.M. (1979) J. Roy. Stat. Soc. Ser. B 41 113-147*

*Berger, J.O., Bernardo J.M., Sun D (2009) Ann. Stat.; Vol. 37, No. 2; 905-938*

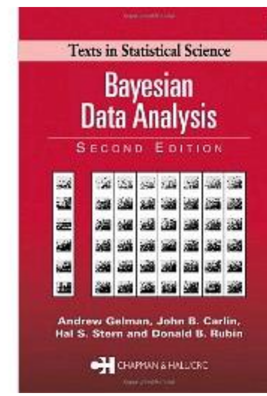
*Bernardo J.M., Ramón J.M. (1998) The Statiscian 47, 1-35*

*Bernardo J.M., "Reference Analysis", http://www.uv.es/~bernardo/RefAna.pdf*

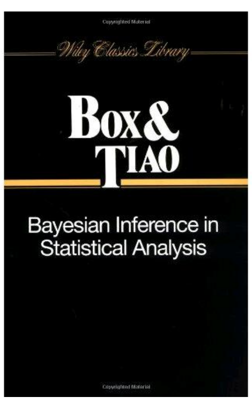
# + many more:




*An Introduction to Bayesian Analysis*
*J.K. Ghosh, M. Delampady, T. Samanta*
*Springer Texts in Statistics*




*Bayesian Data Analysis*
*A. Gelman, J.S. Carlin, H.S. Stern, D.B. Rubin*
*Chapman & Hall*




*Bayesian Inference in Statistical Analysis*
*G.E.P. Box, G.C. Tiao*
*Wiley Classics Library*

## Bayesian inference: Scheme

**1) Model to describe data:**

$$M = \{ p(\boldsymbol{x} \mid \boldsymbol{\lambda}); \quad \boldsymbol{x} \in \Omega_X; \quad \boldsymbol{\lambda} \in \Omega_\Lambda \}$$

$\boldsymbol{\lambda}$: *Unknown parameters about which we want to make inferences from observed data*

*Usually:* $\boldsymbol{\lambda} = \{\boldsymbol{\theta}, \boldsymbol{\varphi}\}$ $\begin{cases} \boldsymbol{\theta} & \text{parameter(s) of interest} \\ \boldsymbol{\varphi} & \text{nuisance parameter(s)} \end{cases}$

**2) Full Model:**

$$p(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\varphi}) = p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\varphi})\, \pi(\boldsymbol{\theta}, \boldsymbol{\varphi})$$

$\pi(\boldsymbol{\theta}, \boldsymbol{\varphi})$ *"Prior density" for the parameters*

**3) Get distribution of unknown parameters conditioned to observed data:**

*Bayes Theorem (Bayes "rule")*

$$p(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\varphi}) = p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\varphi})\, \pi(\boldsymbol{\theta}, \boldsymbol{\varphi}) = p(\boldsymbol{\theta}, \boldsymbol{\varphi} \mid \boldsymbol{x})\, p(\boldsymbol{x})$$

$$p(\boldsymbol{\theta}, \boldsymbol{\varphi} \mid \boldsymbol{x}) = \frac{p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\varphi})\, \pi(\boldsymbol{\theta}, \boldsymbol{\varphi})}{p(\boldsymbol{x})}$$

$supp\{p(\boldsymbol{x})\}$

## *4) Draw inferences on parameters of interest from "posterior" densities:*

$$\pi(\theta,\varphi) = \pi(\varphi \mid \theta)\pi(\theta)$$

***Integrate on nuisance parameters:***

$$p(\theta \mid \mathbf{x}) = \pi(\theta)\int_{\Phi} \frac{p(\mathbf{x} \mid \theta,\varphi)\,\pi(\varphi \mid \theta)}{p(\mathbf{x})}\,d\varphi$$

***…if any, for otherwise simpler expression:***

$$p(\theta \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \theta)\,\pi(\theta)}{p(\mathbf{x})}$$

***… may get conditional densities if needed:***

$$p(\theta \mid \varphi, \mathbf{x}) \propto \frac{p(\mathbf{x} \mid \theta)\,\pi(\theta \mid \varphi)}{\int p(\mathbf{x} \mid \theta)\,\pi(\theta \mid \varphi)d\theta}$$

***Normalisation:***

$$p(\mathbf{x}) = \int_{\Theta\times\Lambda} p(\mathbf{x},\theta,\varphi)d\theta d\varphi = \int_{\Theta\times\Lambda} p(\mathbf{x} \mid \theta,\varphi)\,\pi(\theta,\varphi)d\theta d\varphi$$

***and usually, domain*** $\Theta\times\Lambda$ ***does not depend on parameters so***

$$p(\theta \mid \mathbf{x}) \sim \pi(\theta)\int_{\Phi} p(\mathbf{x} \mid \theta,\varphi)\,\pi(\varphi \mid \theta)d\varphi \qquad p(\theta \mid \mathbf{x}) \sim p(\mathbf{x} \mid \theta)\,\pi(\theta)$$

## 5) Bayesian approach allows, eventually, to make *"predictive inferences"* (… model checking)

**Within same model** $M_X = \{p(\boldsymbol{x}|\boldsymbol{\theta});\quad \boldsymbol{x}\in\Omega_X;\quad \boldsymbol{\theta}\in\Omega_\Theta\}$

**predict new data yet to come** $p(\boldsymbol{x}',\boldsymbol{x},\boldsymbol{\theta}) = p(\boldsymbol{x}',\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \underbrace{p(\boldsymbol{x}'|\boldsymbol{\theta})}\underbrace{p(\boldsymbol{x}|\boldsymbol{\theta})}\,\pi(\boldsymbol{\theta})$

$\boldsymbol{x},\boldsymbol{x}'$ **sampling independent**

$$p(\boldsymbol{x}'|\boldsymbol{x}) = \int_\Theta \frac{p(\boldsymbol{x}',\boldsymbol{x},\boldsymbol{\theta})}{p(\boldsymbol{x})}d\boldsymbol{\theta} = \int_\Theta \overbrace{p(\boldsymbol{x}'|\boldsymbol{\theta})}\left[\frac{\overbrace{p(\boldsymbol{x}|\boldsymbol{\theta})}\,\pi(\boldsymbol{\theta})}{p(\boldsymbol{x})}\right]d\boldsymbol{\theta} =$$

$$= \int_\Theta \underbrace{p(\boldsymbol{x}'|\boldsymbol{\theta})}_{\text{info on } \boldsymbol{x}' \leftarrow \boldsymbol{\theta}}\underbrace{p(\boldsymbol{\theta}|\boldsymbol{x})}_{\boldsymbol{x} \rightarrow \text{info on } \boldsymbol{\theta}}d\boldsymbol{\theta}$$

➡ $\boldsymbol{x},\boldsymbol{x}'$ **related through** $\boldsymbol{\theta}$

# Bayesian inference: Elements

$$p(\boldsymbol{\theta} \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathbf{x})} \propto p(\mathbf{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})$$

***Likelihood:*** ***How data modifies our knowledge on the parameters***

*Experiment affect knowledge on parameters only through likelihood*

*(thus, same likelihoods → same inferences)*

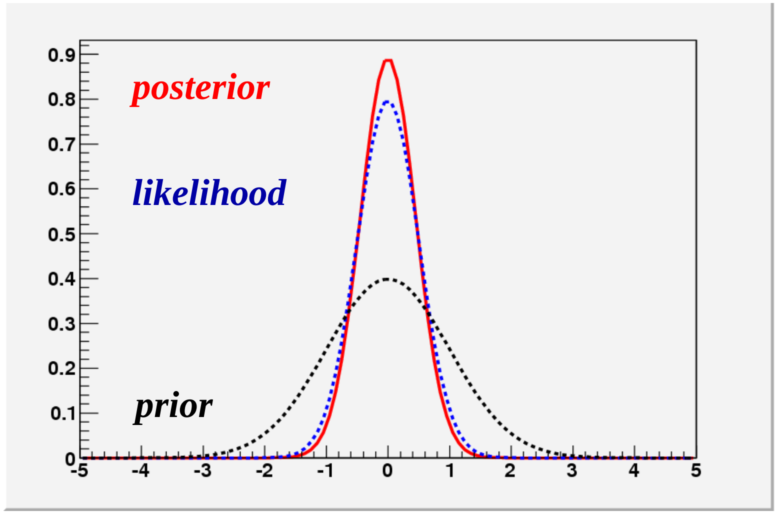
*Defined up to a multiplicative constant*

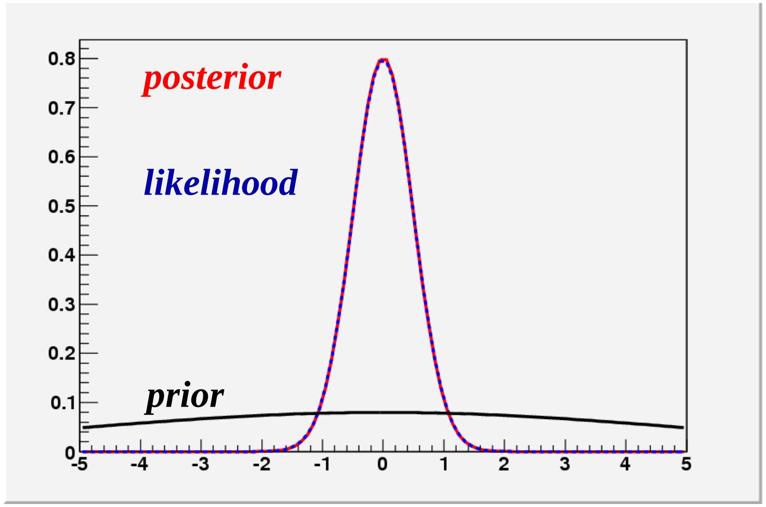
***Prior:*** ***Knowledge ("degree of credibility") about parameters before data is taken***

*Informative:* *Include Prior knowledge (in particular, if trustable info from previous experiments)*

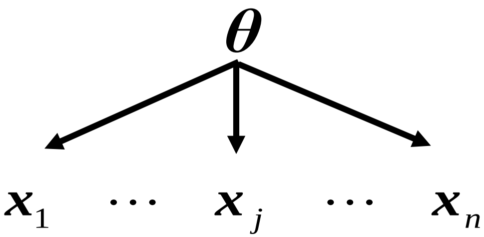
*Non-Informative:* *Relative ignorance before experiment is done (… independent experiment)*

*… posterior is dominated by likelihood*



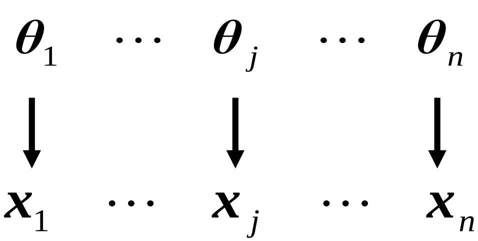



# Bayesian inference: Structures



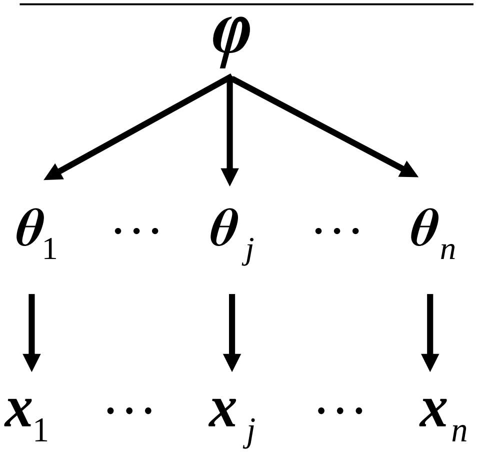

*One experiment (same conditions)*

$$p(\mathbf{x},\theta) = p(\mathbf{x}\,|\,\theta)\,\pi(\theta) = \prod_{i=1}^{n} p(x_i\,|\,\theta)\,\pi(\theta)$$




*Results from isolated experiments ( $\theta$ independent)*

$$p(\mathbf{x},\boldsymbol{\theta}) = p(\mathbf{x}\,|\,\boldsymbol{\theta})\,\pi(\boldsymbol{\theta}) = \prod_{i=1}^{n} p(x_i\,|\,\theta_i)\,\pi(\theta_i)$$




*Parameters* $\{\theta_1,\theta_2,\cdots,\theta_n\}$ *come from a common source governed by hyperparameters* $\boldsymbol{\varphi}$

$$p(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\varphi}) = p(\mathbf{x}\,|\,\boldsymbol{\theta})\;p(\boldsymbol{\theta}\,|\,\boldsymbol{\varphi})\,\pi(\boldsymbol{\varphi})$$

$$= \left\{\prod_{i=1}^{n} p(x_i\,|\,\theta_i)\;p(\theta_i\,|\,\varphi)\right\}\pi(\varphi)$$

**EXAMPLE:** *Acceptance of events after cuts (or efficiency or whatever)*

*Total number of events* **N**

*Number of events that have been observed to pass the cuts* **n**

*Model:*

$$p(n \mid N,\theta) = \binom{N}{n} \theta^{n} (1-\theta)^{N-n}$$

*(B: OK… "a" model)*

## Frequentist / Classical Solution:

*"Estimator" of θ:*
*(maximizing the likelihood)*

$$\theta^{*} = \frac{n}{N}$$

*(B: How Why)*

*$\theta^*$ is a random quantity with:*

*…e(k>>)…*

$$E[\theta^{*}] = \frac{E[n]}{N} = \theta$$

$$\sigma[\theta^{*}] = \frac{\sigma[n]}{\sqrt{N}} \overset{\theta\to\theta^{*}}{\approx} \frac{\sqrt{n(1-n/n)}}{N}$$

*(B: So what?)*
*(B: $\sigma[n]$ ?)*
*(B: did you e(k)?)*
*(B: inferences(e(k)) ?)*
*(B: if n=0, n=N?)*

*Confidence Level Interval:*

$$[\theta^{*} - c_1(n,N), \theta^{*} + c_2(n,N)]$$

*(B: What does it mean?)*

*(B: Does it contain $\theta_{true}$?)*

# Bayes Solution:

**Bayes "rule":** $$p(\theta \mid n, N) \propto [\theta^{n}(1-\theta)^{N-n}]\,\pi(\theta)$$

***(F: θ is a fixed (albeit unknown) number. Can't talk about probabilities…***
***B: Probability is…  + Ramsey, de Finetti, Savage, Lindley, …)***

**Prior density:**

***Bayes-Laplace postulate ("Principle of Insufficient Reason")***
***If no special reason, all possible outcomes are equally likely***

$$\pi(\theta) = \mathbf{1}_{[0,1]}(\theta)$$

**Posterior density:** $$p(\theta \mid n, N) \propto \theta^{n}(1-\theta)^{N-n}\,\mathbf{1}_{[0,1]}(\theta)$$

***(F: Inferences depend on π(θ) ?***

**Inferences on $\theta$ :** $$\theta \sim Be(n+1, N-n+1)$$

$$E[\theta] = \frac{n+1}{N+2} \qquad N >> n >> 1 \qquad \rightarrow \frac{n}{N}$$

***(F: Is biased ?***
***B: So what? n=0,n=N)***

$$mode: \theta_m = \frac{n}{N} \qquad \rightarrow \frac{n}{N}$$

$$V[\theta] = \frac{(n+1)(N-n+1)}{(N+2)^2(N+3)} \qquad \rightarrow \frac{n(1-n/N)}{N^2}$$

$$P[\theta \in (\theta_1, \theta_2)] = \int_{\theta_1}^{\theta_2} p(\theta \mid n, N)\, d\theta$$

# EXAMPLE:

$$p(x\,|\,\tau)=\tau^{-1}e^{-x/\tau} \qquad e(n)=\{x_1,x_2,\ldots,x_n\}$$

## Frequentist / Classical Solution:

**Maximum likelihood:** $\tau^{*}=n^{-1}\sum_{i=1}^{n}x_i$

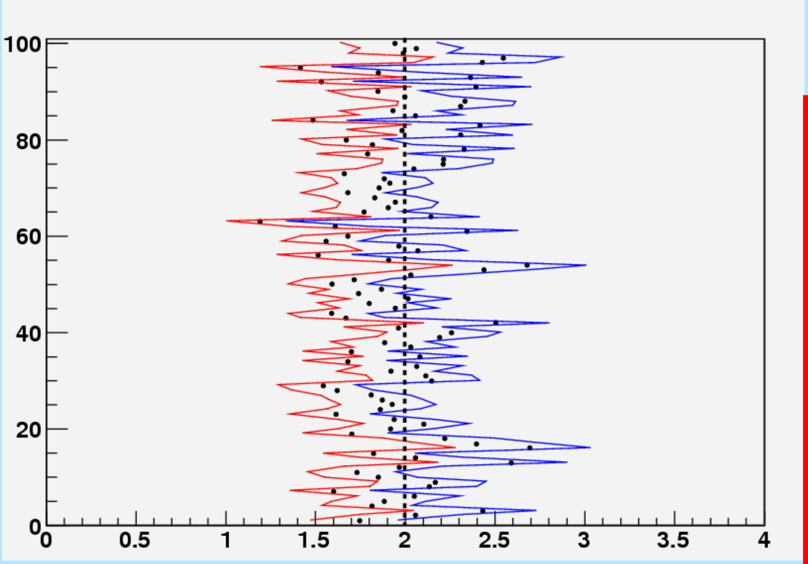
$$p(\tau_*\,|\,\tau,n)=\left(\frac{n}{\tau}\right)^{n}\frac{e^{-n\tau_*/\tau}\,\tau_*^{n-1}}{\Gamma(n)}$$

***100 identical experiments***

$\tau=2 \quad n=50$

$68\%CL \quad intervals \quad \left(c_1(\tau^*,n),c_2(\tau^*,n)\right]$

***…random intervals***




***34 do not contain*** $\tau=2$

***Absolutely different philosophy:***

***Frequentist:*** ***Given the parameters, how likely is the observed sample?,…***

***Great but of hardly any interest;… we are interested in the parameters***

***Bayesian:*** ***Given the data, draw inferences on the parameters,…***

***…you do one sampling (n=50)…***

*In the Bayesian Solution for the Binomial example, we took a* ***uniform prior density:***

**Bayes-Laplace postulate** *("Principle of Insufficient Reason"; J. Bernoulli)*
***If no special reason, all possible outcomes are equally likely***

1) *Not always reasonable choice*

2) *Consistency: Non-linear one-to-one transformation* $\phi = \phi(\theta)$

*Ex:* $\phi = e^{\theta}$ $$\pi(\theta)d\theta \propto \mathbf{1}_{\Theta}(\theta)d\theta \to \pi(\phi)d\phi \propto \mathbf{1}_{\phi(\Theta)}(\phi)\frac{1}{\phi}d\phi$$

***Task:*** ***Reasonable and sound criteria to choose a proper prior density:***

***Several arguments:***

1) *Invariances* *(Jeffreys (1939))*
2) *Conjugated Priors* *(Raiffa+Schlaifer (1961))*
3) *Probability Matching Priors* *(Welch+Peers (1963) … Ghosh, Datta, Mukerjee,…)*
4) *Reference Priors* *(Bernardo (1979), Berger)*
5) *Hierarchical Structures*
6) *… more…*

# *Before we start with priors… two important concepts:*

1) ***Sufficient Statistics***
2) ***Exchangeable sequences***

***(… make life easier)***

***(Frequentist / Classical care about unbiased, consistency, asymptotic efficiency, minimum variance,…)***

**STATISTICS:**

$$X \to e(m) = \{x_1, x_2, \ldots, x_m\} \qquad \boldsymbol{X} = (X_1, X_2, \ldots, X_m)$$

$$x_i \in \Omega_i$$

**Statistic:** $\boldsymbol{t} : (x_1, \ldots, x_m) \in \Omega_1 \times \cdots \times \Omega_m \to R^{k(m)} \qquad \dim(\boldsymbol{t}) = k(m)$

**Sufficient:** ***Statistics that have all relevant info on parameters of interest***

$$\text{iff} \quad \forall m \geq 1 \quad \text{and} \quad \forall \pi(\boldsymbol{\theta}) \qquad p(\boldsymbol{\theta} \mid x_1, x_2, \ldots, x_m) = p(\boldsymbol{\theta} \mid \boldsymbol{t})$$

**Sufficient and Minimal:** $k(m) = \dim(\boldsymbol{t}) < m$

***If they exist, we do not have to work with the whole sample***

***(better if*** $k(m) = k$ ***independent of m )***

**Example:**

$$M_X = \{N(x \mid \mu, \sigma); x \in R; \quad (\mu, \sigma) \in R \times R^+\}$$

$$X \overset{iid}{\to} e(m) = \{x_1, x_2, \ldots, x_m\}$$

$$p(x_1, x_2, \ldots, x_m \mid \mu, \sigma) \propto \sigma^{-m} \exp\left\{-\frac{1}{2}\sum_{i=1}^{m}\left(\frac{x_i - \mu}{\sigma}\right)^2\right\}$$

$$\boldsymbol{t} = \left(m, \sum_{i=1}^{m} x_i, \sum_{i=1}^{m} x_i^{\,2}\right)$$

$$p(data \mid \mu, \sigma) \propto \sigma^{-t_1} \exp\left\{-\frac{1}{2\sigma^2}\left(t_3 - 2\mu t_2 + \mu^2 t_1\right)\right\}$$

***Except some irregular cases, the only distributions that admit a fixed number of minimal sufficient statistics*** $(k(m) = k < m)$ ***belong to the*** ***exponential family***

**Exponential Family** $M = \{p(\boldsymbol{x} \mid \boldsymbol{\theta}); \quad \boldsymbol{x} \in \Omega_X; \quad \boldsymbol{\theta} \in \Omega_\theta\}$

***Belongs to the k-parametric exponential family if:***

$$p(\boldsymbol{x} \mid \boldsymbol{\theta}) = f(\boldsymbol{x}) g(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} c_i \phi_i(\boldsymbol{\theta}) h_i(\boldsymbol{x})\right\}$$

$$g^{-1}(\boldsymbol{\theta}) = \int_{\Omega_X} f(\boldsymbol{x}) g(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} c_i \phi_i(\boldsymbol{\theta}) h_i(\boldsymbol{x})\right\} dx < \infty$$

***"regular" family if*** $\Omega_X$ ***does not depend on*** $\theta$

***… more than what you think …***

***… and less than what we would like…***

## *Few examples:*

$$p(\boldsymbol{x}\mid\boldsymbol{\theta}) = f(\boldsymbol{x})g(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^{k} c_i \phi_i(\boldsymbol{\theta}) h_i(\boldsymbol{x})\right\}$$

$$X \sim Po(n\mid\mu) = \frac{e^{-\mu}\mu^n}{\Gamma(n+1)} = \frac{e^{-(\mu - n\ln\mu)}}{\Gamma(n+1)}$$

$$X \sim Bi(n\mid N,\theta) = \binom{N}{n}\theta^n(1-\theta)^{N-n} = \binom{N}{n}\exp\{n\ln\theta + (N-n)\ln(1-\theta)\}$$

$$X \sim Ga(x\mid a,b) = \frac{a^b}{\Gamma(b)}\exp\{-ax + (b-1)\ln x\}$$

$$\cdots$$

$$X \sim Ca(x\mid\alpha,\beta) \propto \frac{1}{1+\beta(x-\alpha)^2} = \exp\left\{-\ln\left(1+\beta(x-\alpha)^2\right)\right\}$$

$$X \sim p(x\mid a) = \frac{3}{8}(1+x^2) + ax \qquad x\in[-1,1]$$

$$X \sim p(x\mid a) = ap_1(x) + (1-a)p_2(x)$$

***No Sufficient and Minimal Statistics → work with the whole sample***

# 2) EXCHANGEABLE SEQUENCES:

$$X \sim p(x|\theta) \qquad X \rightarrow e(m) = \{x_1, x_2, \ldots, x_m\}$$

**The sequence** $\{x_1, x_2, \ldots, x_m\}$ **is exchangeable if the joint density** $p(x_1, x_2, \ldots, x_m \mid \theta)$ **is invariant under any permutation of indices**

→ ***Symmetry in observations so order in which are taken is irrelevant***

***Any sequence*** $\{x_1, x_2, \ldots, x_m\}$ ***with*** $\boldsymbol{X} = \{X_1, X_2, \ldots, X_m\}$ ***conditionally independent and identically distributed (iid) , is exchangeable***

***Converse not true: exchangeable sequences are identically distributed but not necessarily independent***

$$X \overset{iid}{\rightarrow} e(m) = \{x_1, x_2, \ldots, x_m\}$$

$$p(x_1, x_2, \ldots, x_m \mid \theta) = \prod_{i=1}^{m} p(x_i \mid \theta)$$

$$\log p(x_1, x_2, \ldots, x_m \mid \theta) = \sum_{i=1}^{m} \log p(x_i \mid \theta)$$

# Now …PRIORS

**Model:** $M = \{ p(\boldsymbol{x} \mid \boldsymbol{\theta}); \quad \boldsymbol{x} \in \Omega_X ; \quad \boldsymbol{\theta} \in \Omega_{\theta} \}$

**Posterior:** $p(\theta \mid data) \propto p(data \mid \theta)\, \pi(\theta)$

**Prior ~** ***Knowledge ("degree of credibility") about parameters before data is taken***

***What "prior" info do we have on parameters of interest?***

***Informative:***

***Include Prior knowledge (for instance, if trustable info from previous experiments)***

***Non-Informative:***

***Relative ignorance before experiment is done (or independent analysis,…)***

# *"~Non-Informative" …* ***reference priors***

***1) "Used as standard" such that posterior is dominated by likelihood***

***2) "Non-Informative"….One is never in a state of complete ignorance***
***"knowing little a priori"… is relative to info provided by experiment***

***3) Usually are improper densities:***
***… do not quantify prior knowledge on parameters in a pdf …***
***(sequence of compact coverings, admissible models, …)***
***… do we really need a proper prior?...***

***Task: Specify a prior which provides little info relative to what is expected to be provided by experiment***

***General advise: Prior will have some (hopefully small) effect on inferences so try with several reasonable priors …***

***Priors derived from INVARIANCE ARGUMENTS***

# INVARIANCE: Position and scale parameters

**Position parameter:** $X \sim p(x|\theta) = f(x-\theta)$

$$\pi(\theta) \longrightarrow p(x,\theta)dxd\theta = p(x|\theta)\pi(\theta)dxd\theta = f(x-\theta)\pi(\theta)dxd\theta$$

***New r.q.:*** $Y = X + a\,;\quad J(Y;X) = 1$ $\qquad p(y,\theta)dyd\theta = f(y-a-\theta)\pi(\theta)dyd\theta$

**Reparametrization:** $\theta' = \theta + a\,;\quad J(\theta';\theta) = 1$ $\qquad p(y,\theta')dyd\theta' = f(y-\theta')\pi(\theta'-a)dyd\theta'$

***Model:*** $f(x-\theta)$ $\qquad X \overset{iid}{\longrightarrow} e(m) = \{x_1, x_2, \ldots, x_m\}$ ***inferences on*** $\theta$

***formally analogous to***

***Model:*** $f(y-\theta')$ $\qquad Y \overset{iid}{\longrightarrow} e(m) = \{y_1, y_2, \ldots, y_m\}$ ***inferences on*** $\theta'$

***Prior knowledge on*** $\theta$ ***same as on*** $\theta'$

$$\pi(\theta) = \pi(\theta'-a);\quad \forall a \in R \quad \longrightarrow \quad \boxed{\pi(\theta) = c\,\mathbf{1}_{\Theta}(\theta)}$$

# Scale parameter:

$$X \sim p(x|\theta) = \frac{1}{\theta} f\left(x/\theta\right)$$

$$\pi(\theta) \longrightarrow p(x,\theta)dxd\theta = p(x|\theta)\pi(\theta)dxd\theta = \frac{1}{\theta} f\left(x/\theta\right)\pi(\theta)dxd\theta$$

***New r.q.:*** $Y = aX; \quad J(Y;X) = 1/a \qquad p(y,\theta)dyd\theta = \frac{1}{a\theta} f\left(y/a\theta\right)\pi(\theta)dyd\theta$

***Reparametrization:*** $\theta' = a\theta; \quad J(\theta';\theta) = 1/a \qquad p(y,\theta')dyd\theta' = \frac{1}{\theta'} f\left(y/\theta'\right)\left[\frac{1}{a}\pi\left(\theta'/a\right)\right]dyd\theta'$

***Model:*** $\frac{1}{\theta} f\left(x/\theta\right) \qquad X \xrightarrow{iid} e(m) = \{x_1, x_2, \ldots, x_m\}$ ***inferences on*** $\theta$

***formally analogous to***

***Model:*** $\frac{1}{\theta'} f\left(y/\theta'\right) \qquad Y \xrightarrow{iid} e(m) = \{y_1, y_2, \ldots, y_m\}$ ***inferences on*** $\theta'$

***Prior knowledge on*** $\theta$ ***same as on*** $\theta'$

$$\frac{1}{a}\pi\left(\frac{\theta'}{a}\right) = \pi(\theta); \quad \forall a \in R/\{0\} \longrightarrow \boxed{\pi(\theta) = 1/\theta \, \mathbf{1}_{\Theta}(\theta)}$$

# *For Location and Scale parameters:*

$$\pi(\mu,\theta) = \pi(\mu)\pi(\theta) \propto {}^{1}\!/_{\theta}\,\mathbf{1}_{\Theta}(\theta)\mathbf{1}_{\mathrm{M}}(\mu)$$

***(considered independent)***

***… Many important models with LOCATION AND SCALE parameters:***

**Example:**

$$X \sim Ex(x|\theta) \qquad p(x|\theta) = \theta e^{-x\theta}\mathbf{1}_{[0,\infty)}(x)$$

$\theta$ ***unknown parameter of interest***

$$X \xrightarrow{iid} e(m) = \{x_1, x_2, \ldots, x_m\} \qquad p(x_1, x_2, \ldots, x_n|\theta) = \theta^n e^{-\theta(x_1+x_2+\cdots+x_n)}$$

***Given n,*** $t = \sum_{i=1}^{n} x_i$ ***is sufficient statistic***

$$p(t|\theta, n) = \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t}$$

$\theta^{-1}$ ***is scale parameter*** $\Rightarrow \pi(\theta^{-1}) \propto 1/\theta^{-1} \Rightarrow \pi(\theta) \propto 1/\theta$

$$p(\theta|t,n) \propto p(t|\theta,n)\pi(\theta)$$

$$p(\theta|t,n) = \frac{t^n}{\Gamma(n)}\theta^{n-1}e^{-\theta t} \qquad \Theta \sim Ga(\theta|t,n)$$

---

***... consistency:***

***Model:*** $z = \ln t \qquad \phi = \ln\theta \qquad p(z|\phi,n) = \frac{1}{\Gamma(n)}\exp\{n(\phi - z) - e^{\phi - z}\}$

$\phi = \ln\theta$ ***Position parameter*** $\qquad \pi(\phi) = c \rightarrow \pi(\theta) \propto 1/\theta$

**Example:**

$$X \sim Un(x|0,\theta) = \frac{1}{\theta}\mathbf{1}_{(0,\theta)}(x) \qquad \Omega_X = (0,\theta)$$

$\theta$ ***unknown parameter of interest***

$$X \xrightarrow{iid} e(m) = \{x_1, x_2, ..., x_m\} \qquad p(x_1, x_2, ..., x_m \mid \theta) = \frac{1}{\theta^m}\mathbf{1}_{(0,\theta)}(\max\{x_i\})$$

***Sufficient statistic*** $t = \max\{x_1, x_2, ..., x_n\}$

$$p(t \mid n, \theta) = n\left(\frac{t}{\theta}\right)^{n-1}\frac{1}{\theta}\mathbf{1}_{(0,\theta)}(t)$$

$\theta$ ***Scale parameter*** ➡ $\pi(\theta) \propto 1/\theta$

$$p(\theta \mid t, n) \propto p(t \mid \theta, n)\pi(\theta)$$

$$p(\theta \mid n, t) = n\frac{t^n}{\theta^{n+1}}\mathbf{1}_{(t,\infty)}(\theta) \qquad \Theta \sim Pa(\theta \mid n, t)$$

***(by the way, Frequentists and bias:*** $\theta_{ML}^{freq} = t$ ***and*** $E[\theta] = \frac{n}{n-1}t \overset{n>>}{\approx} t + t/n$***)***

**Problem:**

$$X \sim N(x|\mu,\sigma) \qquad e(n) \xrightarrow{iid} \boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$$

**1) Show that** $\left\{ n; \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \right\}$ **is a set of sufficient statistics**

**2) Being** $\{\mu,\sigma\}$ **location and scale parameters, take** $\pi(\mu,\sigma) \propto 1/\sigma$ **as (improper) prior and show that**

$$T = \sqrt{n-1}\left(\frac{\mu - \bar{x}}{s}\right) \sim St(t\,|\,n-1) \qquad Z = n\left(\frac{s^2}{\sigma^2}\right) \sim \chi^2(z\,|\,n-1)$$

⟶ **inferences on** $\mu$

⟶ **inferences on** $\sigma$

$E[T] = 0 \quad (n > 2)$

$V[T] = (n-1)(n-3)^{-1} \quad (n > 3)$

…

$E[Z] = n-1 \quad (n > 1)$

$V[Z] = 2(n-1) \quad (n > 1)$

…

# Problem: Comparison of means and variances

**Let** $X_1 \sim N(x|\mu_1,\sigma_1)$ **and** $X_2 \sim N(x|\mu_2,\sigma_2)$ **be independent r.q.**

**Consider the samplings** $\boldsymbol{x}_1=\{x_{11},x_{12},\dots,x_{1n_1}\}\ \ (iid)\qquad \boldsymbol{x}_2=\{x_{21},x_{22},\dots,x_{2n_2}\}\ \ (iid)$

**and the sufficient statistics** $\left\{n_k;\quad \bar{x}_k=\frac{1}{n_k}\sum_{i=1}^{n_k}x_{ki};\quad s_k^{\,2}=\frac{1}{n_k}\sum_{i=1}^{n_k}(x_{ki}-\bar{x}_k)^2\right\}_{k=1,2}$

**1) Show that for the priors (improper)**

**Comparison of variances:** $\pi(\mu_1,\sigma_1,\mu_2,\sigma_2)\propto 1/\sigma_1\sigma_2$

$$Z=\frac{n_2(n_1-1)}{n_1(n_2-1)}\left(\frac{\sigma_1^{\,2}/s_1^{\,2}}{\sigma_2^{\,2}/s_2^{\,2}}\right)\sim Sn(z\,|\,n_2-1,n_1-1)$$

**Comparison of means:**

**H1: same (but unknown) variances:** $\sigma_1=\sigma_2=\sigma \qquad \pi(\mu_1,\mu_2,\sigma)\propto 1/\sigma$

$$T=\left(\frac{(\mu_1-\mu_2)-(\bar{x}_1-\bar{x}_2)}{s\,(1/n_1+1/n_2)^{1/2}}\right)\sim St(t\,|\,n_1+n_2-2) \qquad s=\frac{n_1s_1^{\,2}+n_2s_2^{\,2}}{n_1+n_2-2}$$

**H2: unknown and different variances:** $\pi(\mu_1,\sigma_1,\mu_2,\sigma_2)\propto 1/\sigma_1\sigma_2$

$$W=\mu_1-\mu_2\sim \quad p(w\,|\,\boldsymbol{x}_1,\boldsymbol{x}_2)\sim\int_{-\infty}^{\infty}\left(s_1^{\,2}+(\bar{x}_1-w-u)^2\right)^{-(n_1+1/2)}\left(s_2^{\,2}+(\bar{x}_2-u)^2\right)^{-(n_2+1/2)}du$$

**(Behrens-Fisher problem)**

# INVARIANCES: under a Group of Transformations

$X \sim p(x|\theta)$ **Model:** $M = \{p(x|\theta);\, x \in \Omega_X \,;\, \theta \in \Omega_\theta\}$

**Consider a Group of transformations that acts**

**on the sample space:** $x \to x' = gx \quad ; g \in G;\, x, x' \in \Omega_X$

**on the parametric space:** $\theta \to \theta' = g\theta \quad ; g \in G;\, \theta, \theta' \in \Omega_\theta$

| |
|---|
| **The model M is invariant under G if** $\forall g \in G \,;\; \forall \theta \in \Omega_\theta$ **the random quantity** $X' = gX$ **is distributed as** $p(gx|g\theta) = p(x'|\theta')$ |

► ***The action of the group on the sample and parameter space may be different***

**SCHEME:**

1) ***Identify the group of transformations on sample space (if exists)***

***This induces the action of the group G on the parameter space under which the model is invariant***

**Examples:** **Translations:** $x \rightarrow x' = gx = a + x$

**Scaling:** $x \rightarrow x' = gx = bx$

**Affine:** $x \rightarrow x' = gx = a + bx$

**Rotations:** $\mathbf{x} \rightarrow \mathbf{x}' = g\mathbf{x} = R\mathbf{x}$ *(… matrix Transforms)*

**Important Case:**

**Location and Scale Parameters:** $$p(x \mid \mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$$

**Affine group** $G = \{g_{a,b} := (a,b);\, a \in R\,;\, b \in R^{+}\}$

**On sample Space** $x \in R$ $\qquad x \rightarrow x' = g_{a,b}x = a + bx$

**Model Invariant if on Parameter Space** $(\mu, \sigma) \in R \times R^{+}$

$$(\mu, \sigma) \rightarrow (\mu, \sigma)' = g_{a,b}(\mu, \sigma) =$$
$$= (a + b\mu, b\sigma)$$

---

**Group:** $\forall x \in R \rightarrow g_{a,b}x \in R$

$$g_{c,d}(g_{a,b}x) = g_{c,d}(a + bx) = (a + da + dbx) = g_{a+da,db}x$$

$$g_I = (0,1) \qquad g_{a,b}^{-1} = (-a/b, 1/b)$$

## 2) Choose a prior density that is invariant under this group so that:

→ *Initial transformations of data will make no difference on inferences*

→ *Same Prior Beliefs on original and transformed parameters*

*(…if we want to incorporate this as part of prior info)*

*For the action of group G of transformations, there exists an invariant measure $\mu$ (Haar measure) such that*

$$\int_{\Omega} f(gx)d\mu(x) = \int_{\Omega} f(x')d\mu(x')$$

*for any Lebesgue integrable function* $f(x)$

*Alfred Harr (1933)*

► *Unique (up to a multiplicative constant): Von Neuman (1934); Weil and Cartan (1940)*

**In our case:** $x \rightarrow \theta \qquad f(x) \rightarrow p(\bullet|\theta)\,\mathbf{1}_{\Theta}(\theta)$

*and the measure we are looking for:* $\mu \ll \text{Lebesgue} \rightarrow d\mu(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta})d\boldsymbol{\theta}$

*Group acting on the **LEFT**:*

$$\int_{\Delta\subseteq R} p(\bullet|g\theta)\pi(\theta)d\theta = \int_{\Delta'\subseteq R} p(\bullet|\theta')\pi(\theta')d\theta'$$

***RIGHT:***

$$\int_{\Delta\subseteq R} p(\bullet|\theta g)\pi(\theta)d\theta = \int_{\Delta'\subseteq R} p(\bullet|\theta')\pi(\theta')d\theta'$$

**Example:** **Location and Scale Parameters** $\boldsymbol{\theta} = (\mu, \sigma)$

**Group acting on the LEFT:**

$$\underbrace{\int_{\Delta \subseteq R} p(\bullet \mid g\theta)\pi(\theta)d\theta}_{d\mu(\theta)} = \int_{\Delta' \subseteq R} p(\bullet \mid \theta')\underbrace{\pi(\theta')d\theta'}_{d\mu(\theta')}$$

$$\int_{\Delta \subseteq R} p(\bullet \mid g(\mu,\sigma))\pi(\mu,\sigma)d\mu d\sigma = \int_{\Delta' \subseteq R} p(\bullet \mid \mu',\sigma')\pi(\mu',\sigma')d\mu' d\sigma'$$

**Affine group** $G = \left\{ g_{a,b} := (a,b);\ a \in R\ ;\ b \in R^{+} \right\}$

**...on the left:**

$$g_{a,b}(\mu,\sigma) \equiv (\mu',\sigma') = (a + b\mu, b\sigma)$$

$$(\mu,\sigma) = g_{a,b}^{-1}(\mu',\sigma') = \left( \frac{\mu' - a}{b}, \frac{\sigma'}{b} \right)$$

$$J(\mu',\sigma';\mu,\sigma) = \frac{1}{b^2}$$

$$\int_{\Delta \subseteq R} p(\bullet \mid g(\mu,\sigma))\pi(\mu,\sigma)d\mu d\sigma$$

$$\Rightarrow \int_{\Delta' \subseteq R} p(\bullet \mid \mu',\sigma')\left[\pi(g^{-1}(\mu',\sigma'))J(\mu',\sigma';\mu,\sigma)\right]d\mu' d\sigma'$$

$\forall \Delta \subseteq R$

$$\pi(g^{-1}(\mu',\sigma'))J(\mu',\sigma';\mu,\sigma)d\mu' d\sigma' = \pi\left( \frac{\mu' - a}{b}, \frac{\sigma'}{b} \right)\frac{1}{b^2}d\mu' d\sigma' = \pi(\mu',\sigma')d\mu' d\sigma'$$

$$\pi\left(\frac{\mu'-a}{b},\frac{\sigma'}{b}\right)\frac{1}{b^2}d\mu' d\sigma' = \pi(\mu',\sigma')d\mu' d\sigma' \qquad \forall \mu', a \in R \quad ; \quad \forall \sigma', b \in R^+$$

*LEFT Haar Invariant Measure* → $$\pi_L(\mu,\sigma) \propto \frac{1}{\sigma^2}$$

***Group acting on the RIGHT:***

$$\int_{\Delta\subseteq R} p(\bullet \mid \theta g)\pi(\theta)d\theta = \int_{\Delta'\subseteq R} p(\bullet \mid \theta')\pi(\theta')d\theta'$$

$$(\mu',\sigma') = (\mu,\sigma)g_{a,b} = (\mu+\sigma a, \sigma b)$$

$$(\mu,\sigma) = (\mu',\sigma')g_{a,b}^{-1} = \left(\mu'-\sigma'\frac{a}{b},\frac{\sigma'}{b}\right)$$

$$J(\mu',\sigma';\mu,\sigma) = \frac{1}{b}$$

*RIGHT Haar Invariant Measure* → $$\pi_R(\mu,\sigma) \propto \frac{1}{\sigma}$$

►*For the Group of affine transformations, LEFT and RIGHT invariant Haar measures are not the same*

$$\pi_L(\mu,\sigma) \propto \frac{1}{\sigma^2} \qquad\qquad \pi_R(\mu,\sigma) \propto \frac{1}{\sigma}$$

***0) As invariant measures, both are OK.***

***1) For models with one parameter, prior derived from RIGHT invariant measure coincides with Jeffreys' prior (will came in a while)***

***2) Prior derived from RIGHT invariant measure is probability matching (will come in a while²)***

***3) For models with several parameters, prior derived from LEFT Haar measure coincide with Jeffreys'***

***(and, as we shall see, gives problems …although is interesting too…)***

***Common approach:***

►***Usually there is no invariance ("obvious?") of the model but…***

►***If the structure of the model has a group invariance, … RIGHT Haar prior***

## Problem: Bivariate Normal

$$(X_1, X_2) \sim N(x_1, x_2 | 0, 0, \Sigma)$$

$$\boldsymbol{x}^t = (x_1, x_2)$$

$$p(\boldsymbol{x} \mid \boldsymbol{\Sigma}) = (2\pi)^{-1} |\Sigma|^{-1/2} \exp\left\{-1/2\left[\boldsymbol{x}^t \Sigma^{-1} \boldsymbol{x}\right]\right\}$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$|\Sigma| = \sigma_1^2 \sigma_2^2 (1 - \rho^2)$$

1) *Obtain Right and Left Invariant Haar measures under the action of Lower (Upper) Triangular Matrix Group*

---

## Solution:

1) **Cholesky decomposition in *lower triangular matrices***

$$\Sigma^{-1} = \frac{1}{|\Sigma|} \begin{pmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{pmatrix} = A^t A$$

$$A = \begin{pmatrix} \dfrac{1}{\sigma_1} & 0 \\ -\dfrac{\rho}{\sigma_1\sqrt{1-\rho^2}} & \dfrac{1}{\sigma_2\sqrt{1-\rho^2}} \end{pmatrix}$$

$$|\Sigma| = \frac{1}{|\Sigma^{-1}|} = \frac{1}{|A|^2}$$

*New parametrization of the model* $\boldsymbol{A} = \{A_{11}, A_{21}, A_{22}\}$

$$p(\boldsymbol{x} \mid \boldsymbol{A}) = (2\pi)^{-1} |A| \exp\left\{-\frac{1}{2}\left[\boldsymbol{x}^t A^t A \boldsymbol{x}\right]\right\}$$

2) Group of *lower triangular 2x2 matrices*

$$G = \{T \in LT_{2\times 2}; \quad T_{ii} > 0\}$$
$$T^{-1} \in G$$
$$TT^{-1} = T^{-1}T = \mathbf{1}$$

*…and the action on the Sample Space*

**On the Left**

$$\left[\boldsymbol{x}^t (T^t (T^t)^{-1}) A^t A (T^{-1}T) \boldsymbol{x}\right]$$
$$T \circ \boldsymbol{x} \to T\boldsymbol{x} = \boldsymbol{x}'$$
$$M = T$$

**On the Right**

$$\left[\boldsymbol{x}^t ((T^t)^{-1} T^t) A^t A (TT^{-1}) \boldsymbol{x}\right]$$
$$\boldsymbol{x} \circ T \to T^{-1}\boldsymbol{x} = \boldsymbol{x}'$$
$$M = T^{-1}$$

$$M\boldsymbol{x} = \boldsymbol{x}' \quad \boldsymbol{x} = M^{-1}\boldsymbol{x}' \quad \boldsymbol{x}'^t = \boldsymbol{x}^t M^t \quad d\boldsymbol{x} = \frac{1}{|M|} d\boldsymbol{x}'$$

$$p(\boldsymbol{x}' | \boldsymbol{A}) = (2\pi)^{-1} \frac{|A|}{|M|} \exp\left\{-1/2\left[\boldsymbol{x}'^t (AM^{-1})^t (AM^{-1}) \boldsymbol{x}'\right]\right\}$$

► *Model is invariant under G if the action on the Parameter Space is:*

$$G : A \mapsto A' = AM^{-1} \quad A'M = A \quad |A| = |A'||M|$$

$$p(\boldsymbol{x}' | \boldsymbol{A}') = (2\pi)^{-1} |\boldsymbol{A}'| \exp\left\{-1/2\left[\boldsymbol{x}'^t (A'^t A') \boldsymbol{x}'\right]\right\}$$

***Haar Equation:*** $\int p(\bullet \mid gA)\pi(A)dA = \int p(\bullet \mid A')\pi(A')dA'$

$$\int p(\bullet \mid A')\pi(A'M)J(A';A)dA' = \int p(\bullet \mid A')\pi(A')dA'$$

$$\forall M \in G \longrightarrow \boxed{\pi(A'M)J(A';A)dA'_{11}\,dA'_{21}\,dA'_{22} = \pi(A')dA'_{11}\,dA'_{21}\,dA'_{22}}$$

$$M = T = \begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \qquad J(A';A) = a^2 c$$

$$\pi(aA'_{11}, aA'_{21} + bA'_{22}, cA'_{22})(a^2 c)dA'_{11}\,dA'_{21}\,dA'_{22} = \pi(A')dA'_{11}\,dA'_{21}\,dA'_{22}$$

$$\boxed{\pi(A'_{11}, A'_{21}, A'_{22}) \propto \frac{1}{A'^{\,2}_{11} A'_{22}}}$$

$$M = T^{-1} = \frac{1}{ac}\begin{pmatrix} c & 0 \\ -b & a \end{pmatrix} \qquad J(A';A) = \frac{1}{ac^2}$$

$$\pi\left(\frac{A'_{11}}{a}, \frac{cA'_{21} - bA'_{22}}{ac}, \frac{A'_{22}}{c}\right)\left(\frac{1}{ac^2}\right)dA'_{11}\,dA'_{21}\,dA'_{22} = \pi(A')dA'_{11}\,dA'_{21}\,dA'_{22}$$

$$\boxed{\pi(A'_{11}, A'_{21}, A'_{22}) \propto \frac{1}{A'_{11} A'^{\,2}_{22}}}$$

***Transform back to parameters of interest:*** $\{\sigma_1, \sigma_2, \rho\}$

$$A = \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ -\frac{\rho}{\sigma_1\sqrt{1-\rho^2}} & \frac{1}{\sigma_2\sqrt{1-\rho^2}} \end{pmatrix}$$

$$dA_{11}dA_{21}dA_{22} = \frac{1}{\sigma_1^{\ 3}\sigma_2^{\ 2}(1-\rho^2)^2}\, d\sigma_1 d\sigma_2 d\rho$$

$$M = T \qquad T \circ \mathbf{x} \to T\mathbf{x} = \mathbf{x}'$$

$$M = T^{-1} \qquad \mathbf{x} \circ T \to T^{-1}\mathbf{x} = \mathbf{x}'$$

$$\pi_{LH}(\sigma_1, \sigma_2, \rho) \propto \frac{1}{\sigma_1\sigma_2(1-\rho^2)^{3/2}}$$

$$= \pi_J(\sigma_1, \sigma_2, \rho)$$

$$\pi_{RH}(\sigma_1, \sigma_2, \rho) \propto \frac{1}{\sigma_1^{\ 2}(1-\rho^2)}$$

---

***Upper Triangular:***

$$T = \begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$$

$$\pi_{RH}(\sigma_1, \sigma_2, \rho) \propto \frac{1}{\sigma_2^{\ 2}(1-\rho^2)}$$

$$\pi_{LH}(\sigma_1, \sigma_2, \rho) = \pi_J(\sigma_1, \sigma_2, \rho)$$

# INVARIANCE under reparameterisations (Sir H. Jeffreys; 1950)

**Any criteria to specify a prior density $\pi(\theta)$ for $\theta$ should give a consistent result for parameter $\phi = h(\theta)$ with $h(\bullet)$ a single-valued function**

$$d\phi = \left|\frac{\partial\phi}{\partial\theta}\right| d\theta \longrightarrow \pi(\theta)d\theta = \pi(\phi)d\phi = \pi(\phi(\theta))\left|\frac{\partial\phi(\theta)}{\partial\theta}\right| d\theta \longrightarrow \boxed{\pi(\theta) = \pi(\phi(\theta))\left|\frac{\partial\phi(\theta)}{\partial\theta}\right|}$$

**Fisher's Matrix:**

$$I_{ij}(\phi) = \int_{\Omega_X} p(x|\phi)\left(-\frac{\partial^2 \log p(x|\phi)}{\partial\phi_i\partial\phi_j}\right)dx$$

**…obviously if exists:**

**1)** $supp_x\{p(x|\phi)\}$ does not depend on $\phi$

**2)** $p(x|\phi) \in C_k(\phi) \quad (\ldots k \geq 2)$

**3)** well behaved so $\frac{\partial}{\partial\phi}\int_{\Omega_X}(\bullet)dx = \int_{\Omega_X}\frac{\partial(\bullet)}{\partial\phi}dx$

$$X \sim p(x|\phi)$$

$$I(\phi) = \int_{\Omega_X} p(x|\phi)\left(-\frac{\partial^2 \log p(x|\phi)}{\partial^2\phi}\right)dx \qquad (\geq 0)$$

**and under a transformation (s.v. function)**

$$\left.\begin{array}{l} \phi = \phi(\theta) \\ \int_{\Omega_X} p(x|\phi)dx = 1 \end{array}\right\} \quad I(\theta) = I(\phi)\left(\frac{\partial\phi}{\partial\theta}\right)^2$$

$$\boxed{\pi_J(\theta) = [I(\theta)]^{1/2}} = \left[E_X\left(-\frac{\partial^2 \log p(x|\theta)}{\partial^2\theta}\right)\right]^{1/2}$$

**(Not unique: other expressions are also invariant against reparametrizations)**

# *Example: (back to efficiency, acceptance,…)*

$$X \sim Bi(k \mid n, \theta)$$

$$p(X = k \mid n, \theta) = \binom{n}{k} \theta^{k} (1-\theta)^{n-k}$$

$$I(\theta) = E_X\left[ -\frac{\partial^2 \log p(x \mid \theta)}{\partial^2 \theta} \right] = \frac{n}{\theta(1-\theta)}$$

$$\pi(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$$
$$\sim Be(\theta \mid 1/2, 1/2)$$

***proper***

➡

$$p(\theta \mid n, k) \propto p(k \mid n, \theta)\pi(\theta) \propto \theta^{k-1/2}(1-\theta)^{n-k-1/2}$$

$$\theta \sim Be(k + 1/2, n-k+1/2)$$

**By the way…** $I(\theta) = I(\phi)\left(\frac{\partial \phi}{\partial \theta}\right)^2$ **… "Normalization"**

$$e(n) \xrightarrow{iid} \boldsymbol{x} = \{x_1, x_2, \ldots, x_n\} \qquad p(\boldsymbol{x}\,|\,\theta) = \prod_{i=1}^{n} p(x_i\,|\,\theta) \qquad w(\theta\,|\,\bullet) = \log\, p(\boldsymbol{x}\,|\,\theta) = \sum_i \log\, p(x_i\,|\,\theta)$$

**Taylor expansion around max likelihood $\theta_m$:**

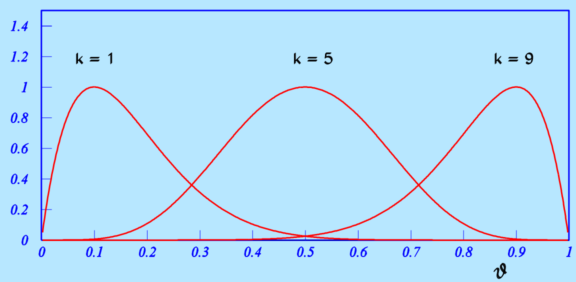
$$w(\theta\,|\,\boldsymbol{x}) = w(\theta_m\,|\,\boldsymbol{x}) - \frac{n}{2}\underbrace{\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2(-\log p(\boldsymbol{x}\,|\,\theta))}{\partial^2\theta}\right]_{\theta_m}}_{n \gg \xrightarrow{LLN} I(\theta_m)} (\theta-\theta_m)^2 + \ldots \quad \Rightarrow \quad p(\theta\,|\,\boldsymbol{x}) \sim N\left(\theta\,|\,\theta_m, \sigma^{-2} = nI(\theta_m)\right)$$

**Transformation** $\phi = \phi(\theta)$ **such that** $\Rightarrow$ $p(\phi\,|\,\boldsymbol{x}) \sim N\left(\phi\,|\,\phi_0, \sigma = const\right)$

$$I^{1/2}(\phi) = I^{1/2}(\theta)\left(\frac{d\theta}{d\phi}\right) = const \qquad I^{1/2}(\theta)d\theta \propto d\phi$$

**"Data translated likelihood"**

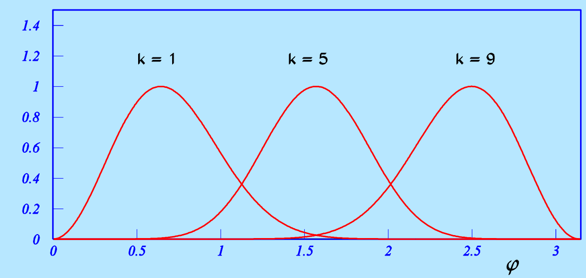
**(G.E.P. Box, G.C. Tiao)**




**EX: Binomial Distribution**

$$\phi = \int I^{1/2}(\theta)d\theta = \int \theta^{-1/2}(1-\theta)^{-1/2}\,d\theta =$$

$$= 2\;atan\sqrt{\theta/_{1-\theta}}$$



**Example:**

$$X \sim Po(k \mid \mu) \qquad P(X=n \mid \mu) = \frac{e^{-\mu}\mu^{n}}{\Gamma(n+1)} \qquad \mu > 0$$

$$I(\mu) = E_X\left[-\frac{\partial^2 \log p(x \mid \mu)}{\partial^2 \mu}\right] = \frac{1}{\mu} \quad \rightarrow \quad \pi(\mu) \propto \mu^{-1/2}$$ ***(improper)***

***Posterior density*** $p(\mu \mid n) = Ga(1, n + 1/2)$ ***(proper)***

---

**Example:**

$$X \sim Pa(x \mid \alpha, x_0) \qquad p(x \mid \alpha, x_0) = \frac{\alpha}{x_0}\left(\frac{x_0}{x}\right)^{\alpha+1} \mathbf{1}_{(x_0,\infty)}(x)$$

$\alpha > 0$

$x_0$ *known*

$$I(\alpha) = E_X\left[-\frac{\partial^2 \log p(x \mid \alpha, x_0)}{\partial^2 \alpha}\right] = \frac{1}{\alpha^2} \quad \rightarrow \quad \pi(\alpha) \propto \alpha^{-1}$$ ***(improper)***

***(*** $t = \ln(x) \rightarrow \alpha$ ***scale parameter)***

***Posterior density*** $p(\alpha \mid x, x_0) = \ln(x) x^{-\alpha}$ ***(proper)***

# *Jeffrey's in n dimensions:*

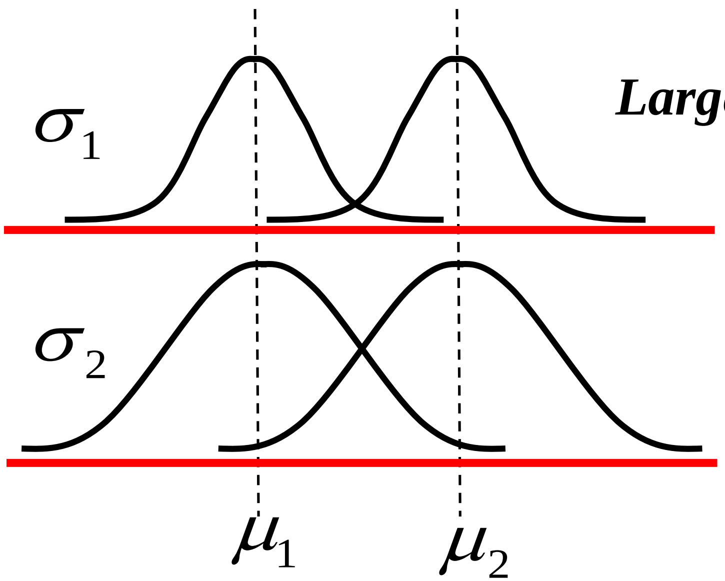
## *Fisher's Matrix:*

$$I_{ij}(\boldsymbol{\theta}) = E_X\left[\frac{\partial \ln p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \theta_i}\frac{\partial \ln p(\boldsymbol{x}|\boldsymbol{\theta})}{\partial \theta_j}\right]$$

***smooth one-to-one transformation*** $\boldsymbol{\varphi} = \boldsymbol{\varphi}(\boldsymbol{\theta}) \longrightarrow I_{ij}(\boldsymbol{\varphi}) = \frac{\partial \theta_k}{\partial \varphi_i}\frac{\partial \theta_l}{\partial \varphi_j} I_{kl}(\boldsymbol{\theta})$

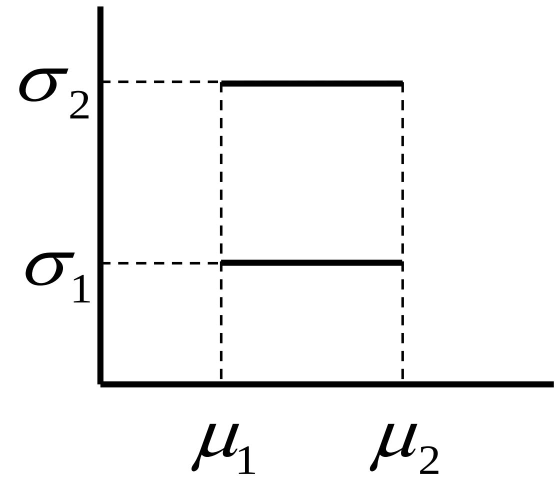
***… like a covariant tensor***

## *Connection: Geometry, Information and Probability*




***Largest difference***

***…and same Euclidean distance***




$\boldsymbol{\theta} \in \Omega \subseteq \Re^n$ ***(If..) The parameter space of a distribution Riemannian Manifold***

$I_{ij}(\boldsymbol{\theta}) = g_{ij}(\boldsymbol{\theta})$ ***metric tensor*** $\quad (ds)^2 = g_{ij}(\theta)\, d\theta^i d\theta^j$

$$g^{ik}(\theta)\, g_{kj}(\theta) = \delta^i_j$$

*Let* $\boldsymbol{\varphi} = \boldsymbol{\varphi}(\boldsymbol{\theta})$ *be a smooth one-to-one transformation such that at* $\boldsymbol{\varphi}_0 = \boldsymbol{\varphi}(\boldsymbol{\theta}_0)$

*the Fisher's matrix becomes* $[\boldsymbol{F}_{ij}(\boldsymbol{\varphi}_0)] = \boldsymbol{I}$ *(identity matrix)*

*(location)*

*In a neighborhood of* $\boldsymbol{\varphi}_0$ *, the geometry is Euclidean* → $\pi(\varphi)d\varphi \propto d\varphi$

$$d\boldsymbol{\varphi} = \left[\det\left|\frac{\partial\theta_i}{\partial\varphi_j}\right|\right]^{-1} d\boldsymbol{\theta} = \left[\det[\boldsymbol{F}(\boldsymbol{\theta})]\right]^{1/2} d\boldsymbol{\theta} = \pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$\boldsymbol{F}_{ij}(\boldsymbol{\varphi}) = \frac{\partial\theta_k}{\partial\varphi_i}\frac{\partial\theta_l}{\partial\varphi_j}\boldsymbol{F}_{kl}(\boldsymbol{\theta})$$

$$\pi(\boldsymbol{\theta}) \propto \left[\det[\boldsymbol{F}(\boldsymbol{\theta})]\right]^{1/2}$$

$$\pi_J(\theta) = \left[I(\theta)\right]^{1/2}$$

**EXAMPLE:**

**Gamma Distribution:** $p(x \mid a,b) = e^{-ax}x^{b-1}\frac{a^b}{\Gamma(b)}\boldsymbol{1}_{[0,\infty)}(x) \qquad a,b \in R^+$

*1) Fisher's matrix elements are:* $\boldsymbol{F}_{aa} = \frac{b}{a^2} \quad \boldsymbol{F}_{ab} = -\frac{1}{a} \quad \boldsymbol{F}_{bb} = \Psi'(b)$

$$\det[\boldsymbol{F}] = (b\Psi'(b) - 1)a^{-2} \qquad \pi_J(a,b) \propto a^{-1}\sqrt{b\Psi'(b) - 1}$$

## But $\pi(\boldsymbol{\theta}) \propto [\det[\boldsymbol{F}(\boldsymbol{\theta})]]^{-1/2}$ *may not be the best choice…*

*(as pointed out by Jeffreys)*

## Example: Normal distribution

$$p(x|a,b) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$F_{\mu\mu} = \sigma^{-2} \quad ; \quad F_{\sigma\sigma} = 2\sigma^{-2} \quad ; \quad F_{\mu\sigma} = 0$$

$\sigma$ **known** $\rightarrow$ $\{\mu\}$ $\qquad \pi(\mu) \propto F_{\mu\mu}^{1/2} = c$ **OK** ***(Jeffreys' / location)***

$\mu$ **known** $\rightarrow$ $\{\sigma\}$ $\qquad \pi(\sigma) \propto F_{\sigma\sigma}^{1/2} \propto \sigma^{-1}$ **OK** ***(Jeffreys' / scale)***

$\mu, \sigma$ **unknown** $\rightarrow$ $\{\mu, \sigma\}$ $\qquad F_{\mu\sigma} = 0$

➡ $\pi(\boldsymbol{\theta}) \propto [\det[\boldsymbol{F}(\boldsymbol{\theta})]]^{-1/2}$ $\quad |F| = 2\sigma^{-4} \rightarrow \pi(\mu,\sigma) \propto \sigma^{-2}$ **(L-Haar)**

➡ ***assumed independent*** $\quad \pi(\mu,\sigma) = \pi(\mu)\pi(\sigma) \propto \sigma^{-1}$ **(R-Haar)**

$$\pi(\mu,\sigma) \propto \sigma^{-a} \qquad\qquad E[ns^2\sigma^{-2}] = n-1$$

$a=1 \qquad T = \sqrt{n-1}\left(\frac{\mu-\bar{x}}{s}\right) \sim St(t|n-1) \qquad Z = n\left(\frac{s^2}{\sigma^2}\right) \sim \chi^2(z|n-1)$ **OK** $\quad E[Z] = n-1$

$a=2 \qquad T = \sqrt{n}\left(\frac{\mu-\bar{x}}{s}\right) \sim St(t|n) \qquad Z = n\left(\frac{s^2}{\sigma^2}\right) \sim \chi^2(z|n) \qquad E[Z] = n$

# PRACTICAL SUMMARY
# From Invariance arguments:

***For ONE parameter Models:*** $\pi_J(\theta) = [I(\theta)]^{1/2} = \left[ E_X\left( -\frac{\partial^2 \log p(x|\theta)}{\partial^2 \theta} \right) \right]^{1/2}$

***(if exists)***

***LOCATION parameters:*** $\pi(\mu) \propto c\mathbf{1}_{\mathrm{M}}(\mu)$

***SCALE parameters:*** $\pi(\theta) \propto 1/\theta \, \mathbf{1}_{\Theta}(\theta)$

***If $\exists$ Group of Transformations:*** $\pi_{RH}(\theta)$ $\pi_{LH}(\theta)$

***Priors from***

***CONJUGATED FAMILY***

# Conjugated Priors

*Make life easier … Simplify calculations (?)*

*Let $\mathcal{S}$ be a class of sampling distributions* $p(x \mid \theta)$

*Let $\mathcal{P}$ be a class of prior distributions for parameter* $\theta$ $p(\theta)$

***The class $\mathcal{P}$ is conjugated for $\mathcal{S}$ if***

***for all*** $p(x \mid \theta) \in \mathcal{S}$ ***and*** $p(\theta) \in \mathcal{P}$ ➡ $p(\theta \mid x) \in \mathcal{P}$

***Natural Conjugated:***

***If $\mathcal{P}$ is the class of prior distributions with same functional form as $\mathcal{S}$***

⟶ ***Conjugated Prior Distributions are "closed under sampling"***

---

***Likelihood functions*** $p(x \mid \theta)$ ***for which there is a conjugated family of priors are those which belong to the exponential family***

$$p(x_1, x_2, \ldots, x_n \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i) g^n(\boldsymbol{\theta}) \exp\left\{ \sum_{j=1}^{k} c_j \phi_j(\boldsymbol{\theta}) \sum_{i=1}^{n} h_j(x_i) \right\}$$

$$\boxed{p(\boldsymbol{\theta} \mid \tau_0, \tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{K(\boldsymbol{\tau})} g^{\tau_0}(\boldsymbol{\theta}) \exp\left\{ \sum_{j=1}^{k} c_j \phi_j(\boldsymbol{\theta}) \tau_j \right\}}$$

$$K(\boldsymbol{\tau}) = \int_{\Theta} g^{\tau_0}(\boldsymbol{\theta}) \exp\left\{ \sum_{j=1}^{k} c_j \phi_j(\boldsymbol{\theta}) \tau_j \right\} d\boldsymbol{\theta} < \infty$$

**General Scheme:** **Model** $M = \left\{ p(\boldsymbol{x} \mid \boldsymbol{\theta}) \,;\, \boldsymbol{x} \in X \,;\, \boldsymbol{\theta} \in \Theta \right\}$

1) **Choose a class of priors** $\pi(\boldsymbol{\theta} \mid \boldsymbol{\varphi})$ **that reflect the structure of the model**

2) **Choose a prior distribution for the "hyperparameters"** $\pi(\boldsymbol{\varphi})$

3) **Posterior density** $p(\boldsymbol{\theta}, \boldsymbol{\varphi} \mid \boldsymbol{x}) \propto p(\boldsymbol{x} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \boldsymbol{\varphi}) \pi(\boldsymbol{\varphi})$

**and marginalise for parameters of interest:** $p(\boldsymbol{\theta} \mid \boldsymbol{x}) \propto \int_{\Phi} p(\boldsymbol{x} \mid \boldsymbol{\theta}) \pi(\boldsymbol{\theta} \mid \boldsymbol{\varphi}) \pi(\boldsymbol{\varphi}) d\boldsymbol{\varphi}$

**… 2) How do we choose the "hyperprior"** $\pi(\boldsymbol{\varphi})$ **?**

***Option 1):***

***From marginal density:*** $$p(\boldsymbol{\varphi}, \boldsymbol{x}) = \pi(\boldsymbol{\varphi})\int_{\Theta} p(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \boldsymbol{\varphi})d\boldsymbol{\theta} = \pi(\boldsymbol{\varphi})\, p(\boldsymbol{x} \mid \boldsymbol{\varphi})$$

$$p(\boldsymbol{x} \mid \boldsymbol{\varphi}) = \int_{\Theta} p(\boldsymbol{x} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \boldsymbol{\varphi})d\boldsymbol{\theta}$$ ⟶ ***reference prior*** $\pi(\boldsymbol{\varphi})$ ***for model*** $p(\boldsymbol{x} \mid \boldsymbol{\varphi})$

⟶ $$\pi(\boldsymbol{\theta}) = \int_{\Phi} \pi(\boldsymbol{\theta} \mid \boldsymbol{\varphi})\pi(\boldsymbol{\varphi})d\boldsymbol{\varphi}$$

---

***Option 2): "Empirical method": assign numeric values to hyperparameters***

***from*** $p(\boldsymbol{x} \mid \boldsymbol{\varphi})$ ***(for instance MaxLik)*** $< \delta_{\varphi^*}, p(\boldsymbol{\theta}, \boldsymbol{\varphi} \mid \boldsymbol{x}) >$

***No variability (uncertainty) of hyperparameters but may help to guess***

***Option 3): … Consider "reasonable hyperpriors"***

***(proper posteriors; may look at Option 2 for indication)***

**EXAMPLE:** $X \sim Po(k \mid \mu)$

***express in the form:***

$$p(n \mid \mu) = \frac{e^{-\mu}\mu^{n}}{\Gamma(n+1)} = \frac{e^{-(\mu - n\ln\mu)}}{\Gamma(n+1)}$$

$$p(x_1, x_2, \ldots, x_n \mid \boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i) g^{n}(\boldsymbol{\theta}) \exp\left\{ \sum_{j=1}^{k} c_j \phi_j(\boldsymbol{\theta}) \sum_{i=1}^{n} h_j(x_i) \right\}$$

$$f(n) = \frac{1}{\Gamma(n+1)} \qquad g(\mu) = 1 \qquad k = \{1,2\}$$

$$k = 1 \quad c_1 = -1 \quad \phi_1(\mu) = \mu \quad h_1(n) = 1$$

$$k = 2 \quad c_2 = 1 \quad \phi_2(\mu) = \ln\mu \quad h_2(n) = n$$

$$p(\boldsymbol{\theta} \mid \tau_0, \tau_1, \tau_2, \ldots, \tau_k) = \frac{1}{K(\boldsymbol{\tau})} g^{\tau_0}(\boldsymbol{\theta}) \exp\left\{ \sum_{j=1}^{k} c_j \phi_j(\boldsymbol{\theta}) \tau_j \right\} \longrightarrow p(\mu \mid \tau_1, \tau_2) = \frac{\tau_1^{\tau_2}}{\Gamma(\tau_2)} e^{-\tau_1 \mu} \mu^{\tau_2 - 1}$$

$$p(\mu, \tau_1, \tau_2 \mid n) \propto \frac{\tau_1^{\tau_2}}{\Gamma(\tau_2)} e^{-(\tau_1+1)\mu} \mu^{n+\tau_2-1} \pi(\tau_1, \tau_2)$$

$$p(\mu \mid n) \propto \int_{\mathrm{T}} p(\mu, \tau_1, \tau_2 \mid n) d\boldsymbol{\tau}$$

$$p(\tau_1, \tau_2 \mid n) = \int_{0}^{\infty} p(\mu, \tau_1, \tau_2 \mid n) d\mu \propto \underbrace{\left[ \frac{\Gamma(n+\tau_2)}{\Gamma(\tau_2)} \frac{\tau_1^{\tau_2}}{(1+\tau_1)^{n+\tau_2}} \right]}_{p(n \mid \tau_1, \tau_2)} \pi(\tau_1, \tau_2) \propto p(n \mid \tau_1, \tau_2) \pi(\tau_1, \tau_2)$$

# Example: Dirichlet and Generalised Dirichlet

**Conjugated priors for Multinomial:**

$$p(\mathbf{d}\,|\,\boldsymbol{\theta}) = C\prod_{i=1}^{no}\theta_i^{n_i}$$

$$\boldsymbol{\Theta} = \{\theta_1,\ldots,\theta_n\} \sim Di(\boldsymbol{\theta}\,|\,\boldsymbol{\alpha})$$

$$p(\boldsymbol{\theta}\,|\,\boldsymbol{\alpha}) = D(\boldsymbol{\alpha})\,\theta_1^{\alpha_1-1}\theta_2^{\alpha_2-1}\cdots\theta_n^{\alpha_n-1}$$

$$\alpha_k > 0 \qquad D(\boldsymbol{\alpha}) = \Gamma(\alpha_0)\left[\prod_{k=1}^{n}\Gamma(\alpha_k)\right]^{-1} \qquad \alpha_0 = \sum_{k=1}^{n}\alpha_k$$

$$\theta_k \in [0,1] \qquad \sum_{k=1}^{n}\theta_k = 1 \;\longrightarrow\; \theta_n = 1 - \sum_{k=1}^{n-1}\theta_k$$

***Degenerated Distribution:***

$$p(\boldsymbol{\theta}\,|\,\boldsymbol{\alpha}) = D(\boldsymbol{\alpha})\left(\prod_{i=1}^{n-1}\theta_i^{\alpha_i-1}\right)\left(1-\sum_{k=1}^{n-1}\theta_k\right)^{\alpha_n-1}$$

$$E[\Theta_i] = \frac{\alpha_i}{\alpha_0} \qquad V[\Theta_i,\Theta_j] = \frac{\alpha_i\alpha_0\delta_{ij} - \alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}$$

***Control on E or V; not both:*** $\longrightarrow$ ***Generalised Dirichlet***

# *EXAMPLE: UNFOLDING a simple distribution*

***Interest in distribution of*** $X$

***DATA:*** $Y$ ***observed quantity in categories*** $data = \{n_1, n_2, \ldots, n_{no}\}$ $n_E = \sum_{i=1}^{no} n_i$

***MC simulation of detector response and selection***

$$p_{obs}(y) = \int_X R(y \mid x, \zeta)\, p_G(x)\, dx \qquad p(y, x) = R(y \mid x, \zeta)\, p_G(x)$$

***"Migration" (unfolding) Matrix***

$$P(i \mid j) = \frac{\int_{\Lambda_j} p_G(x)dx \int_{\Omega_i} R(y \mid x, \zeta)dy}{\int_{\Lambda_j} p_G(x)dx}$$

***MODEL:***

$$Po: \quad p(\mathbf{d} \mid \boldsymbol{\varphi}) = C \prod_{i=1}^{no} \phi_i^{\,n_i} e^{-n_E \phi_i} \qquad \boldsymbol{\varphi} = \{\phi_1, \phi_2, \ldots, \phi_{no}\}$$

$$Mn: \quad p(\mathbf{d} \mid \boldsymbol{\varphi}) = C \prod_{i=1}^{no} \phi_i^{\,n_i}$$

***Interest in:*** $\boldsymbol{\theta} = \{\theta_1, \theta_2, \ldots, \theta_{nt}\}$ $\phi_i = \sum_{j=1}^{nt} P(i \mid j)\, \theta_j$

$n_t < n_o$

$$\boldsymbol{\Theta} = \{\theta_1, \ldots, \theta_n\} \sim GDi(\boldsymbol{\theta} \mid \boldsymbol{\alpha}, \boldsymbol{\beta})$$

$$p(\boldsymbol{\theta} \mid \boldsymbol{\alpha}) = \prod_{i=1}^{n-1} \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} \theta_i^{\alpha_i - 1} \left(1 - \sum_{k=1}^{i} \theta_k\right)^{\gamma_i}$$

$$0 < \theta_i < 1 \qquad \sum_{k=1}^{n-1} \theta_k < 1 \qquad \theta_n = 1 - \sum_{k=1}^{n-1} \theta_k$$

$$\alpha_k > 0 \qquad \beta_k > 0 \qquad \gamma_i = \begin{cases} \beta_i - \alpha_{i+1} - \beta_{i+1}\ ; & i = 1, 2, \ldots, n-2 \\ \beta_{n-1} - 1; & i = n-1 \end{cases}$$

$$E[\Theta_i] = \frac{\alpha_i}{\alpha_i + \beta_i} S_i$$

$$V[\Theta_i] = E[\Theta_i]\left(\frac{\alpha_i + \delta_{ij}}{\alpha_i + \beta_i + 1} T_i - E[\Theta_i]\right)$$

$$S_i = \prod_{k=1}^{i-1} \beta_k (\alpha_k + \beta_k)^{-1}$$

$$T_i = \prod_{k=1}^{i-1} (\beta_k + 1)(\alpha_k + \beta_k + 1)^{-1}$$

***PRIOR:*** $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta} \mid \boldsymbol{\alpha}) = Di(\boldsymbol{\theta} \mid \boldsymbol{\alpha})$ $\quad ...GDi(\boldsymbol{\theta} \mid \boldsymbol{\alpha})$

***POSTERIOR:***

$$p(\boldsymbol{\theta} \mid \mathbf{d}, \boldsymbol{\alpha}) \propto p(\mathbf{d} \mid \boldsymbol{\varphi}(\boldsymbol{\theta}))\pi(\boldsymbol{\theta} \mid \boldsymbol{\alpha})$$

$$p(\boldsymbol{\theta} \mid \mathbf{d}, \boldsymbol{\alpha}) \propto p(\mathbf{d} \mid \boldsymbol{\varphi}(\boldsymbol{\theta}))\pi(\boldsymbol{\theta} \mid \boldsymbol{\alpha})\pi(\boldsymbol{\alpha})$$

***(Still more work…***
***See later "Priors with Partial Information")***

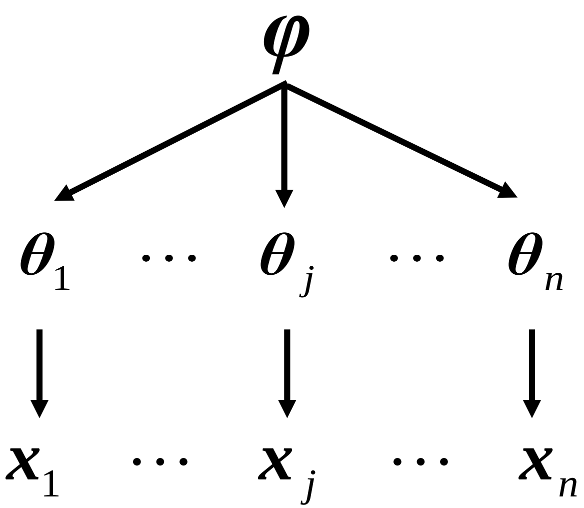
***The structure*** $\pi(\theta,\varphi)=\pi(\theta\mid\varphi)\pi(\varphi)$

***…hierarchical structures and "hyperpriors"***

# *Priors for*
# *HIERARCHICAL STRUCTURES*

# Hierarchical Structures

**Parameters $\{\theta_1 \quad \theta_2 \quad \cdots \quad \theta_n\}$ are drawn from a common "super-population" governed by "hyperparameters" $\boldsymbol{\varphi}$**




$p(\boldsymbol{\varphi})$

$p(\boldsymbol{\theta} \mid \boldsymbol{\varphi})$

$p(\boldsymbol{x} \mid \boldsymbol{\theta})$

$$M_{\Theta} = \{p(\boldsymbol{\theta} \mid \boldsymbol{\varphi})\,;\, \boldsymbol{\theta} \in \Theta\,;\, \boldsymbol{\varphi} \in \boldsymbol{\Phi}\}$$

- **considers variability of $\theta_i$**
- **usually, conjugate models simplify life**

$$M_X = \{p(\boldsymbol{x} \mid \boldsymbol{\theta})\,;\, \boldsymbol{x} \in X\,;\, \boldsymbol{\theta} \in \Theta\}$$

$$p(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\varphi}) = p(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{\varphi})\, p(\boldsymbol{\theta}, \boldsymbol{\varphi}) = \\ = p(\boldsymbol{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \boldsymbol{\varphi})\, p(\boldsymbol{\varphi})$$

**Under exchangeability**

$$p(\theta_1, \ldots, \theta_n \mid \boldsymbol{\varphi}) = \prod_{i=1}^{n} p(\theta_i \mid \boldsymbol{\varphi})$$

$$p(x_1, \ldots, x_n \mid \theta_1, \ldots \theta_n) = \prod_{i=1}^{n} p(x_i \mid \theta_i)$$

**Posterior:**

$$p(\boldsymbol{\theta}, \boldsymbol{\varphi} \mid \boldsymbol{x}) \propto p(\boldsymbol{x} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid \boldsymbol{\varphi})\, p(\boldsymbol{\varphi})$$

$$= p(\boldsymbol{\varphi}) \prod_{i=1}^{n} p(x_i \mid \theta_i)\, p(\theta_i \mid \boldsymbol{\varphi})$$

**Parameterisation** $\{\theta, \varphi\}$ … **usually, parameters of interest** $\{\theta\}$

**1) What data tell us about** $\varphi$ **?**

$$p(\mathbf{x},\varphi) = p(\varphi \mid \mathbf{x})\,p(\mathbf{x}) = \int_{\Theta} p(\mathbf{x},\theta,\varphi)d\theta = \int_{\Theta} p(\mathbf{x}\mid\theta)\,p(\theta\mid\varphi)\,p(\varphi)d\theta$$

$$p(\varphi \mid \mathbf{x}) \propto p(\mathbf{x}\mid\varphi)\,p(\varphi) = p(\varphi)\underbrace{\int_{\Theta} p(\mathbf{x}\mid\theta)\,p(\theta\mid\varphi)d\theta}_{p(\mathbf{x}\mid\varphi)}$$

$M_X$ —— $M_\Theta$

*(… make sure it is proper…)*

**2) Inferences on** $\theta$

$$p(\theta \mid \mathbf{x},\varphi) = \frac{p(\theta,\mathbf{x},\varphi)}{p(\mathbf{x},\varphi)} = \frac{p(\mathbf{x}\mid\theta)\,p(\theta\mid\varphi)}{p(\mathbf{x}\mid\varphi)}$$

$$p(\theta \mid \mathbf{x},\varphi) \propto p(\mathbf{x}\mid\theta)\,p(\theta\mid\varphi)$$

**… nicely suited for MC sampling:**

**1) Draw** $\varphi$ **from** $p(\varphi \mid \mathbf{x})$

**2) Draw** $\theta$ **from** $p(\theta \mid \mathbf{x},\varphi)$

*(usually can be drawn in dependently)*

$$p(\theta \mid \mathbf{x},\varphi) = \prod_i p(\theta_i \mid x_i,\varphi)$$

## 3) Interest in Marginal… $\theta$

**3.1)** $$p(\boldsymbol{\theta},\mathbf{x}) = p(\boldsymbol{\theta}\mid\mathbf{x})p(\mathbf{x}) = \int_{\Phi} p(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi} = p(\mathbf{x}\mid\boldsymbol{\theta})\int_{\Phi} p(\boldsymbol{\theta}\mid\boldsymbol{\varphi})p(\boldsymbol{\varphi})d\boldsymbol{\varphi}$$

$$\boxed{p(\boldsymbol{\theta}\mid\mathbf{x}) \propto p(\mathbf{x}\mid\boldsymbol{\theta})\int_{\Phi} p(\boldsymbol{\theta}\mid\boldsymbol{\varphi})p(\boldsymbol{\varphi})d\boldsymbol{\varphi}}$$

$M_X$ $M_\Theta$

**3.2)** $$p(\boldsymbol{\theta},\mathbf{x}) = p(\boldsymbol{\theta}\mid\mathbf{x})p(\mathbf{x}) = \int_{\Phi} p(\mathbf{x},\boldsymbol{\theta},\boldsymbol{\varphi})d\boldsymbol{\varphi} = \int_{\Phi} p(\boldsymbol{\theta}\mid\mathbf{x},\boldsymbol{\varphi})p(\boldsymbol{\varphi}\mid\mathbf{x})p(\mathbf{x})d\boldsymbol{\varphi} =$$

$$= \int_{\Phi} \frac{p(\boldsymbol{\theta},\mathbf{x},\boldsymbol{\varphi})}{p(\mathbf{x},\boldsymbol{\varphi})} p(\boldsymbol{\varphi}\mid\mathbf{x})p(\mathbf{x})d\boldsymbol{\varphi} = \int_{\Phi} \frac{p(\mathbf{x}\mid\boldsymbol{\theta})p(\boldsymbol{\theta}\mid\boldsymbol{\varphi})}{p(\mathbf{x}\mid\boldsymbol{\varphi})} p(\boldsymbol{\varphi}\mid\mathbf{x})p(\mathbf{x})d\boldsymbol{\varphi}$$

$$\boxed{p(\boldsymbol{\theta}\mid\mathbf{x}) \propto \int_{\Phi} p(\mathbf{x}\mid\boldsymbol{\theta})p(\boldsymbol{\theta}\mid\boldsymbol{\varphi})\frac{p(\boldsymbol{\varphi}\mid\mathbf{x})}{p(\mathbf{x}\mid\boldsymbol{\varphi})}d\boldsymbol{\varphi}}$$

***Usually draw inferences on each*** $\theta_i$

$$p(\theta_i\mid\mathbf{x}) \propto \int_{\Phi} \frac{p(\mathbf{x}\mid\theta_i)p(\theta_i\mid\boldsymbol{\varphi})}{p(\mathbf{x}\mid\boldsymbol{\varphi})} p(\boldsymbol{\varphi}\mid\mathbf{x})d\boldsymbol{\varphi}$$

## 4) Predictive distribution

$$p(\boldsymbol{\theta},\mathbf{x},\mathbf{x}_{new}) = p(\mathbf{x}_{new}\mid\boldsymbol{\theta},\mathbf{x})p(\boldsymbol{\theta},\mathbf{x}) = p(\mathbf{x}_{new}\mid\boldsymbol{\theta})p(\boldsymbol{\theta}\mid\mathbf{x})p(\mathbf{x})$$

$$\boxed{p(\mathbf{x}_{new}\mid\mathbf{x}) = \int_{\Theta} p(\mathbf{x}_{new}\mid\boldsymbol{\theta})p(\boldsymbol{\theta}\mid\mathbf{x})d\boldsymbol{\theta}}$$

***(independent samplings correlated through model+ posterior)***

# *Example: Comparing sample means*

***We have see the case for two means:***

***differences between two means:*** ***Student's t-Distribution***

***Differences among a group of means: Generalize Student's t-Distribution***

***Classical Approach: ANOVA (R.A. Fisher):***

***To compare means, consider variances within groups***
***and variances between groups***

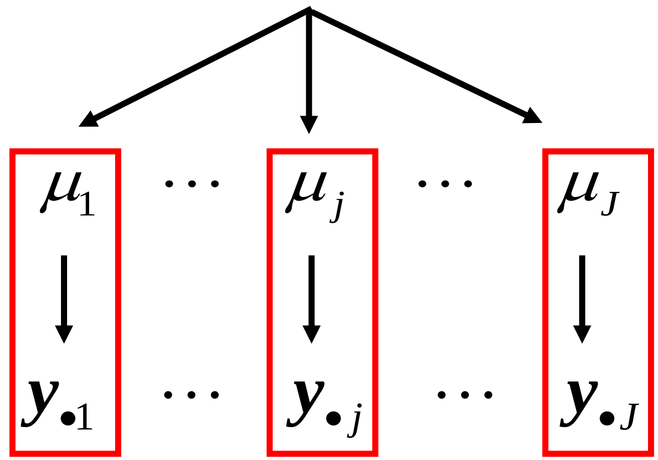
***Classically… under hypothesis of Normality***
***→ transformations to get "close" to Normal***

***Clear and general scheme as***
***Hierarchical Bayesian Structure***

# Hierarchical Structure

*(assume Normality in this example)*

$\varphi = \{\mu, \sigma, \sigma_\mu\}$ $\qquad \pi(\mu, \sigma^2, \sigma_\mu^2)$




$\mu_j \sim N(\mu, \sigma_\mu^2)$

$y_{ij} \sim N(\mu_j, \sigma^2)$

**1) J groups of observations:** $j = \quad 1 \quad \cdots \quad j \quad \cdots \quad J$ $\qquad \boldsymbol{y}_{\bullet j} = \{y_{1j}, \ldots, y_{n_j j}\}$

**2) For each group** $j = 1, \ldots, J$ **, observations** $\{y_{1j}, y_{2j}, \ldots, y_{n_j j}\}$ **are considered as an exchangeable sequence (within groups; conditional on** $\{\mu_j, \sigma\}$ **) drawn from the model** $\qquad y_{ij} \sim N(\mu_j, \sigma^2)$

**3) For the J groups, the parameters** $\{\mu_1, \mu_2, \ldots, \mu_J\}$ **are considered as an exchangeable sequence conditional on** $\{\mu, \sigma_\mu\}$ **drawn from the model** $\qquad \mu_j \sim N(\mu, \sigma_\mu^2)$

**4) Set up prior** $\qquad \pi(\mu, \sigma^2, \sigma_\mu^2)$

$$y_{ij} \sim N(\mu_j, \sigma^2) \qquad p(\mathbf{y}_{\bullet j} \mid \mu_j, \sigma) \sim \prod_{i=1}^{n_j} \sigma^{-1} e^{-\frac{(y_{ij}-\mu_j)^2}{2\sigma^2}}$$

$$\mu_j \sim N(\mu, \sigma_\mu^2) \qquad p(\mathbf{y}_{\bullet j}, \mu_j \mid \mu, \sigma, \sigma_\mu) \sim \prod_{j=1}^{J} \sigma_\mu^{-1} e^{-\frac{(\mu_j-\mu)^2}{2\sigma_\mu^2}} \left\{ \prod_{i=1}^{n_j} \sigma^{-1} e^{-\frac{(y_{ij}-\mu_j)^2}{2\sigma^2}} \right\}$$

$$\{\mathbf{y}_{\bullet j}, \mu_j, \mu, \sigma, \sigma_\mu\} \sim \qquad p(\mathbf{y}_{\bullet j}, \mu_j, \mu, \sigma, \sigma_\mu) \sim p(\mathbf{y}_{\bullet j}, \mu_j \mid \mu, \sigma, \sigma_\mu)\pi(\mu, \sigma, \sigma_\mu)$$

***For the J groups:***

$$\mu_j \sim N\left( \frac{\sigma_\mu^2 \sum_{i=1}^{n_j} y_{ij} + \mu\sigma^2}{\sigma_\mu^2 n_j + \sigma^2}, \frac{\sigma_\mu^2 \sigma^2}{\sigma_\mu^2 n_j + \sigma^2} \right)$$

***For*** $\{\mu, \sigma, \sigma_\mu\}$***:***

$$\pi(\mu) = N(\mu_0, \sigma_0^2) \qquad \mu \sim N\left( \frac{\sigma_\mu^2 \mu_0 + \sigma_0^2 \sum_{j=1}^{J} \mu_j}{\sigma_\mu^2 + \sigma_0^2 J}, \frac{\sigma_\mu^2 \sigma_0^2}{\sigma_\mu^2 + \sigma_0^2 J} \right)$$

$$\eta = \sigma_\mu^{-2}$$
$$\pi(\eta) = Ga(c, d)$$

$$\eta = \sigma_\mu^{-2} \sim Ga\left(\frac{1}{2}\sum_{j=1}^{J}(\mu_j - \mu)^2 + c, \frac{J}{2} + b\right)$$

$$\phi = \sigma^{-2}$$
$$\pi(\phi) = Ga(a, b)$$

$$\phi = \sigma^{-2} \sim Ga\left(\frac{1}{2}\sum_{j=1}^{J}\sum_{i=1}^{n_j}(y_{ij} - \mu_j)^2 + a, \frac{1}{2}\sum_{j=1}^{J} n_j + b\right)$$

***(…Inverse Gamma Distribution… )***

***(Uniform priors for*** $\sigma$ ***and*** $\sigma_\mu$ ***correspond to*** $a = c = 0 \quad ; \quad b = d = -\frac{1}{2}$ ***)***

## Gibbs Sampling:

***(see lecture on MC sampling)***

***0) Set*** $\{\mu_0, \sigma_0, a, b, c, d\}$

***1) Draw from marginal densities***

***1):*** $\{\mu_1, \mu_2, \ldots, \mu_J\}$ $\quad \mu_j \sim N(\bullet, \bullet)$

***2):*** $\{\mu\}$ $\quad \mu \sim N(\bullet, \bullet)$

***3):*** $\{\eta\}$ $\quad \eta \sim Ga(\bullet, \bullet) \rightarrow \sigma_\mu = \eta^{-1/2}$

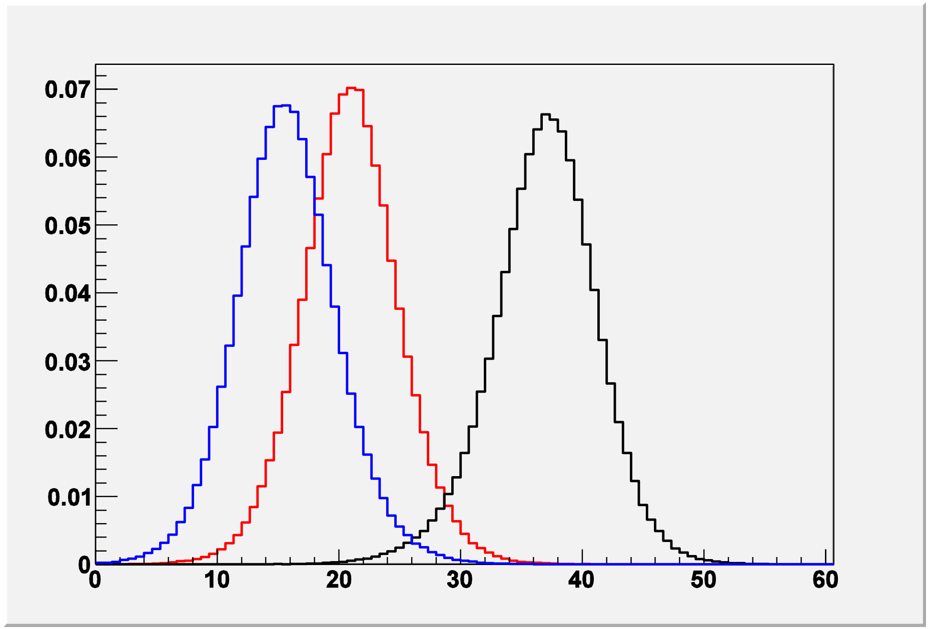
***4):*** $\{\phi\}$ $\quad \phi \sim Ga(\bullet, \bullet) \rightarrow \sigma = \phi^{-1/2}$

***Repeat step 1) until equilibrium is reached and then take 1 every 5 or 10 samples***

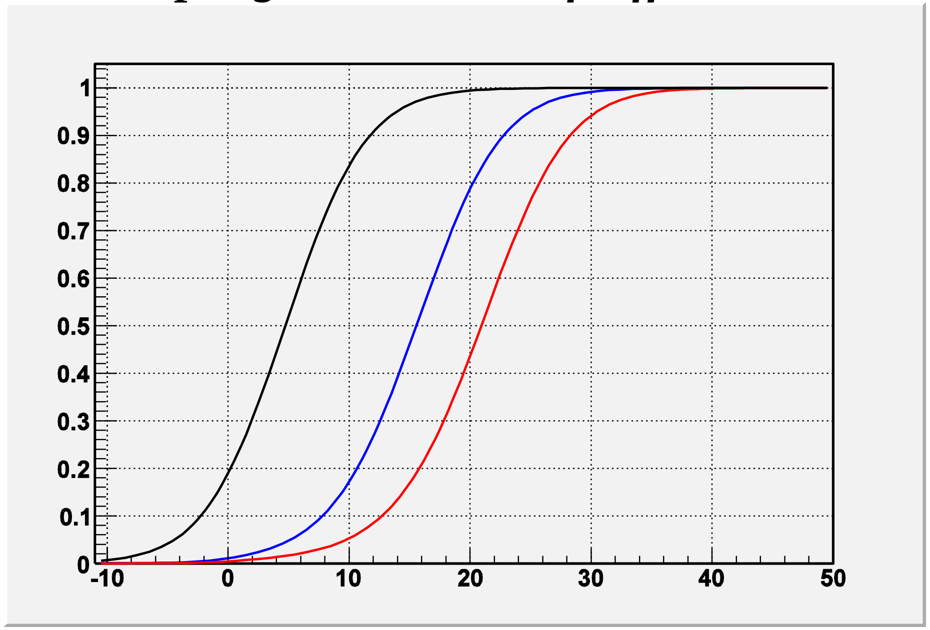
***For this example I took, without further checks, (all densities are proper)***

$$\{\mu_0, \sigma_0\}$$ ***sample mean and rms***

$$a = c = \varepsilon \quad ; \quad b = d = -(\frac{1}{2} + \varepsilon)$$




***Sampling Distribution of differences***




***Real data (now a days irrelevant and forgotten) on scaled transmittance of aerogel tiles under different pressure and temperature***

## *Predictive Distribution:*
## *SAMPLING from POSTERIOR ("Bayesian Bootstrapping")*

***As starting hypothesis, we consider a model:*** $M_X = \{p(\boldsymbol{x}\,|\,\boldsymbol{\theta})\,;\,\boldsymbol{x}\in X\,;\boldsymbol{\theta}\in\Theta\}$

***…but any model is an approximate description of nature***

***Is the model reasonable to describe the observed data?***

***From the posterior*** $p(\boldsymbol{\theta}\,|\,\boldsymbol{x})$ ***we make inferences but do not have a feeling on the adequacy of the model***

***If doubts:***

***1) Check other models (obvious)*** $M_X^{i)} = \{p_i(\boldsymbol{x}\,|\,\boldsymbol{\theta})\,;\,\boldsymbol{x}\in X\,;\boldsymbol{\theta}\in\Theta\}$

***2) Get a feeling of the adequacy of the model from predictive density***

$$p(\boldsymbol{x}^{new}\,|\,\boldsymbol{x}^{obs}) = \int_{\Theta} p(\boldsymbol{x}^{new}\,|\,\boldsymbol{\theta})\,p(\boldsymbol{\theta}\,|\,\boldsymbol{x}^{obs})d\boldsymbol{\theta}$$

***usually MC sampling under the assumption that model is adequate and test statistics (for instance order statistics)***

# *PROBABILITY MATCHING PRIORS*

*Welch B. , Pears H. (1963) J. Roy. Stat. Soc. B 25 318-329*
*Ghosh M., Mukerjee R., Biometrika 84 970-975*

# Probability Matching Priors

***Prior distribution such that one-sided credible intervals derived from posterior distribution coincide (to a certain degree of accuracy) with those derived from frequentist approach***

***Expand*** $p(\theta \mid x) \propto p(x \mid \theta)\pi(\theta)$ ***around*** $\theta_m = \max_\theta p(x \mid \theta)$ ***up to*** $O\left((\theta-\theta_m)^4\right)$

$$T = \sqrt{nI(\theta_m)}(\theta - \theta_m)$$

$$P(T \le z \mid \mathbf{x}) = P(z) - \frac{Z(z)}{\sqrt{n}}\left[\left(\frac{I^{-1/2}(\theta)}{\pi(\theta)}\frac{\partial \pi(\theta)}{\partial \theta}\right)_{\theta_m} + \frac{z^2+2}{3}\left(\frac{\partial I^{-1/2}(\theta)}{\partial \theta}\right)_{\theta_m}\right] + O\left(1/n\right)$$

$$Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$P(x) = \int_{-\infty}^{x} Z(t)\,dt$$

***"Frequentists":*** $\theta_0 = \theta_m + O\left(1/\sqrt{n}\right)$ ***and, for a sequence of priors that shrink to*** $\theta_m \approx \theta_0$

$$P_F(T \le z \mid \mathbf{x}) = P(z) - \frac{Z(z)}{\sqrt{n}}\left[\frac{z^2+1}{3}\left(\frac{\partial I^{-1/2}(\theta)}{\partial \theta}\right)_{\theta_m}\right] + O\left(1/n\right)$$

***same Probability if***

$$\left(\frac{I^{-1/2}(\theta)}{\pi(\theta)}\frac{\partial \pi(\theta)}{\partial \theta}\right)_{\theta_m} = -\left(\frac{\partial I^{-1/2}(\theta)}{\partial \theta}\right)_{\theta_m} \longrightarrow \pi(\theta) = \left[I(\theta)\right]^{1/2}$$

***…Jeffrey's prior is PMP***

**n-dimensional case:** $p(\boldsymbol{x} \mid \theta_1, \theta_2, \ldots, \theta_n)$ **…same argument…**

**Parameter of interest:** $\theta_1$ **ordered parameterization** $\{\theta_1, \theta_2, \ldots, \theta_n\}$

**1)** $\boldsymbol{S} = \boldsymbol{F}^{-1}$

**2)** $\sum_{k=1}^{p} \frac{\partial}{\partial \theta_k} \left[ \pi(\boldsymbol{\theta}) \frac{\boldsymbol{S}_{k1}}{\boldsymbol{S}_{11}^{1/2}} \right] = 0$ **any solution** $\pi(\boldsymbol{\theta})$ **will do the job**

**EXAMPLE:**

**Gamma Distribution:** $p(x \mid a, b) = e^{-ax} x^{b-1} \frac{a^b}{\Gamma(b)} \boldsymbol{1}_{[0,\infty)}(x) \qquad a, b \in R^+$

**1) Fisher's matrix elements are:** $\boldsymbol{F}_{aa} = \frac{b}{a^2} \quad \boldsymbol{F}_{ab} = -\frac{1}{a} \quad \boldsymbol{F}_{bb} = \Psi'(b)$

**2) ordering** $\{b, a\}$ $\quad \pi_{PM}(b, a) \propto a^{-1} b^{-1/2} \sqrt{b\Psi'(b) - 1}$

$\{a, b\}$ $\quad \pi_{PM}(a, b) \propto a^{-1} \sqrt{\Psi'(b)} \left( \sqrt{b\Psi'(b) - 1} \right)$

**($a$ is a scale parameter)** $\quad \pi_J(a, b) \propto a^{-1} \sqrt{b\Psi'(b) - 1}$

## PROBLEM: Correlation Coefficient of the Bivariate Normal Model

$$p(\boldsymbol{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Sigma}) \propto \det[\boldsymbol{\Sigma}]^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^t\,\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

$$\boldsymbol{x}=\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \boldsymbol{\mu}=\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \boldsymbol{\Sigma}=\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

$$e(n) \xrightarrow{iid} \boldsymbol{x}=\{(x_{11},x_{21}),(x_{12},x_{22}),\ldots,(x_{1n},x_{2n})\}$$

1) ***Show that a probability matching prior with*** $\rho$ ***the parameter of interest is given by***

$$\pi(\rho,\mu_1,\mu_2,\sigma_1,\sigma_2) \propto \sigma_1^{-1}\sigma_2^{-1}(1-\rho^2)^{-1}$$

2) ***Show that the posterior for the correlation coefficient is:***

$$\pi(\rho\,|\,\boldsymbol{x}) \propto (1-\rho^2)^{(n-3)/2}(1-r\rho)^{-(n-3/2)}\,F\left(1/2, 1/2, n-1/2, \frac{1+r\rho}{2}\right)$$

***sample correlation*** $r=\dfrac{\sum_i (x_{1i}-\bar{x}_1)(x_{1i}-\bar{x}_1)}{\left[\sum_i (x_{1i}-\bar{x}_1)^2 \sum_i (x_{2i}-\bar{x}_2)^2\right]^{1/2}}$ ***is a sufficient statistic for*** $\rho$

# *REFERENCE PRIORS*

# Reference Priors

Bernardo, J.M. (1979)
J. Roy. Stat. Soc. Ser. B 41 113-147

$$M_X = \{p(\boldsymbol{x}\,|\,\boldsymbol{\theta})\,;\,\boldsymbol{x}\in X\,;\boldsymbol{\theta}\in\Theta\}$$

$e(1) \rightarrow \{\boldsymbol{x}_1\}$ **Expected amount of info on parameter $\theta$ provided by one observation of the model $p(\boldsymbol{x}\,|\,\theta)$ relative to prior knowledge $p(\theta)$**

$$I[e(1), p(\theta)] = \int_{\Theta,X} p(\theta,\boldsymbol{x})\log\frac{p(\theta,\boldsymbol{x})}{p(\boldsymbol{x})\,p(\theta)}\,d\boldsymbol{x}d\theta = \int_{\Theta} p(\theta)d\theta\int_X p(\boldsymbol{x}\,|\,\theta)\log\frac{p(\theta\,|\,\boldsymbol{x})}{p(\theta)}\,d\boldsymbol{x}$$

---

***Expected Mutual Information:***

***Given two continuous random quantities X and Y with density $p(x,y)$, the info we expect to get on X if we observe Y is***

$$I(X:Y) = \int_X dx\int_Y dy\; p(x,y)\ln\left(\frac{p(x,y)}{p(x)\,p(y)}\right) = \int_Y p(y)dy\left\{\int_X dx\; p(x\,|\,y)\ln\left(\frac{p(x\,|\,y)}{p(x)}\right)\right\}$$

***Kullback-Leibler Discrepancy between two distributions*** $\quad D_{KL}(g|f) = \int_X f(x)\ln\left(\frac{f(x)}{g(x)}\right)dx$

***A sequence $\{p_i(x)\}_{i=1}^{\infty}$ of pdf's converges logarithmically to a pdf $p(x)$ iff:*** $\quad \lim_{i\rightarrow\infty} D_{KL}(p\,|\,p_i) = 0$

**For k independent observations** $\quad \boldsymbol{X} \xrightarrow{iid} e(k) \rightarrow \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$

***Expected amount of info* on parameter** $\theta$ **provided by *k independent observations* of the model** $p(\boldsymbol{x}|\theta)$ ***relative to prior knowledge*** $p(\theta)$

$$\boldsymbol{z}_k = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$$

$$d\boldsymbol{z}_k = d\boldsymbol{x}_1 d\boldsymbol{x}_2 \cdots d\boldsymbol{x}_k$$

$$\int_X p(\boldsymbol{z}_k|\theta) d\boldsymbol{z}_k = 1$$

$$I[e(k), p(\theta)] = \int_\Theta p(\theta) d\theta \int_X p(\boldsymbol{z}_k|\theta) \log \frac{p(\theta|\boldsymbol{z}_k)}{p(\theta)} d\boldsymbol{z}_k =$$

$$= \int_\Theta p(\theta) d\theta \left\{ \int_X p(\boldsymbol{z}_k|\theta) \log p(\theta|\boldsymbol{z}_k) d\boldsymbol{z}_k - \log p(\theta) \right\} = \int_\Theta p(\theta) \log \frac{f_k(\theta)}{p(\theta)} d\theta$$

$$I[e(k), p(\theta)] = \int_\Theta p(\theta) \log \frac{f_k(\theta)}{p(\theta)} d\theta$$

$$f_k(\theta) = \exp\left\{ \int_X p(\boldsymbol{z}_k|\theta) \log p(\theta|\boldsymbol{z}_k) d\boldsymbol{z}_k \right\}$$

***"Perfect info"*** $\quad I[e(\infty), p(\theta)] = \lim_{k\to\infty} I[e(k), p(\theta)] \quad (if\ \exists)$

**measures *maximum info on*** $\theta$ ***that could be obtained from model*** $p(\boldsymbol{x}|\theta)$
***relative to prior knowledge*** $p(\theta)$

## Central Idea:

**Reference prior relative to model** $M_X = \{ p(\mathbf{x}|\boldsymbol{\theta}) ; \mathbf{x} \in X ; \boldsymbol{\theta} \in \Theta \}$

… *that for which "perfect info" is maximal*

… *"less informative" for this model*

**Calculus of Variations** $p(\theta) = p^*(\theta) + \varepsilon \eta(\theta)$

$$\int_\Theta p(\theta) d\theta = \int_\Theta p^*(\theta) d\theta = 1 \longrightarrow \int_\Theta \eta(\theta) d\theta = 0$$

$$\log f_k(\theta) = \log f_k^*(\theta) + O(\varepsilon^2)$$

$$\log p(\theta) = \log p^*(\theta) + \varepsilon \frac{\eta(\theta)}{p^*(\theta)} + O(\varepsilon^2)$$

$$I[e(k), p(\theta)] = I[e(k), p^*(\theta)] + \varepsilon \int_\Theta \eta(\theta) \left\{ \log \frac{f_k^*(\theta)}{p^*(\theta)} - 1 \right\} d\theta + O(\varepsilon^2)$$

$$\Rightarrow \quad \forall \theta \in \Theta \rightarrow p(\theta) \propto f_k(\theta)$$

## Problems:

1) *implicit equation* **since** $f_k(\theta)$ **depends on** $p(\theta)$ **through posterior** $p(\theta|\mathbf{x}_k)$

2) *usually* $\lim_{k\to\infty} f_k(\theta)$ *divergent*: $\infty$ **info needed to know** $\theta$

3) *In general, info not defined on unbounded sets* → **sequence of priors** $\pi_k(\theta)$

*(Propper priors, defined on a sequence of compacts* $\Lambda_k \subset \Theta$ $\quad \Lambda_k \xrightarrow{k\to\infty} \Theta$ *)*

## ➡ *Define what is a "Reference Prior"*

*Berger, J.O., Bernardo J.M., Sun D (2009)*
*Ann. Stat.; Vol. 37, No. 2; 905-938*

### *Def.: Permissible Prior:*

***A strictly positive prior function*** $\pi(\theta)$ ***is a permissible prior for the model***
$M_X = \{p(\mathbf{x}\,|\,\theta)\,;\,\mathbf{x} \in X\,;\,\theta \in \Theta\}$ ***if :***

***1)*** $\forall \mathbf{x} \in X \qquad \int_{\Theta} p(\mathbf{x}\,|\theta)\pi(\theta)d\theta < \infty$ ***→ proper posterior***

***2) For some approximating compact sequence,*** $\Lambda_k \subset \Theta \qquad \Lambda_k \xrightarrow{k\to\infty} \Theta$
***the sequence of posteriors*** $\pi_k(\theta\,|\,\mathbf{x}) \propto p(\mathbf{x}\,|\,\theta)\pi_k(\theta)$
***converges logarithmically ($D_{KL}$) to*** $\pi(\theta\,|\,\mathbf{x}) \propto p(\mathbf{x}\,|\,\theta)\pi(\theta)$

### *Def.: Reference Prior*

***A permissible prior that maximizes "perfect info"*** ***(for model $M_X$)***

# Explicit form of the reference prior:

Berger, J.O., Bernardo J.M., Sun D (2009)
Ann. Stat.; Vol. 37, No. 2; 905-938

1) Consider a continuous strictly positive function $\pi^*(\theta)$ such that the corresponding posterior

$$\pi^*(\theta \mid \mathbf{z}_k) = \frac{p(\mathbf{z}_k \mid \theta)\pi^*(\theta)}{\int_\Theta p(\mathbf{z}_k \mid \theta)\pi^*(\theta)d\theta}$$

is proper and asymptotically consistent

(… the easier the better …)

2) Obtain $f_k^*(\theta) = \exp\left\{\int_X p(\mathbf{z}_k \mid \theta)\log \pi^*(\theta \mid \mathbf{z}_k)d\mathbf{z}_k\right\}$ and for any interior point $\theta_0$ of $\Theta$ define $h_k(\theta, \theta_0) = \frac{f_k^*(\theta)}{f_k^*(\theta_0)}$

3) If

3.1) each $f_k^*(\theta)$ is continuous

3.2) For any fixed $\theta$ and large k $h_k(\theta, \theta_0)$ is either monotonic in k or bounded above by some $h(\theta)$ which is integrable on any compact set

3.3) $f(\theta) = \lim_{k\to\infty} h_k(\theta, \theta_0)$ is a permissible prior function

Then $\pi(\theta) = f(\theta)$ is a reference prior for the model $M_X = \{p(\mathbf{x} \mid \theta); \mathbf{x} \in X; \theta \in \Theta\}$

Easy to implement a MC. algorithm

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

Usually are PMP

## EXAMPLE: Scale Parameter

**Show that** $\pi(\theta) \propto \frac{1}{\theta}$

**is a reference prior for the model** $p(x|\theta) = \frac{1}{\theta} f\left(\frac{x}{\theta}\right) \qquad x \in R; \quad \theta \in R^+$

**Solution:**

$$X \xrightarrow{iid} e(k) \rightarrow \boldsymbol{x}_k = \{x_1, x_2, \ldots, x_k\} \qquad p(x_1, x_2, \ldots, x_k|\theta) = \frac{1}{\theta^k} \prod_{i=1}^{k} f\left(\frac{x_i}{\theta}\right)$$

**Take, for instance:** $\pi^*(\theta) = \frac{1}{\theta^a}$

➡ **Proper posterior:** $\pi_k^*(\theta|\boldsymbol{x}) = \frac{1}{I(\boldsymbol{x})} \theta^{-(k+a)} \prod_{i=1}^{k} f\left(\frac{x_i}{\theta}\right) \qquad I(\boldsymbol{x}) = \int_0^\infty \theta^{-(k+a)} \prod_{i=1}^{k} f\left(\frac{x_i}{\theta}\right) d\theta$

$$\log \pi_k^*(\theta|\boldsymbol{x}) = -(k+a)\log\theta - \log I(\boldsymbol{x}) + \sum_{i=1}^{k} \log f\left(\frac{x_i}{\theta}\right)$$

$$\int_X p(x_1, x_2, \ldots, x_k|\theta) \log \pi_k^*(\theta|\boldsymbol{x}) dx_1 dx_2 \cdots dx_k = -(k+a)\log\theta + J_1(k) - J_2(k,\theta,a)$$

$$J_1(k)=\theta^{-k}\int_X \prod_{i=1}^{k} f\left(x_i\theta^{-1}\right)\sum_{j=1}^{j}\log f\left(x_j\theta^{-1}\right)dx_1dx_2\cdots dx_k = k\int_W f(w)\log f(w)dw = k\tilde{J}_1$$

$$J_2(k,\theta,a)=\theta^{-k}\int_X \prod_{i=1}^{k} f\left(x_i\theta^{-1}\right)\log\left[\int_0^\infty \theta'^{-(k+a)}\prod_{i=1}^{k} f\left(x_i\theta'^{-1}\right)d\theta'\right]dx_1dx_2\cdots dx_k =$$

$$=\int_X \prod_{i=1}^{k} f(w_i)\log\left[\theta^{-(k+a-1)}\int_0^\infty s^{-(k+a)}\prod_{i=1}^{k} f(sw_i)ds\right]dw_1dw_2\cdots dw_k =$$

$$=\tilde{J}_2(k,a)-(k+a-1)\log\theta$$

$$\int_{X^k} p(\mathbf{x}\,|\,\theta)\log\pi_k^*(\theta\,|\,\mathbf{x})d\mathbf{x} = -\log\theta + k\tilde{J}_1 - \tilde{J}_2(k,a) = -\log\theta + G(k,a)$$

$$f_k(\theta,\bullet)=\exp\left\{\int_{X^k} p(\mathbf{x}\,|\,\theta)\log\pi_k^*(\theta\,|\,\mathbf{x})d\mathbf{x}\right\}=\frac{1}{\theta}\exp\{G(k,a)\}$$

$$\pi(\theta)\propto\lim_{k\to\infty}\frac{f_k(\theta,\bullet)}{f_k(\theta_0,\bullet)}\propto\frac{1}{\theta}$$

## PROBLEM: Poisson Distribution

$$p(n\,|\,\theta) = e^{-\theta}\,\frac{\theta^n}{\Gamma(n+1)}$$

**Consider** $X \xrightarrow{iid} e(k) \rightarrow \mathbf{x}_k = \{n_1, n_2, \ldots, n_k\}$ **and** $\pi^*(\theta) = 1$

**Show that** $f_{k\to\infty}(\theta,\bullet) \propto (k\theta)^{-1/2}$ **and, in consequence** $\pi(\theta) \propto \lim_{k\to\infty} \dfrac{f_k(\theta,\bullet)}{f_k(\theta_0,\bullet)} \propto \dfrac{1}{\theta^{1/2}}$

## PROBLEM: Binomial Distribution

$$p(n\,|\,\theta, N) = \binom{N}{n}\theta^n (1-\theta)^{N-n}$$

**Consider** $X \xrightarrow{iid} e(k) \rightarrow \mathbf{x}_k = \{n_1, n_2, \ldots, n_k\}$ **and** $\theta \in (0,1) \rightarrow \pi^*(\theta) \propto \theta^{a-1}(1-\theta)^{b-1}$

**Show that** $\pi(\theta) \propto \lim_{k\to\infty} \dfrac{f_k(\theta,\bullet)}{f_k(\theta_0,\bullet)} \propto \dfrac{1}{\theta^{1/2}(1-\theta)^{1/2}}$

---

*Hint: Analize the behaviour of* $f_k(\theta,\bullet)$ *expanding* $\log\Gamma(z,\bullet)$ *around* $E[z]$

*and considering the asymptotic behaviour of the Polygamma Function*

$\Psi^{n)}(z) \sim a_n z^{-n} + a_{n+1} z^{-(n+1)} + \ldots$ *, the moments of the Distribution,…*

## For n>1 dimensions

**1) arrange parameters in order of importance** $\{\theta_1, \theta_2, \theta_3, \ldots\}$

**2) proceed sequentially with conditionals** $\pi(\theta_n \mid \theta_1, \theta_2, \ldots, \theta_{n-1}) \cdots \pi(\theta_2 \mid \theta_1) \pi(\theta_1)$

**(… check if they are proper)**

## Ex: n=2

$X \sim p(x|\theta, \phi)$

**parameter of interest:** $\theta$
**nuisance parameter:** $\phi$

**ordered parameterization** $\{\theta, \phi\}$

$$p(\theta, \phi, x) = p(x \mid \theta, \phi) \pi(\phi \mid \theta) \pi(\theta)$$

**1) get conditional prior** $\pi(\phi \mid \theta)$ **(reference prior for** $\phi$ **keeping** $\theta$ **fixed)**

**2) find the marginal model:** $p(\theta, x) = \pi(\theta) \underbrace{\int_\Phi p(x \mid \theta, \phi) \pi(\phi \mid \theta) d\phi}_{} = \pi(\theta) p(x \mid \theta)$

**3) get reference prior** $\pi(\theta)$ **from marginal model** $p(x \mid \theta)$

# *EXAMPLE: Multinomial*

$$\boldsymbol{X} \sim Mn(\boldsymbol{n} \mid \boldsymbol{\theta}) \qquad p(\boldsymbol{n} \mid \boldsymbol{\theta}) \propto \theta_1^{\,n_1}\theta_2^{\,n_2}\cdots\theta_k^{\,n_k}(1-\delta_k)^{n_{k+1}} \qquad \delta_k = \sum_{j=1}^{k}\theta_j$$

***ordered parameterization*** $\{\theta_1, \theta_2, \cdots, \theta_k\}$

$$\pi(\theta_1, \theta_2, \ldots, \theta_k) = \pi(\theta_k \mid \theta_{k-1}, \ldots, \theta_1)\pi(\theta_{k-1} \mid \theta_{k-2}, \ldots, \theta_1)\cdots\pi(\theta_1)$$

$$\pi(\theta_m \mid \theta_{m-1}, \ldots, \theta_1) \propto \theta_m^{-1/2}(1-\delta_m)^{-1/2} \quad \Rightarrow \quad \pi_R(\theta_1, \theta_2, \ldots, \theta_k) \propto \prod_{i=i}^{k}\theta_i^{-1/2}(1-\delta_i)^{-1/2}$$

***All are proper***

$$\boxed{p(\boldsymbol{\theta} \mid \boldsymbol{n}) \propto \left[\prod_{i=i}^{k}\theta_i^{\,n_i-1/2}(1-\delta_i)^{-1/2}\right](1-\delta_k)^{n_{k+1}}}$$

*… 2) proceed sequentially with conditionals …*

**Usually** $\pi(\phi\,|\,\theta)$ **is improper** → $\int_{\Phi} p(x\,|\,\theta,\phi)\pi(\phi\,|\,\theta)d\phi$ **divergent**

→ **Define** $\pi_k(\phi\,|\,\theta)$ **over a sequence** $\Omega_k$ **of compact sets of the full parameter space such that in the limit tend to** $\Omega$

**For instance:** $\Omega = (0,\infty)$ $\quad \Omega_k = [1/k, k]$ $\quad \bigcup_{k=1}^{\infty} \Omega_k = \Omega$

**One nuisance parameter with Joint Posterior Asymptotic Normality:**

*Bernardo J.M., Ramón J.M. (1998)*
*The Statiscian 47, 1-35*

**If:** $\Lambda$ **independent of** $\theta$

$F(\theta,\lambda)$ **such that** $\left[F_{2,2}(\theta,\lambda)\right]^{1/2} = a_1(\theta)\,b_1(\lambda)$

$S(\theta,\lambda) = F^{-1}(\theta,\lambda)$ **such that** $\left[S_{1,1}(\theta,\lambda)\right]^{-1/2} = a_0(\theta)\,b_0(\lambda)$

**Then**

$$\left.\begin{aligned} &\pi(\lambda\,|\,\theta) \propto b_1(\lambda) \\ &\pi(\theta) \propto a_0(\theta) \end{aligned}\right\} \quad \pi(\theta,\lambda) \propto a_0(\theta) b_1(\lambda)$$

**even if conditional reference priors are not proper**

**EXAMPLE:**

**Gamma Distribution:** $$p(x\,|\,a,b) = e^{-ax}x^{b-1}\frac{a^b}{\Gamma(b)}\mathbf{1}_{[0,\infty)}(x) \qquad a,b \in R^+$$

**1) Fisher's matrix elements are:** $$F_{aa} = \frac{b}{a^2} \quad F_{ab} = -\frac{1}{a} \quad F_{bb} = \Psi'(b)$$

**2) Jeffrey's prior** $$\pi_J(a,b) \propto a^{-1}\sqrt{b\Psi'(b)-1}$$

**(** $a$ **is a scale parameter)**

**2) ordering** $\{b,a\}$ $$\pi_R(b,a) \propto a^{-1}b^{-1/2}\sqrt{b\Psi'(b)-1}$$

$$\pi_{PM}(b,a) \propto a^{-1}b^{-1/2}\sqrt{b\Psi'(b)-1}$$

$\{a,b\}$ $$\pi_R(a,b) \propto a^{-1}\sqrt{\Psi'(b)}$$

$$\pi_{PM}(a,b) \propto a^{-1}\sqrt{\Psi'(b)}\left(\sqrt{b\Psi'(b)-1}\right)$$

## PROBLEM: Negative Binomial

$$X \sim p(x|\theta,a) = \frac{\Gamma(a+x)}{\Gamma(x+1)\Gamma(a)}\theta^{a}(1-\theta)^{x}$$

*a: number of failures until experiment is stopped (fixed)* $a > 0$

*X: number of successes observed* $\Omega_X = \{0,1,2,\ldots\}$

*θ: probability of failure* $0 < \theta \le 1$

$$E[X] = a(1-\theta)\theta^{-1}$$

$$F(\theta) = a\theta^{-2}(1-\theta)^{-1}$$

$$\pi_R(\theta) = \pi_J(\theta) \propto \theta^{-1}(1-\theta)^{-1/2}$$

$$p(\theta|x,a) = \frac{\theta^{a-1}(1-\theta)^{x-1/2}}{Be(a,x+1/2)}$$

## PROBLEM: Weibull Distribution

$$p(x|\alpha,\beta) = \alpha\beta(\beta x)^{\alpha-1}\exp\{-(\beta x)^{\alpha}\}\mathbf{1}_{[0,\infty)}(x)$$

$\alpha, \beta \in R^{+}$

***1) Find the transformations*** $Z = Z(X)$ ***and*** $\boldsymbol{\varphi} = \boldsymbol{\varphi}(\alpha,\beta)$ ***such that the new parameters are location and scale parameters and transform them back to get the corresponding (improper) prior***

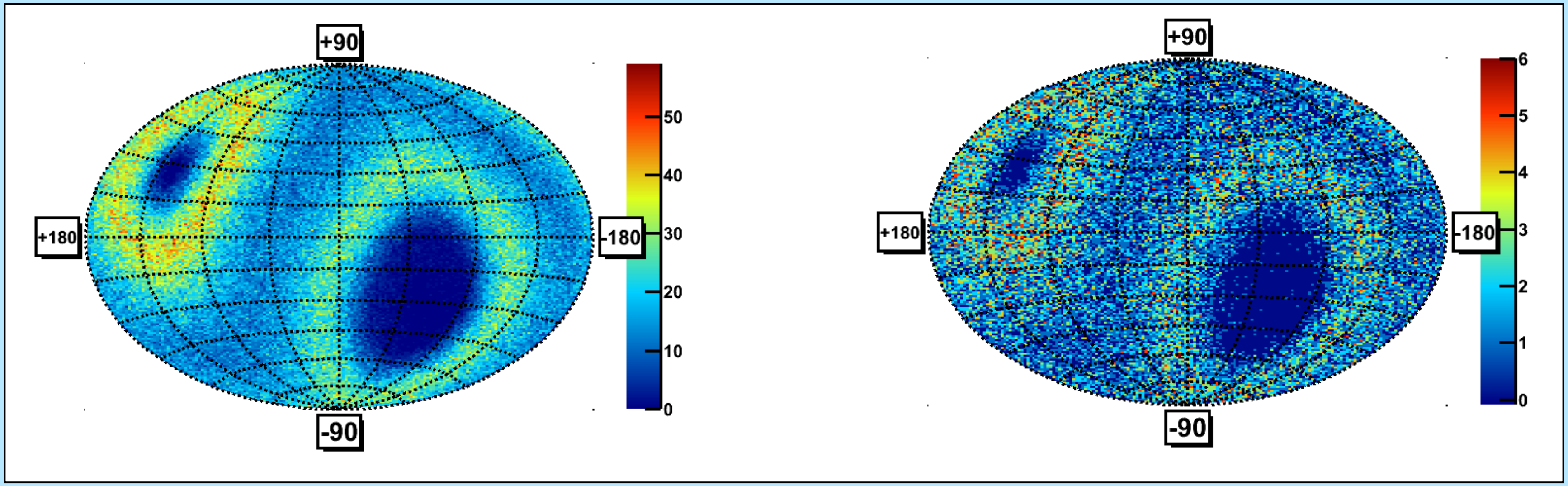
***2) Obtain the Fisher's matrix and the Jeffrey's prior***

***3) Find the reference prior***

$$\pi(\alpha,\beta) \propto \alpha^{-2}\beta^{-1}$$

***4) Show that it is a Probability Matching Prior***

# *EXAMPLE: Ratio of Poisson Parameters*




Sky maps showing the arrival directions of selected 16–350 GeV electrons (left) and positrons (right) in galactic coordinates using a Hammer-Aitoff projection. The color code reflects the number of events per bin. J. Casaus et al.; 33rd ICRC-2013

***Model:***

$$X_1 \sim Po(n_1 \mid \mu_1)$$
$$X_2 \sim Po(n_2 \mid \mu_2)$$

***independent***

***Parameter of interest:*** $\theta = \mu_1 \mu_2^{-1}$

***Nuisance parameter:*** $\phi = \mu_2$

***Ordered parameterisation:*** $\{\theta, \phi\}$

***Previous case with same efficiencies and exposure times:***

$$p(n_1, n_2 \mid \theta, \phi) = \frac{e^{-\phi(1+\theta)} \theta^{n_1} \phi^{n_1+n_2}}{\Gamma(n_1+1)\Gamma(n_2+1)}$$

**Prior:**

$$F=\begin{pmatrix}\phi\theta^{-1} & 1\\ 1 & (1+\theta)\phi^{-1}\end{pmatrix} \qquad F^{-1}=\begin{pmatrix}\theta(1+\theta)\phi^{-1} & -\theta\\ -\theta & \phi\end{pmatrix}$$

$$\left[F_{2,2}(\theta,\phi)\right]^{1/2}=\left[(1+\theta)\phi^{-1}\right]^{1/2}=a_1(\theta)\,b_1(\phi) \qquad \pi(\phi\,|\,\theta)\propto b_1(\phi)=\phi^{-1/2}$$

$$\left[S_{1,1}(\theta,\phi)\right]^{-1/2}=\left[\theta(1+\theta)\phi^{-1}\right]^{-1/2}=a_0(\theta)\,b_0(\phi) \qquad \pi(\theta)\propto a_0(\theta)=\left[\theta(1+\theta)\right]^{-1/2}$$

$$\pi(\theta,\phi)=\pi(\phi\,|\,\theta)\pi(\theta)\propto\phi^{-1/2}\left(\theta(1+\theta)\right)^{-1/2}$$

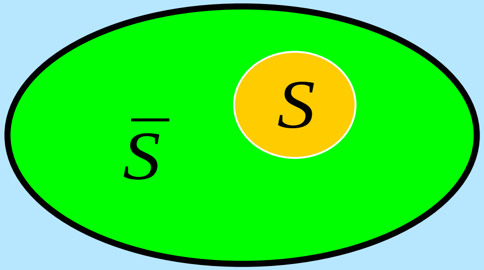
**Posterior:**

$$p\left(\theta\,|\,n_1,n_2\right)=\int p\left(\theta,\phi\,|\,n_1,n_2\right)d\phi$$

$$p\left(\theta\,|\,n_1,n_2\right)=\frac{\Gamma(n_1+n_2+1)}{\Gamma(n_1+1/2)\Gamma(n_2+1/2)}\,\frac{\theta^{n_1-1/2}}{(1+\theta)^{n_1+n_2+1}}$$

$$E[\theta^m]=\frac{\Gamma(n_1+1/2+m)\Gamma(n_2+1/2-m)}{\Gamma(n_1+1/2)\Gamma(n_2+1/2)} \qquad E[\theta\,|\,n_2\geq 1]=\frac{n_1+1/2}{n_2-1/2}$$

# Example: Characterization of a source




**region** $S$ *Events are produced with a rate* $\lambda_s$

*and collected during an "exposure" time* $t_s$

**Model:** *Events produced at S follow* $n_s \sim Po(n_s \mid \lambda_s t_s)$

*Produced events are detected with probability* $\varepsilon_s$

*Observed events:* $s \sim Bi(s \mid n_s, \varepsilon_s)$

*Poisson-Binomial model* $s \sim \sum_{n_s = s}^{\infty} Bi(s \mid n_s, \varepsilon_s) Po(n_s \mid \lambda_s t_s) = Po(s \mid \lambda_s t_s \varepsilon_s)$

**region** $\bar{S}$

*Background inferred from observations at* $b \sim Po(b \mid \lambda_b t_b \varepsilon_b)$

**Model:** $p(s, b \mid \bullet) = Po(s \mid \lambda_s t_s \varepsilon_s) Po(b \mid \lambda_b t_b \varepsilon_b)$

**2) Prior:** *Parameters of interest: Ratio of Poisson Parameters*

$$\left\{\theta = \frac{\lambda_s \varepsilon_s}{\lambda_b \varepsilon_b}, \phi = \lambda_b \varepsilon_b\right\}$$

**(usually** $\varepsilon_s = \varepsilon_b$ **)**

$$p(s,b\,|\,\bullet) \propto e^{-t_b \phi (1+a\theta)} \theta^s \phi^{b+s}$$

$$n = b+s \qquad a = \frac{t_s}{t_b}$$

$$F = \begin{pmatrix} a t_b \phi \theta^{-1} & a t_b \\ a t_b & t_b (1+a\theta)\phi^{-1} \end{pmatrix} \qquad F^{-1} = \begin{pmatrix} \theta(1+a\theta)(\phi a t_b)^{-1} & -\theta t_b^{\ -1} \\ -\theta t_b^{\ -1} & \phi t_b^{\ -1} \end{pmatrix}$$

$$\pi(\theta,\phi) = \pi(\phi\,|\,\theta)\pi(\theta) \propto \phi^{-1/2}\left(\theta(1+a\theta)\right)^{-1/2}$$

**3) Posterior:** $$p(\theta,\phi\,|\,s,b,a,t_b) \propto e^{-t_b \phi (1+a\theta)} \theta^{s-1/2} (1+a\theta)^{-1/2} \phi^{b+s-1/2}$$

**4) Integrate nuisance parameter:**

$$p(\theta\,|\,s,b,a,t_b) \propto I_0^{-1} \frac{\theta^{s-1/2}}{(1+a\theta)^{n+1}}$$

$$I_m = I_0 E[\theta^m] = \frac{1}{a^{m+s+1/2}} \frac{\Gamma(n-s-m+1/2)\Gamma(s+m+1/2)}{\Gamma(n+1)}$$

$$m < n-s$$

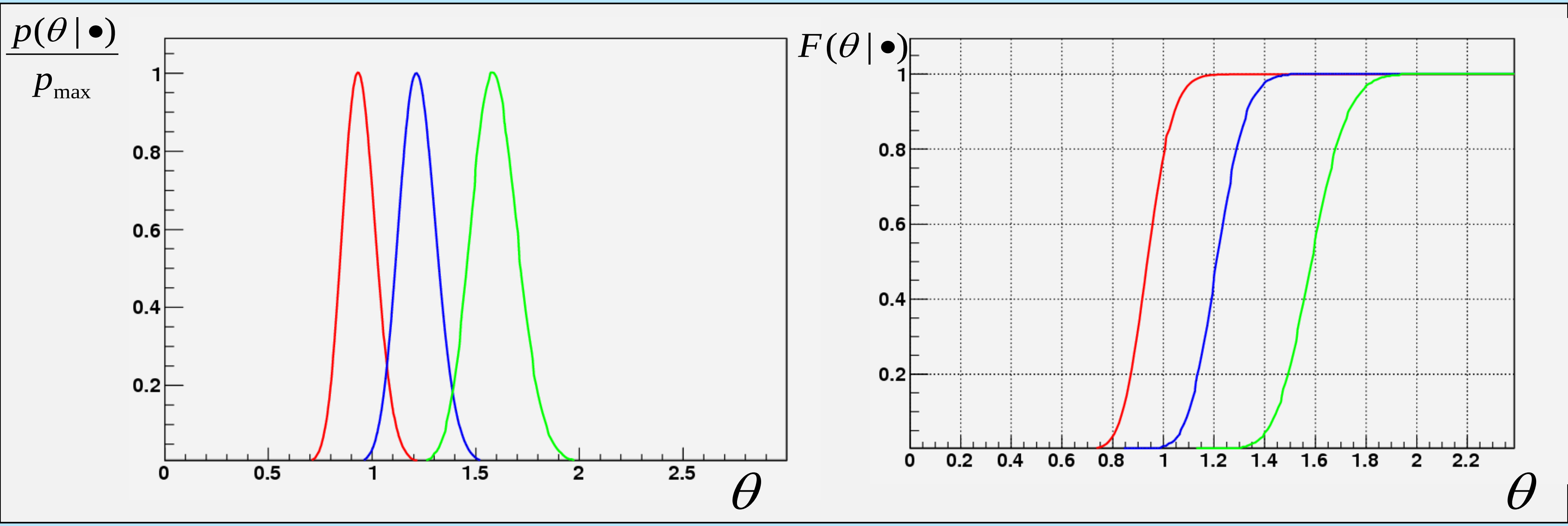
## Sensitivity to existence of a source:

MC simulation with $\varepsilon_s = \varepsilon_b$

$\lambda_b = 4.0; t_B = 200$

$t_s = 50$

$\lambda_S=4.0$ $\lambda_S=5.0$ $\lambda_S=6.0$




***If Source…*** $\lambda_s = \lambda_b + \mu$ ⟶ ***Ordered Parameterization*** $\{\mu, \lambda_b\}$

***Info on*** $\lambda_b$***: Get*** $p(\lambda_b \mid \bullet)$ ***from*** $\bar{S}$ ***region***

$$p(\lambda_s \mid n, n_B, t, t_B) \propto \pi(\mu) \int_0^\infty Po(s \mid (\lambda_b + \mu) t_s \varepsilon_s) Ga(\lambda_b \mid t_b \varepsilon_b, b + 1/2) d\lambda_b$$

$$\propto \pi(\mu)\, e^{-\mu t_s \varepsilon} \sum_{k=0}^{s} \binom{s}{k} \mu^k a_k$$

$$a_k = (t\varepsilon)^k \Gamma(n - k + 1/2)$$

$\varepsilon = \varepsilon_s = \varepsilon_b$

$t = t_s + t_b$

$n = b + s$

# *PRIORS WITH PARTIAL INFORMATION*

# …Including Partial Information in the PRIOR…

$$\pi(\theta) \qquad \int_{\Theta} g_j(\theta)\pi(\theta)d\theta = a_j \qquad X \sim p(x|\theta)$$

***1) get reference prior*** $\pi_0(\theta)$ ***from model*** $p(x|\theta)$

***2) find the prior*** $\pi(\theta)$ ***for which*** $\pi_0(\theta)$ ***is the best approximation in the Kullback-Leibler sense with Lagrange multipliers for the constraints:***

$$I = \int_{\Theta} \pi(\theta)\ln\frac{\pi(\theta)}{\pi_0(\theta)}d\theta + \sum_j \lambda_j \left[\int_{\Theta} g_j(\theta)\pi(\theta)d\theta - a_j\right]$$

$$\pi(\theta) = \pi_s(\theta) + \varepsilon\eta(\theta) \qquad I = I_s + \varepsilon\int_{\Theta}\eta(\theta)\left[\ln\frac{\pi_S(\theta)}{\pi(\theta)} + 1 + \sum_j \lambda_j g_j(\theta)\right]d\theta + O(\varepsilon^2)$$

$$\pi_S(\theta) \propto \pi(\theta)e^{\sum_j \lambda_j g_j(\theta)} \qquad \lambda_j \,\Big|\, \int_{\Theta} g_j(\theta)p_0(\theta)d\theta = a_j$$

# EXAMPLE: Unfolding (2)

**PRIOR:** $$\boldsymbol{\theta} \sim \pi_0(\boldsymbol{\theta} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) = GDi(\boldsymbol{\theta} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) \qquad \sum_{i=1}^{nt} \theta_i = 1$$

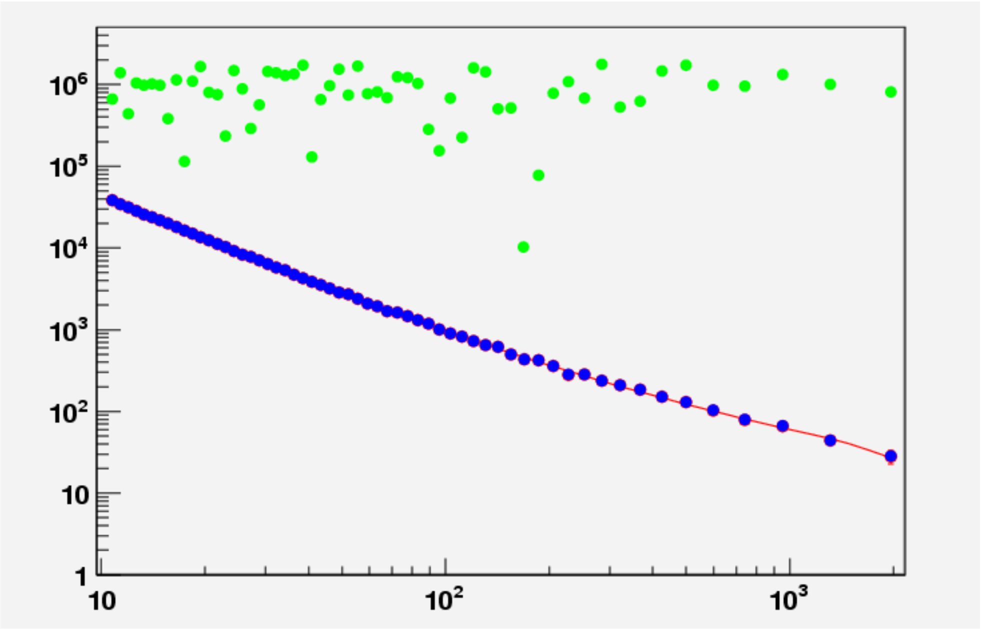
**Additional constraints:**

**for instance existence of 1st derivative** $g_j(\theta) = (\varepsilon_{j,j-1}\theta_{j+1} + \varepsilon_{j+1,j}\theta_{j-1} - \theta_j \varepsilon_{j+1,j-1})^2 = 0$

$$\varepsilon_{i,k} = x_i - x_k$$

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) = \pi_0(\boldsymbol{\theta} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}) e^{\sum_j \lambda_j g_j(\boldsymbol{\theta})}$$

**POSTERIOR:** $$p(\boldsymbol{\theta} \mid \mathbf{d}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \propto p(\mathbf{d} \mid \boldsymbol{\varphi}(\boldsymbol{\theta}))\pi(\boldsymbol{\theta} \mid \boldsymbol{\alpha}, \boldsymbol{\beta})$$




***MCM to draw sampling of $\boldsymbol{\theta}$***

***MC simulation of AMS-02 Cosmic Ray spectrum for p***

# *PRIORS for MODELS WITH DISCRETE PARAMETERS*

# Discrete Parameter Prior

## 1) Hierarchical Model:

$\theta$ **Discrete (or continuous)** **Model:** $p(\boldsymbol{x} \mid \theta)$

**Specify:** $p(\theta \mid \phi)$ $\phi$ **: continuous hyperparameter** $p(\theta, \phi) = p(\theta \mid \phi) p(\phi)$

$$p(\theta, \phi, \boldsymbol{x}) = p(\boldsymbol{x} \mid \theta, \phi) p(\theta, \phi) = p(\boldsymbol{x} \mid \theta) p(\theta \mid \phi) p(\phi)$$

$$p(\phi, \boldsymbol{x}) = \int p(\theta, \phi, \boldsymbol{x}_k) d\theta = p(\phi) \int p(\boldsymbol{x} \mid \theta) p(\theta \mid \phi) d\theta = p(\boldsymbol{x} \mid \phi) p(\phi)$$

$\int \leftrightarrow \sum$

➡ **Model:** $p(\boldsymbol{x} \mid \phi) = \int p(\boldsymbol{x} \mid \theta) p(\theta \mid \phi) d\theta$ ➡ **Prior:** $p(\phi)$

$p(\theta, \phi) = p(\theta \mid \phi) p(\phi)$ ➡ **Prior:** $\boxed{p(\theta) = \int p(\theta, \phi) d\phi = \int p(\theta \mid \phi) p(\phi) d\phi}$

## *2) Simpler approach:*
## *Continuum embedding of parameter space*

***Example: CHARGE Determination with a Cherenkov Detector***

***Observed Number of Cherenkov Photons*** $$X_\gamma \sim Po(n_\gamma \mid n_0 Z^2)$$

$n_0$ : ***(known) Expected number of photons for a Z=1 particle***

$$\pi(Z) \sim c \qquad p(Z \mid n_\gamma, n_0) \propto e^{-n_0 Z^2} Z^{2 n_\gamma}$$

$$P(Z = i \mid n, n_0) = \gamma_+(i) - \gamma_-(i)$$

$$\gamma_\pm(i) = \frac{\gamma(n + 1/2, n_0 (i \pm \varepsilon_\pm)^2)}{\Gamma(n + 1/2)}$$

***(Check accuracy)***

***Last,***

***Another kind of problems you may be interested in:***

***Regression Problems***

# Regression Problems: linear regression

**Data:** $\{(x_i, y_i); i = 1, \ldots, n\}$

**Linear Model:** $y_i = a + bx_i + \varepsilon_i \qquad \varepsilon_i \sim N(\varepsilon \,|\, 0, \sigma) \qquad \sigma \quad \textit{unknown}$

**Model:** $p(\mathbf{y} \,|\, \mathbf{x}, a, b, \sigma) \propto \sigma^{-n} \exp\left\{ \frac{-1}{2\sigma^2} \sum_{i=1}^{n} (y_i - a - bx_i)^2 \right\}$

***0) Parameters of interest:*** $\{a, b, \sigma\}$

***1) Less messy in matrix form ( and easier to generalize)***

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \boldsymbol{P} = \begin{pmatrix} a \\ b \end{pmatrix} \quad \boldsymbol{D} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \quad p(\mathbf{y} \,|\, \mathbf{x}, a, b, \sigma) \propto \sigma^{-n} \exp\left\{ \frac{-1}{2\sigma^2} (\boldsymbol{Y} - \boldsymbol{DP})^t (\boldsymbol{Y} - \boldsymbol{DP}) \right\}$$

***2) Express the exponent in a more convenient form: Find the minimum*** $\boldsymbol{P}^* = \begin{pmatrix} a^* \\ b^* \end{pmatrix}$

$$\boldsymbol{D}^t \boldsymbol{D} \, \boldsymbol{P}^* = \boldsymbol{D}^t \, \boldsymbol{Y} \qquad \boldsymbol{P}^* = (\boldsymbol{D}^t \boldsymbol{D})^{-1} \boldsymbol{D}^t \, \boldsymbol{Y}$$

$$S = (\boldsymbol{Y} - \boldsymbol{DP})^t (\boldsymbol{Y} - \boldsymbol{DP}) = (\boldsymbol{Y} - \boldsymbol{DP}^*)^t (\boldsymbol{Y} - \boldsymbol{DP}^*) + \underbrace{(\boldsymbol{P} - \boldsymbol{P}^*)^t \boldsymbol{D}^t \boldsymbol{D} (\boldsymbol{P} - \boldsymbol{P}^*)}_{\textbf{dependence on } \{a, b\}}$$

**Model:** $$p(\boldsymbol{y}\,|\,\boldsymbol{x},a,b,\sigma) \propto \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} S(a,b,\boldsymbol{y},\boldsymbol{x})\right\}$$

$$S(a,b,\boldsymbol{y},\boldsymbol{x}) = \left(\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{P}^*\right)^t \left(\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{P}^*\right) + \left(\boldsymbol{P} - \boldsymbol{P}^*\right)^t \boldsymbol{D}^t \boldsymbol{D}\left(\boldsymbol{P} - \boldsymbol{P}^*\right)$$

**Prior:** $$\boxed{\pi(a,b,\sigma) \propto \frac{1}{\sigma}}$$

**Posterior:** $$p(a,b,\sigma\,|\,\boldsymbol{y},\boldsymbol{x}) \propto \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2} S(a,b,\boldsymbol{y},\boldsymbol{x})\right\}$$

$$\boxed{p(a,b\,|\,\boldsymbol{y},\boldsymbol{x}) = \int_0^\infty p(a,b,\sigma\,|\,\boldsymbol{y},\boldsymbol{x})\,d\sigma \propto \left\{1 + \frac{\left(\boldsymbol{P} - \boldsymbol{P}^*\right)^t \boldsymbol{D}^t \boldsymbol{D}\left(\boldsymbol{P} - \boldsymbol{P}^*\right)}{\left(\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{P}\right)^t \left(\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{P}\right)}\right\}^{-n/2}} \quad (*)$$

---

***1) Different and known variances:*** $\varepsilon_i \sim N(\varepsilon\,|\,0,\sigma_i)$ $\quad \sigma_i$ *unknown*

***2) Other functional form:*** $y_i = g(x_i) + \varepsilon_i$

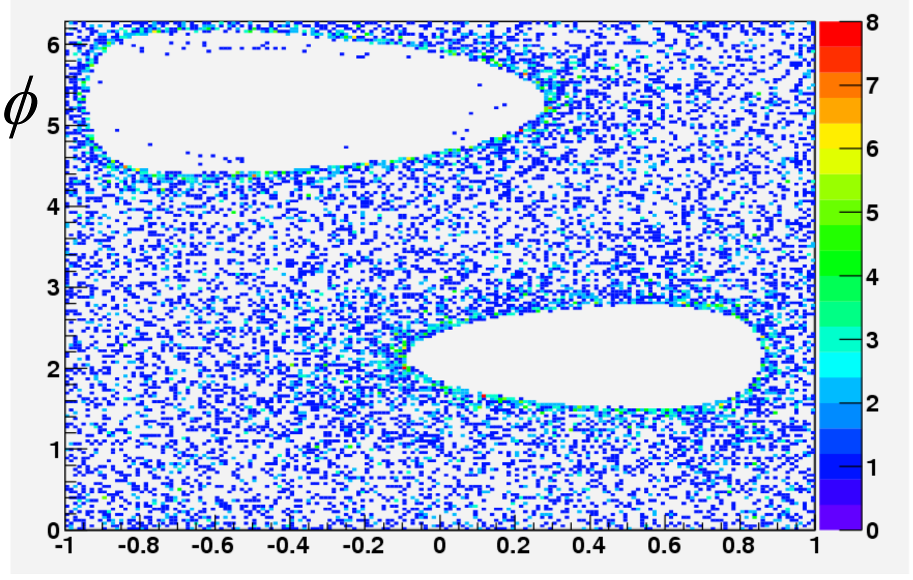
***Regression function***

- ***Prior specification quite involved***
- ***may work… decompose the regression function in a linear combination of orthogonal polynomials…***

***(*) (Same formal solution for*** $y_i = b_0 + b_1 x_{i1} + \ldots + b_n x_{in} + \varepsilon_i$ ***)***

# Example: Spherical harmonics

*… different approaches…*




$$p(\theta,\phi\,|\,\boldsymbol{a}) = \frac{1}{4\pi}\left(1 + a^{lm}Y_{lm}(\theta,\phi\,|\,\boldsymbol{a})\right)$$

$$\forall(\theta,\phi)\in\Omega \quad p(\theta,\phi\,|\,\boldsymbol{a}) \geq 0 \quad \xrightarrow{l=1} \quad \sum_{m=-1}^{1} a_{1m}^2 \leq \frac{4\pi}{3}$$

$$\pi(a_{1-1},a_{10},a_{11}) = Un\left(S_3(r = \frac{4\pi}{3})\right)$$

**Real basis in Ω:**

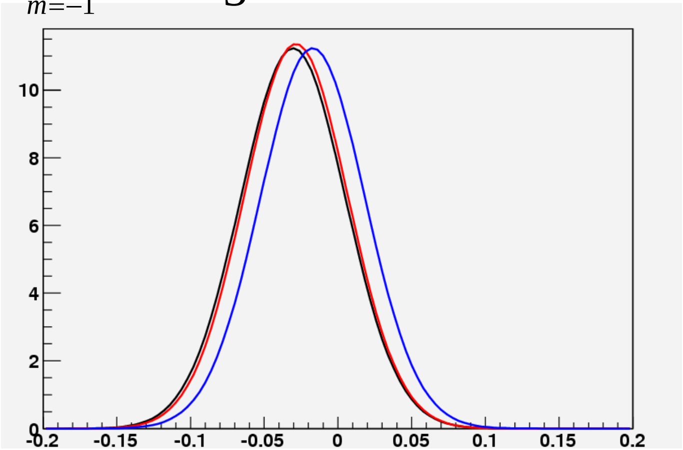
$$\int_\Omega Y_{lm}(\theta,\phi)Y_{l'm'}(\theta,\phi)d\mu = \delta_{ll'}\delta_{mm'}$$

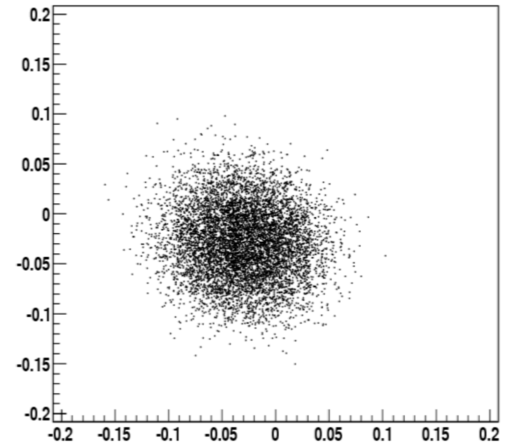
$$\int_\Omega Y_{lm}(\theta,\phi)d\mu = \sqrt{4\pi}\,\delta_{l0}$$

$$d\mu = \sin\theta\, d\theta d\phi$$

$$p(r\,|\,n_1,n_2) = N(n_1,n_2)\frac{r^{n_1-1/2}}{(1+r)^{n+1}}$$

$$r_i = \sum_{l=0}^{\infty}\sum_{m=-l}^{l} a_{lm}\int_{\Omega_i} Y_{lm}(\theta,\phi)d\mu$$





# Problem: Linear Regression

***(with uncertainty in x and y)***

**Data:** $\{(x_i, y_i); i = 1, \ldots, n\}$

**Linear Model:** $y = a + bx$

**Model:** $(X_i, Y_i) \sim N(x_i, y_i \mid x_i^0, y_i^0, \sigma_{xi}, \sigma_{yi})$

**Assume precisions** $(\sigma_{xi}, \sigma_{yi})$ **are known and show that:**

$$p(a,b \mid \boldsymbol{x}, \boldsymbol{y}) \sim \pi(a,b) \prod_{i=1}^{n} \left( \frac{\sigma_{xi}\sigma_{yi}}{\sqrt{\sigma_{yi}^2 + b^2 \sigma_{xi}^2}} \right) \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \frac{(y_i - a - bx_i)^2}{\sigma_{yi}^2 + b^2 \sigma_{xi}^2} \right\}$$

**Take** $\pi(a,b) = \pi(a)\pi(b) = c$ **and obtain** $p(a,b \mid \boldsymbol{x}, \boldsymbol{y})$

$$\sigma_y^2 + \left( \frac{\partial y}{\partial x} \right)^2 \sigma_x^2$$

# *TEST OF HYPOTHESIS*

***…Evaluate evidence in favour of a scientific theory***

***Again fundamental conceptual difference:***

***F: What is the probability to get this data given a specific model?***
***B: What are the probabilities of the various hypothesis given the data?***

## *DECISION THEORY*

# DECISION THEORY

**PROBLEM: How to choose the optimal action among a set of different alternatives** ***(… Games Theory)***

***For a given problem, we have to specify:***

$\Omega_\theta$ ***Set of all possible "states of nature"*** ***(Parameter Space)***

$\Omega_X$ ***Set of all possible experimental outcomes*** ***(Sample Space)***

$\Omega_A$ ***Set of all possible actions to be taken***

---

***Example: Disease Test***

$\Omega_\theta$ ***{sane, ill}***

$\Omega_X$ ***{test +, test -}***

$\Omega_A$ ***{apply treatment, do not apply treatment}***

*In nontrivial situations, we can't take any action without potential losses*

**2) Loss Function**

**Basic element in Decision Theory: Loss Function**

$$l(a\,|\,\theta): \quad (\theta,a)\in\Omega_\theta\times\Omega_A\rightarrow\Re^+ +\{0\}$$

*Quantifies the loss associated to take an action (or decision)* $a$
*when the "state of nature" is* $\theta$

*…What do we know about* $\theta$ *?*

*The knowledge we have on the "state of nature" is quantified by the posterior density* $p(\theta\,|\,x)$

**3) Risk Function**

$$R(a\,|\,x)=E_\theta[l(a\,|\,\theta)]=\int_{\Omega_\theta} l(a\,|\,\theta)\;p(\theta\,|\,x)\,d\theta$$

*Risk associated to take the action* $a$
*having observed the data* $x$

***4) Bayesian Decision:*** ***Take the action*** $a(x)$ ***that minimises the Risk (Minimum Expected loss)***

$$a(x) \quad | \quad \min\{R(a \mid x)\}$$

## *Two types of problems:*

***Hypothesis Testing:*** $\Omega_A =$ ***{accept, reject} an hypothesis***

***Inference:***

$$\Omega_A \subseteq \Re$$

$a(x)$ ***statistic that we shall take as an estimator of*** $\theta$

# Hypothesis Testing: Example with two alternatives

$$P(H_i \mid data) = \frac{P(data \mid H_i)P(H_i)}{P(data)}$$

***Hypothesis are exclusive and exhaustive***

## Possible actions:

$a_1$, $a_2$ ***action to be taken if we decide for hypothesis*** $H_1$, $H_2$

## Loss function

| | ***take action*** | ***(i.e. choose hypothesis)*** | ***when "state of nature" is*** |
|---|---|---|---|
| $l(a_1 \mid H_1) = l_{11} \geq 0$ | $a_1$ → | $H_1$ → | $H_1$ |
| $l(a_2 \mid H_2) = l_{22} \geq 0$ | $a_2$ → | $H_2$ → | $H_2$ |
| $l(a_1 \mid H_2) = l_{12} > 0$ | $a_1$ → | $H_1$ → | $H_2$ |
| $l(a_2 \mid H_1) = l_{21} > 0$ | $a_2$ → | $H_2$ → | $H_1$ |

**Risk Function:** $$R(a_i \mid data) = \sum_{j=1}^{2} l(a_i \mid H_j)\, p(H_j \mid data)$$

$$R(a_1 \mid data) = l_{11}\, p(H_1 \mid data) + l_{12}\, p(H_2 \mid data)$$

$$R(a_2 \mid data) = l_{21}\, p(H_1 \mid data) + l_{22}\, p(H_2 \mid data)$$

**Bayesian Decision: Take the action that minimises the Risk**

***take action*** $a_1$ ***(choose hypothesis*** $H_1$ ***) if*** $R(a_1 \mid data) < R(a_2 \mid data)$

$$p(H_1 \mid data)(l_{11} - l_{21}) < p(H_2 \mid data)(l_{22} - l_{12})$$

***If we take*** $l_{11} = l_{22} = 0$ ➡

$$\boxed{\frac{p(H_1 \mid data)}{p(H_2 \mid data)} > \frac{l_{12}}{l_{21}}}$$

*(same for* $a_2$ *)*

$$\frac{p(H_2 \mid datos)}{p(H_1 \mid datos)} > \frac{l_{21}}{l_{12}}$$

## Bayes Factor:

$$P(H_i \mid data) = \frac{P(data \mid H_i) P(H_i)}{P(data)}$$

***take action*** $a_1$ ***(decide for hypothesis*** $H_1$ ***) if***

$$\frac{p(H_1 \mid data)}{p(H_2 \mid data)} = \frac{p(data \mid H_1)}{p(data \mid H_2)} \frac{p(H_1)}{p(H_2)} > \frac{l_{12}}{l_{21}}$$

***Posterior odds*** ***Evidence from data*** ***Prior odds***

***Change of prior beliefs…***

***How strongly data favours one model over the other***

$$B_{12} \equiv \frac{p(data \mid H_1)}{p(data \mid H_2)}$$

***Ratio of likelihoods***

→ ***Choose hypothesis*** $H_1$ ***if*** $B_{12} > \frac{l_{12}\, p(H_2)}{l_{21}\, p(H_1)}$

---

***Usually:*** $l_{12} = l_{21}$ ***( … 0-1 loss function)*** ***decide for hypothesis*** $H_1$ ***if***

$p(H_1) = p(H_2)$ $B_{12} > 1$

# HYPOTHESIS TESTING: General cases:

**Hypothesis:**

- **SIMPLE:** *specify everything about the model/models, including values of parameters*

$$\theta \in \Theta = \Theta_1 \bigcup \Theta_2 \quad ; \Theta_i = \{\theta_i\}$$

- **COMPOSITE:** *parameter values are not specified by hypothesis*

***(+ different models and nuisance parameters)***

**S-S:**

$$\frac{p(H_1 \mid data)}{p(H_2 \mid data)} \quad \ldots \quad \frac{\int_\Phi p_1(x \mid \theta_1, \phi)\pi_1(\theta_1, \phi)d\phi}{\int_\Phi p_2(x \mid \theta_2, \phi)\pi_2(\theta_2, \phi)d\phi}$$

$$\ldots \quad \left( \frac{p(x \mid \theta_1)\pi(\theta_1)}{p(x \mid \theta_2)\pi(\theta_2)} \right)$$

**C-C:**

$$\frac{\int_{\Theta_1} \left[ \int_\Phi p_1(x \mid \theta, \phi)\pi_1(\theta, \phi)d\phi \right] d\theta}{\int_{\Theta_2} \left[ \int_\Phi p_2(x \mid \theta, \phi)\pi_2(\theta, \phi)d\phi \right] d\theta}$$

$$\ldots \quad \left( \frac{p(x \mid M_1)}{p(x \mid M_2)} \right)$$

***Marginal likelihoods of data for the two models***

**Example:**

**S-S:**

$H$ : ***particle with Z=2 is a $^3$He nucleus***

$\overline{H}$ : ***particle with Z=2 is a $^4$He nucleus***

$$B_{12} = \frac{p(m_{obs} \mid H)}{p(m_{obs} \mid \overline{H})} = \frac{N(m_{obs} \mid m_1, \sigma_1)}{N(m_{obs} \mid m_2, \sigma_2)} = \frac{\sigma_2}{\sigma_1} \exp\left\{ \frac{(m_{obs} - m_2)^2}{2\sigma_2^{\,2}} - \frac{(m_{obs} - m_1)^2}{2\sigma_1^{\,2}} \right\}$$

$$l_{12} = l_{21}$$

$$l_{11} = l_{22} = 0$$

***In favour of H if:***

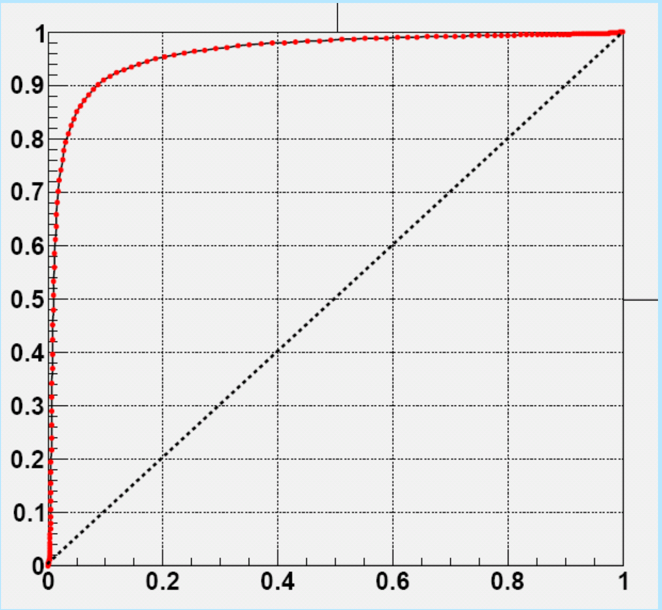
$$B_{12} > \frac{p(\overline{H})}{p(H)} \rightarrow \ln B_{12} > \ln \frac{p(\overline{H})}{p(H)}$$

***Critical mass value:***

$$\sigma_1^{\,2}(m_C - m_2)^2 - \sigma_2^{\,2}(m_C - m_1)^2 = 2\sigma_1^{\,2}\sigma_2^{\,2} \ln \frac{P(\overline{H})\sigma_1}{P(H)\sigma_2}$$

$$m_o < m_C \rightarrow {}^3He$$

$$m_o > m_C \rightarrow {}^4He$$




***Usually, keep p(H) and P(Hbar)***

# Receiver Operating Characteristic

**For each value of $m_c$**

| | $H$ | $\overline{H}$ |
|---|---|---|
| + | $P(+\|H,c)$ | $P(+\|\overline{H},c)$ |
| − | $P(-\|H,c)$ | $P(-\|\overline{H},c)$ |

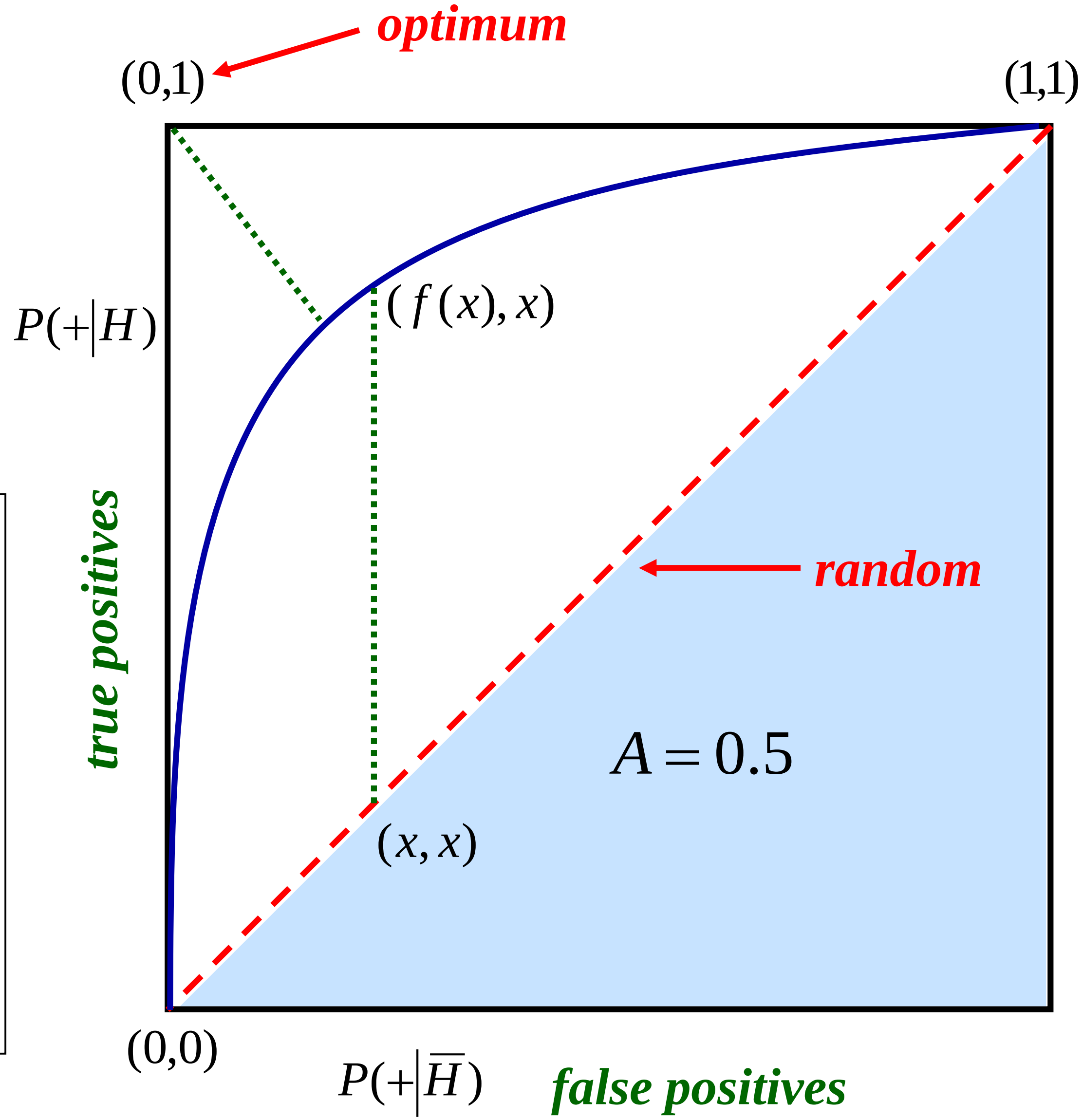
$$P(+|H,c) = F_1(c) = \int_{-\infty}^{c} f_1(u)du$$

$$P(+|\overline{H},c) = F_2(c) = \int_{-\infty}^{c} f_2(u)du$$

$$x = F_2(c) \rightarrow f(x) = F_1(F_2^{-1}(x))$$

$$A = \int_0^1 f(x)dx = \int_0^1 F_1(F_2^{-1}(x))dx$$

$$A = \int_{-\infty}^{\infty} F_1(u) f_2(u)du = \int_{-\infty}^{\infty} f_2(u)du \int_{-\infty}^{u} f_1(x)dx$$




$$d^2 = (f(x) - x)^2 \rightarrow \max(d)$$

**Example:** $X \sim Un(x|0,\theta) \quad \Omega_X = (0,\theta) \qquad X \overset{iid}{\to} e(m) = \{x_1, x_2, ..., x_m\}$

***We have seen already that:***

*Sufficient statistic* $t = \max\{x_1, x_2, ..., x_n\}$

$\theta$ *Scale parameter* $\pi(\theta) \propto 1/\theta$

$$p(\theta \mid n,t) = n\frac{t^n}{\theta^{n+1}} \mathbf{1}_{(t,\infty)}(\theta)$$

$$\Theta \sim \quad Pa(\theta \mid n,t)$$

**C-C: Hypothesis:**

- $H_o$ ***(null hypothesis)*** $\theta \in (0,\alpha]$
- $H_1$ ***(alternative hypothesis)*** $\theta \in (\alpha,\infty)$

$$P(H_0): \int_0^\alpha p(\theta \mid n,t)\mathbf{1}_{(t,\infty)}(\theta)d\theta = \left[1-\left(t/\alpha\right)^n\right]\mathbf{1}_{(t,\infty)}(\alpha)$$

$$P(H_1): \int_\alpha^\infty p(\theta \mid n,t)\mathbf{1}_{(t,\infty)}(\theta)d\theta = \mathbf{1}_{(0,t)}(\alpha) + \left(t/\alpha\right)^n \mathbf{1}_{(t,\infty)}(\alpha)$$

***0-1 Loss:***

$$\frac{P(H_0)}{P(H_1)} = \begin{cases} 0 & if \quad \alpha \le t \\ \left(\alpha/t\right)^n - 1 & if \quad \alpha > t \end{cases}$$

$$if \quad \alpha > t \quad P(\theta \in [\alpha,\infty)) = \left(t/\alpha\right)^n$$

# *STATISTICAL INFERENCE*

## *Point Estimation*

*"…when you cannot express it in numbers,*
*your knowledge is of a meagre and unsatisfactory kind."*
*(Lord W.T. Kelvin)*

# INFERENCE: (Point estimation)

**1) Quadratic Loss:** $l(a\,|\,\theta)=(\theta-a)^2$

$$R(a\,|\,x)=E_\theta[l(a\,|\,\theta)]=\int_{\Omega_\theta}(\theta-a)^2\;p(\theta\,|\,x)\,d\theta$$

$$a\,|\,\min[R(a\,|\,x)]\rightarrow\int_{\Omega_\theta}(\theta-a)\;p(\theta\,|\,x)\,d\theta=0$$

***mean***

$a=E[\theta]$

**2) Lineal Loss:** $l(a\,|\,\theta)=c_1(a-\theta)\boldsymbol{1}_{(-\infty,a]}(\theta)+c_2(\theta-a)\boldsymbol{1}_{(a,\infty)}(\theta)$

$$R(a\,|\,x)=c_1\int_{-\infty}^{a}(a-\theta)\;p(\theta\,|\,x)\,d\theta+c_2\int_{a}^{\infty}(\theta-a)\;p(\theta\,|\,x)\,d\theta$$

$$a\,|\,\min[R(a\,|\,x)]\rightarrow P(\theta\le a)=\frac{c_2}{c_1+c_2}$$

$c_1=c_2$ ***median***

**3) Zero-one Loss:** $l(a\,|\,\theta)=1-\boldsymbol{1}_{B(a;\varepsilon)}(\theta)$

***mode***

$$\min\int_{\Omega_\theta}(1-\boldsymbol{1}_{B(a;\varepsilon)}(\theta))\;p(\theta\,|\,x)\,d\theta=\max\int_{B(a;\varepsilon)}p(\theta\,|\,x)\,d\theta$$

$\theta\,|\,\max\{p(\theta\,|\,x)\}$

# CREDIBLE REGIONS

$$[\theta_1, \theta_2] \in \Theta \quad | \int_{\theta_1}^{\theta_2} p(\theta \mid \mathbf{x})\, d\theta = \alpha$$

**HPD (Highest Probability Density): Smallest possible volume** $V_\alpha \subset \Theta$ **in parameter space such that** $\int_{V_\alpha} p(\theta \mid \mathbf{x})\, d\theta = \alpha$

**Equivalent definitions:**

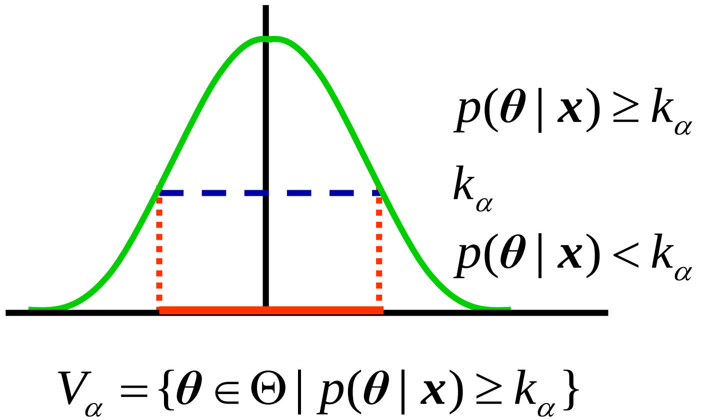
**1)** $V_\alpha = \{\boldsymbol{\theta} \in \Theta\}$ **such that**

**1)** $\int_{V_\alpha} p(\theta \mid \mathbf{x})\, d\theta = \alpha$

**2)** $\forall \boldsymbol{\theta}_1 \in V_\alpha \quad and \quad \boldsymbol{\theta}_2 \notin V_\alpha$
$\rightarrow p(\boldsymbol{\theta}_1 \mid \mathbf{x}) \geq p(\boldsymbol{\theta}_2 \mid \mathbf{x})$

**2)** $V_\alpha = \{\boldsymbol{\theta} \in \Theta \mid p(\boldsymbol{\theta} \mid \mathbf{x}) \geq k_\alpha\}$

**where** $k_\alpha$ **is the largest constant for which** $P(\boldsymbol{\theta} \in V_\alpha) \geq \alpha$




**HPD regions may not be connected (for instance more than one mode)**

**One dimension:**

$$\underbrace{\lambda = -\frac{1}{p(\theta_2 \mid x)} \qquad \lambda = -\frac{1}{p(\theta_1 \mid x)}}_{p(\theta_1 \mid x) = p(\theta_2 \mid x)} \qquad \underbrace{\int_{\theta_1}^{\theta_2} p(\theta \mid x)\, d\theta = \alpha}_{\theta_1 \neq \theta_2}$$

***If distribution has one mode and is symmetric*** ➡ $\theta_2 = 2E[\theta] - \theta_1$

***Some useful properties:*** ***Usually, HPD regions determined from MC sampling***

***1)*** $p(\boldsymbol{\theta}_1 \mid \mathbf{x}) = p(\boldsymbol{\theta}_2 \mid \mathbf{x}) \rightarrow$ ***both are included*** $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in HPD_\alpha$

***or excluded*** $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \notin HPD_\alpha$

***2)*** ***For a given CL*** $\alpha$ ***there is a*** $p_\alpha^*$ ***value such that*** $HPD_\alpha = \{\forall \boldsymbol{\theta} \in \Theta \mid p(\boldsymbol{\theta} \mid \mathbf{x}) \geq p_\alpha^*\}$
***(Def. 2)***

***3)*** $\phi = \phi(\theta)$ ***one to one:***

➡ ***Region with probability content*** $\alpha$ ***in*** $\theta$ ***has probability content*** $\alpha$ ***in*** $\phi$

***… but in general is not HPD unless linear relation***

***4)*** ***In general equal tailed credible intervals may not have the smallest size (are not HPD)***

## *EXAMPLE: $^3He$ and $^4He$ nuclei (AMS-01)*

***Observed (identified as):*** $n_3 = 97 \qquad n = 115$

***Identification criteria:*** $P(4|4), \quad P(3|3)$ ***Probability that an event of type i=3,4 is identified correctly***

***Parameter of interest:*** $\theta$ ***Probability that a nucleus is of $^3He$***

***Probability to identify a nucleus as $^3He$:***

$$\phi(\theta, p_{33}, p_{34}) = P(3|3)\theta + P(3|4)(1-\theta) =$$

$$= (1 - p_{44}) - \theta(1 - p_{33} - p_{44})$$

***Model:*** $X_3 \sim P(n_3, n_4 \mid \theta, p_{33}, p_{34}) \propto \phi^{n_3}(1-\phi)^{n-n_3}$

***Ordered parametrisation:*** $\{\theta, p_{33}, p_{34}\}$

⟶ ***prior:*** $\pi(\phi, p_{33}, p_{34})$

$\{\theta, p_{33}, p_{34}\}$

$$\pi(\phi, p_{33}, p_{34}) = \pi(p_{33})\pi(p_{34})\pi(\phi)$$

***Knowledge on:***

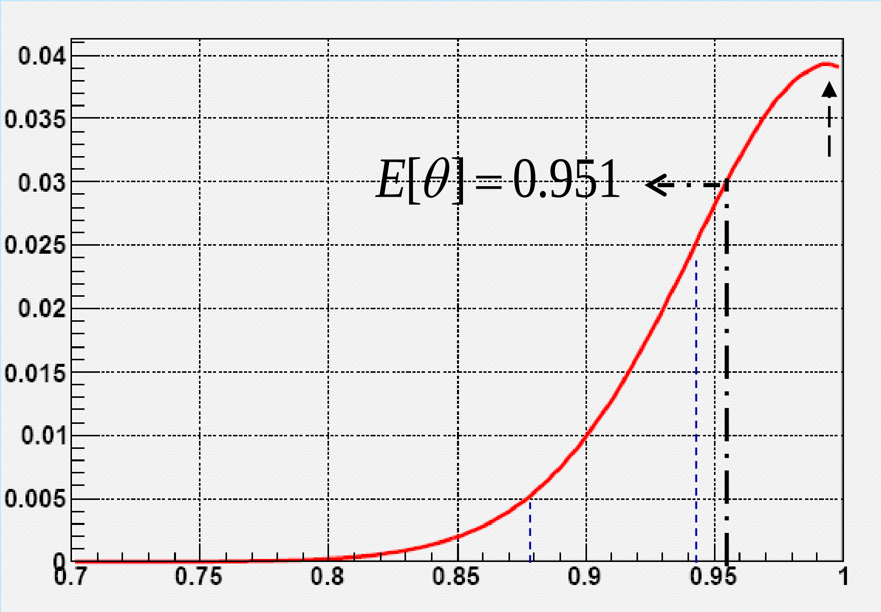
$\pi(p_{33}) \sim N(p_{33} \mid p^0_{33}, \sigma_{33})$ ***(from MC sampling)***

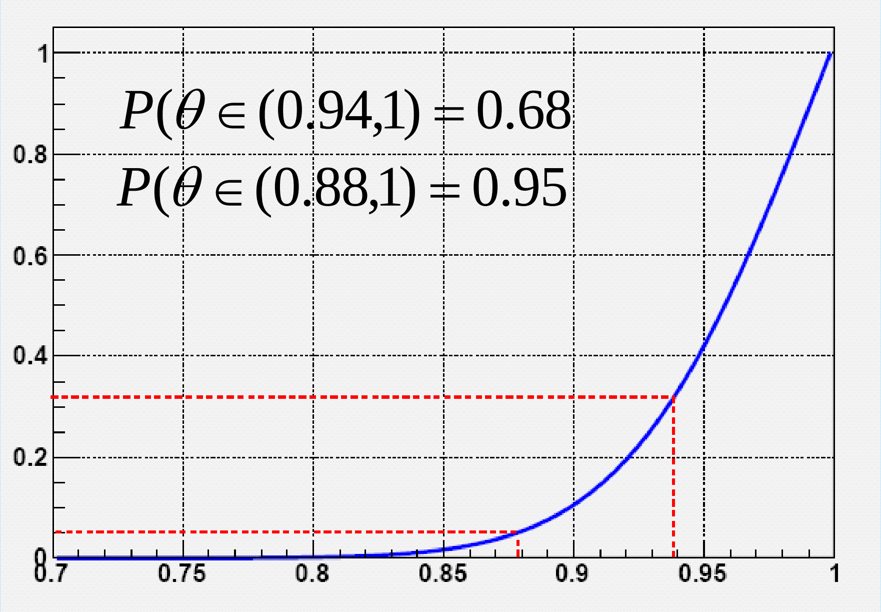
$\pi(p_{44}) \sim N(p_{44} \mid p^0_{44}, \sigma_{44})$ ***(from MC sampling)***

$(p^0_{33} = 0.852, p^0_{44} = 0.803, \sigma_{ii} \approx 0.005)$

$$\pi(\phi) \propto \left[\phi(1-\phi)\right]^{-1/2}$$

$$p(\theta \mid n_3, n_4) \propto \int_{p_{33}, p_{34}} \phi(\theta, p_{33}, p_{34})^{n_3-1/2}(1-\phi(\theta, p_{33}, p_{34}))^{n_4-1/2}\pi(p_{33})\pi(p_{34})dp_{33}dp_{34}$$



***The END***

*… of $L_2$*