David Giles

Bayesian Econometrics

2. Constructing Prior Distributions

- Constructing a Prior Distribution to reflect our *a priori* information / beliefs about the values of parameters is a key component of Bayesian analysis.
- This can be challenging!
- Prior information may be Data-based, or Non-data-based.
- Recall - we need to do this before we observe the *current sample* of data.
- One way to proceed is by using subjective "Betting Odds" .

Example

- Suppose we have 2 analysts wishing to construct a prior p.d.f. for a parameter, $\theta \in (-\infty, \infty)$.
- Decide to use a Normal prior.
- **A**: $p_A(\theta) = \frac{1}{20\sqrt{2\pi}} \exp\{-\frac{1}{2}\left(\frac{\theta-900}{20}\right)^2\}$
- **B**: $p_B(\theta) = \frac{1}{80\sqrt{2\pi}} \exp\{-\frac{1}{2}\left(\frac{\theta-800}{80}\right)^2\}$
- In the case of analyst **A**:

$$Pr.[860 < \theta < 940] = Pr.[-2 < Z < 2] = 0.95$$

Only if offered odds of at least 20:1 would she bet that θ differs from 900 by more than 40.

- In the case of analyst **B**:

$$Pr.[700 < \theta < 900] = Pr.[-1.25 < Z < 1.25] = 0.8$$

Only if offered odds of at least 5:1 would she bet that θ differs from 800 by more than 100.

Check the odds:

- x:1

Bet is \$1: Lose it if wrong

Collect \$$x$ (including stake) if correct

Prior expected payoff is

\$[(x - 1)Pr.(Correct) - (1)Pr.(Wrong)]

Least acceptable payoff is \$0.

- In the case of analyst **A**:

$0 = [(x - 1)(5/100) - (1)(95/100)]$

So, $x = 20$

(Need odds of at least 20:1 before she would bet)

- In the case of analyst **B**:

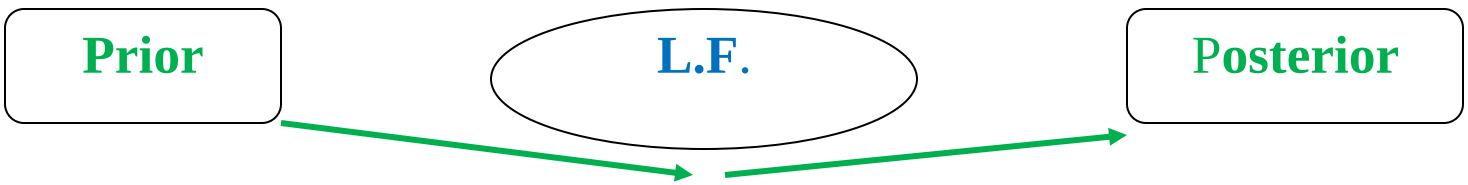
$0 = [(x - 1)(20/100) - (1)(80/100)]$

So, $x = 5$

(Need odds of at least 5:1 before she would bet)

(Natural) Conjugate Priors

- We've already seen examples of this.
- Advantage – computation simplicity - no need for nasty integration!
- Disadvantage – may be unrealistic in particular cases.
- Basic idea:



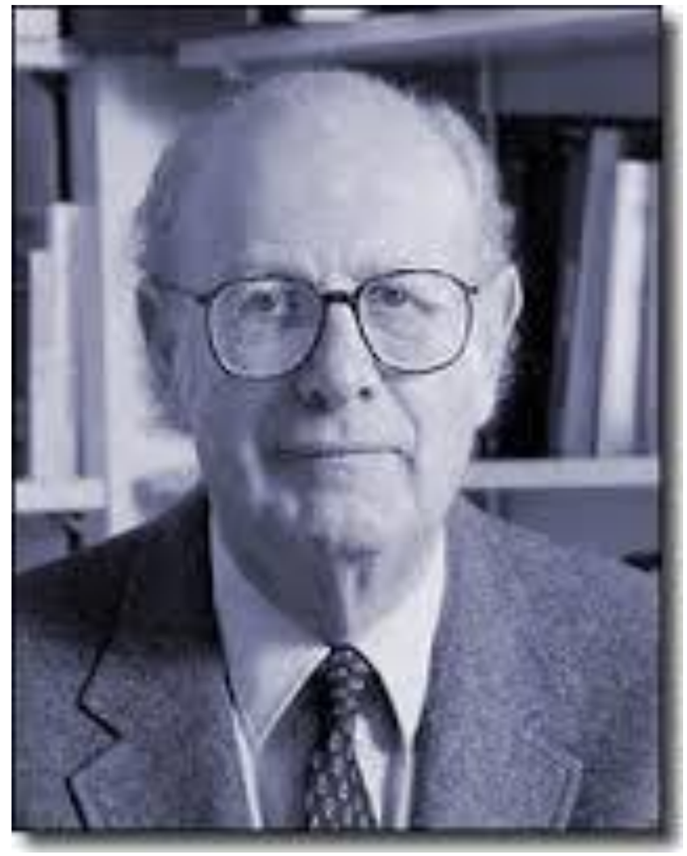
- Note: we *can't always construct a Conjugate Prior*.
- All distributions in the Exponential Family have conjugate priors.
- See CourseSpaces:

  (i) *A Compendium of Conjugate Priors*

  (ii) *Conjugate Prior Relationships*
- Also, https://en.wikipedia.org/wiki/Conjugate_prior

Introduced by Raiffa and Schlaifer, *Applied Statistical Decision Theory*.

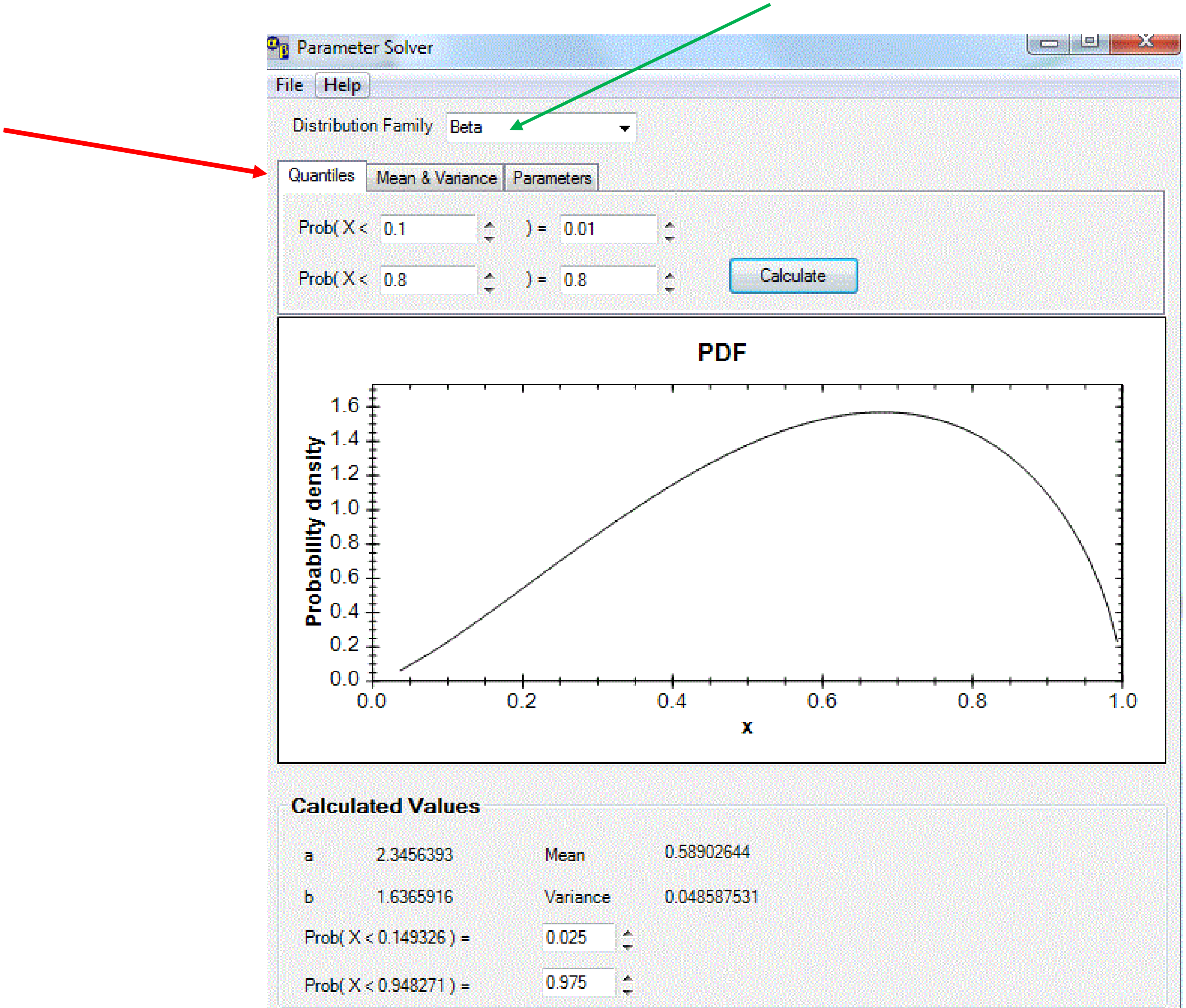
Robert Schlaifer (1919-1994)

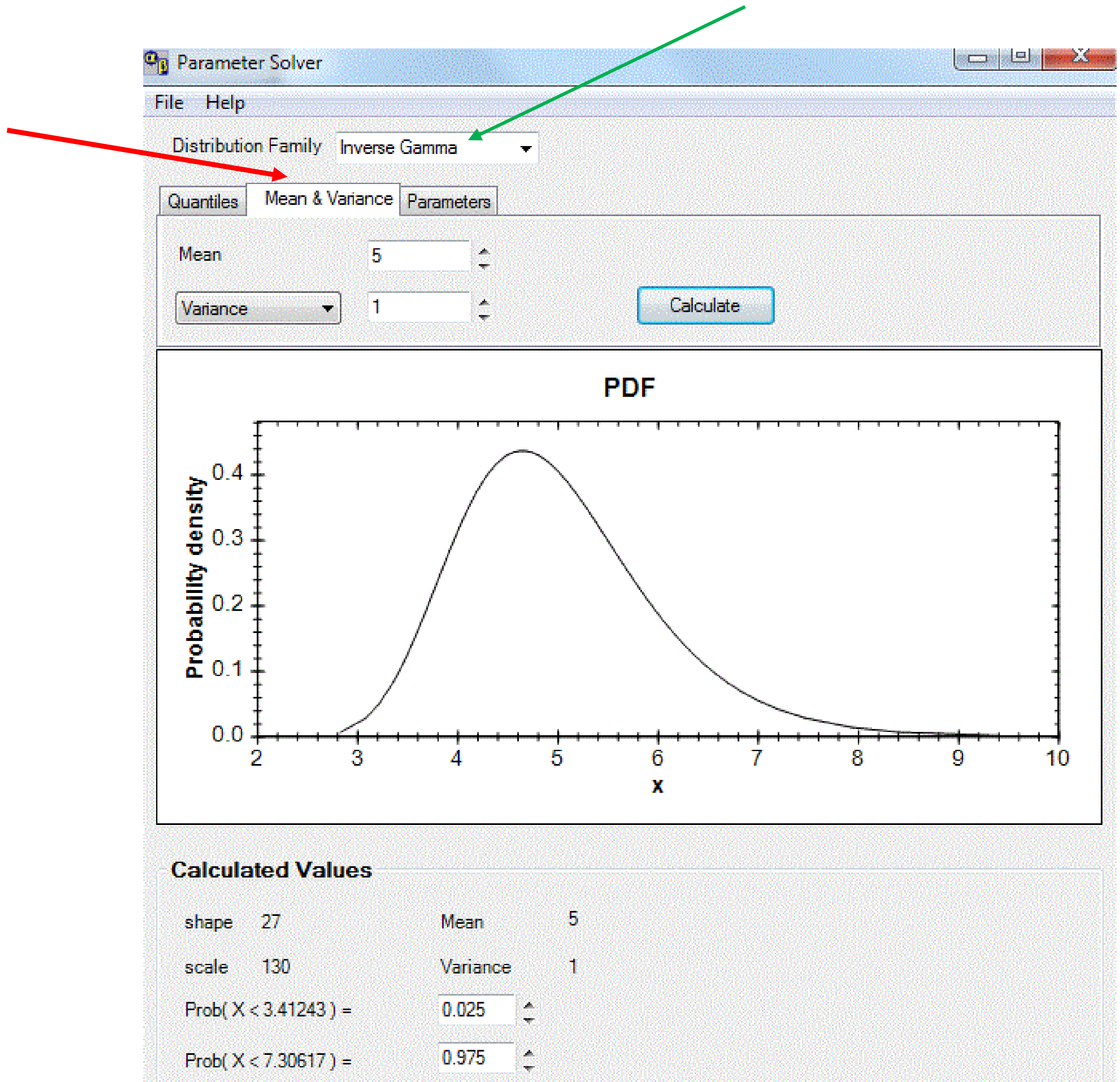


Howard Raiffa (Born, 1924)

John Cook's Material

- See his notes, “Determining Distribution Parameters From Quantiles”.
- “Parameter Solver” – free software.
- Both are on CourseSpaces.
- Bayesian statistics often requires eliciting prior probabilities from subject matter experts who are unfamiliar with statistics.
- Most people have an intuitive understanding of the mean of a probability distribution.
- Fewer people understand variance as well, particularly in the context of asymmetric distributions.
- Prior beliefs may be more accurately captured by asking experts for quantiles rather than for means and variances.


Parameter Solver
File
Help
Distribution Family
Beta
Quantiles
Mean & Variance
Parameters
Prob(X < 0.1) = 0.01
Prob(X < 0.8) = 0.8
Calculate
PDF
Probability density
x
Calculated Values
a 2.3456393
b 1.6365916
Mean 0.58902644
Variance 0.048587531
Prob(X < 0.149326) = 0.025
Prob(X < 0.948271) = 0.975


Parameter Solver
File Help
Distribution Family Inverse Gamma
Quantiles Mean & Variance Parameters
Mean 5
Variance 1
Calculate
PDF
Probability density
0.0 0.1 0.2 0.3 0.4
2 3 4 5 6 7 8 9 10
x
Calculated Values
shape 27
scale 130
Mean 5
Variance 1
Prob(X < 3.41243) = 0.025
Prob(X < 7.30617) = 0.975

Representing Prior Ignorance

- What if we approach a new inference problem without having any prior information about the parameters?
- Our Bayesian analysis can handle this situation.
- It's just a matter of formulating the Prior Distribution appropriately, and then we proceed as usual.
- Note - typically the results we obtain will differ (to some degree) from what we would obtain by using just the sample information (*i.e.*, just the Likelihood Function).

Jeffrey's Priors -

1. All we know is that $-\infty < \theta < \infty$

Assign $p(\theta) \propto constant$

i.e., $p(\theta)d\theta \propto d\theta$

- This prior is "diffuse" over the full real line.
- It is "improper" (it doesn't integrate to one):

$$\int_{-\infty}^{\infty} p(\theta)d\theta \propto \int_{-\infty}^{\infty} d\theta = [\theta]_{-\infty}^{\infty} = \infty$$

2. All we know is that $0 < \phi < \infty$

Assign $p(\phi) \propto 1/\phi$

i.e., $p(\phi)d\phi \propto d\phi/\phi$

- It is "improper" (it doesn't integrate to one):

$$\int_0^\infty p(\phi)d\phi \propto \int_0^\infty (\frac{1}{\phi})d\phi = [log|\phi|]_0^\infty = \infty$$

- This prior is "diffuse" over the positive real half-line.

- To see where this comes from:

Let $\theta = log(\phi)$; $-\infty < \theta < \infty$

Assign $p(\theta) \propto constant$

So, $p(\phi) = p(\theta)\left|\frac{d\theta}{d\phi}\right| \propto \phi^{-1}$; recall that $\phi > 0$

- Note that $p(\phi)$ is *invariant to power transformations*:

  Let $\varphi = \phi^m$, so that $\left(\frac{d\varphi}{d\phi}\right) = m\phi^{m-1}$

  So, $(d\varphi/\varphi) = m(d\phi/\phi) \propto (d\phi/\phi)$

  For instance, it doesn't matter if we work with σ or with σ^2.

- If $-\infty < \theta < 0$ then just re-parameterize, and define $\varphi = -\theta$.

- Note that even though both of the diffuse priors we've considered are "improper", when we apply Bayes' Theorem the posterior p.d.f. will be "proper" - it will integrate to one.

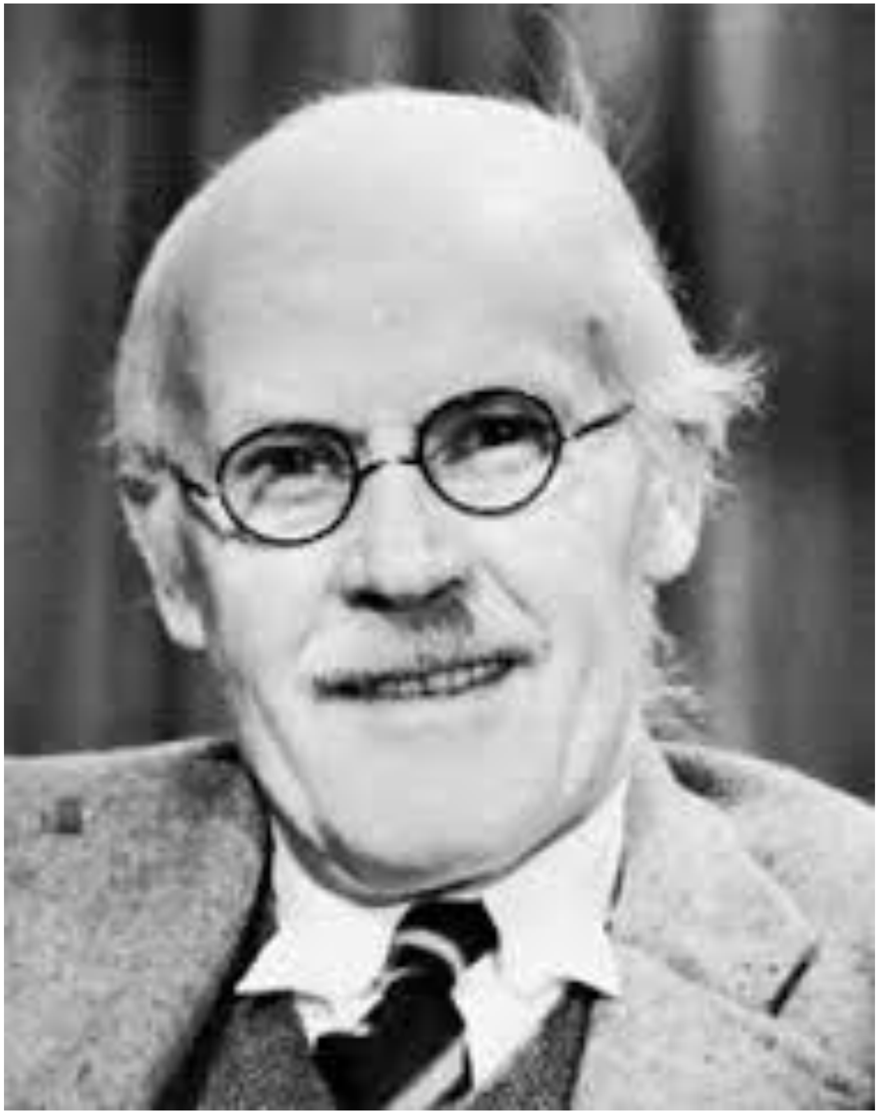
In what sense do these priors represent "total ignorance"?

All we know is that $-\infty < \theta < \infty$

- $p(\theta) \propto constant$
- $Pr.[a < \theta < b] \,/\, Pr.[c < \theta < d] = (0/0)$; *indeterminate*

All we know is that $0 < \phi < \infty$

- $p(\phi) \propto 1/\phi$
- $\int_0^a p(\phi)d\phi \propto \int_0^a d\phi/\phi = [log|\phi|]_0^a = \infty$
- $\int_a^\infty p(\phi)d\phi \propto \int_a^\infty d\phi/\phi = [log|\phi|]_a^\infty = \infty$
- $Pr.[0 < \theta < a] \,/\, Pr.[a < \theta < \infty] = (\infty/\infty)$; *indeterminate*



Sir Harold Jeffreys (1891-1989)

Jeffreys' Priors – More Formally

$$p(\boldsymbol{\theta}) \propto \sqrt{det.(I(\boldsymbol{\theta}))}$$ (*Invariant under re-parameterization*)

where $I(\boldsymbol{\theta})$ is Fisher's Information matrix, which can be written as

$$I(\boldsymbol{\theta}) = -E\left[\frac{\partial^2 logL}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]$$

$$= E\left[(\partial logL/\partial\boldsymbol{\theta})\left((\partial logL/\partial\boldsymbol{\theta})\right)'\right]$$ (OPG)

Example

$$p(y) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2\sigma^2}(y-\mu)^2\right] \quad ; \quad -\infty < \mu < \infty\, ; \;\; 0 < \sigma < \infty$$

$n = 1$

$$logL = -\log(\sigma) - \frac{1}{2}\log(2\pi) - \frac{1}{2\sigma^2}(y-\mu)^2$$

(i) $\partial logL/\partial\mu = (y-\mu)/\sigma^2$

$I(\mu) = E[\{(y-\mu)/\sigma^2\}^2] = 1/\sigma^2$

So,

$p(\mu) = \sqrt{1/\sigma^2} = 1/\sigma \propto constant$ (doesn't depend on μ)

(ii) $\partial logL/\partial\sigma = -1/\sigma + (y-\mu)^2/\sigma^3$

$$I(\sigma) = E\left[\left\{-\frac{1}{\sigma} + (y-\mu)^2/\sigma^3\right\}^2\right] = 2/\sigma^2$$

So,

$p(\sigma) = \sqrt{2/\sigma^2} = \sqrt{2}/\sigma \propto 1/\sigma$