Invariance Property

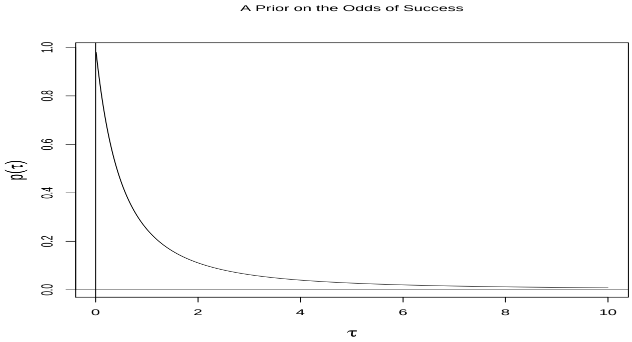
- A problem with the uniform prior is that its "lack of information" is **not invariant** under transformation.
- **Example 1 again**: Consider the **odds** of success $\tau = \dfrac{\theta}{1-\theta}$.
- Then if $p(\theta) = 1$, with the Jacobian

$$J = \left|\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\frac{\tau}{1+\tau}\right)\right| = \frac{1}{(1+\tau)^2},$$

$$\text{then } p(\tau) = \frac{1}{(1+\tau)^2}, \quad 0 < \tau < \infty:$$

Invariance Property

- Picture:



- This same prior is now an “informative” prior for the odds.
- (However, note that $P(0 < \tau < 1) = P(\tau > 1) = 0.5$.)

Jeffreys Prior

- Jeffreys (1961) developed a class of priors that were invariant under transformation.
- For a single parameter θ and data having joint density $f(\mathbf{x}|\theta)$, the Jeffreys prior

$$p_J(\theta) \propto \left[-E\left(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\ln f(\mathbf{x}|\theta)\right)\right]^{1/2} = [I(\theta)]^{1/2}$$

  (square root of Fisher information)
- For a parameter vector $\boldsymbol{\theta}$:

$$p_J(\boldsymbol{\theta}) \propto \left[E\left\{\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f(\mathbf{x}|\boldsymbol{\theta})\right]'\left[\frac{\partial}{\partial\boldsymbol{\theta}}\ln f(\mathbf{x}|\boldsymbol{\theta})\right]\right\}\right]^{1/2}$$

Jeffreys Prior

- **Example 1 yet again**: For $X_1, X_2, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta)$,

$$f(\mathbf{x}|\theta) = \binom{n}{y} \theta^y (1-\theta)^{n-y}, \ \ 0 \le \theta \le 1,$$

where $y = \sum_{i=1}^{n} x_i$.

$$\Rightarrow \ln f(\mathbf{x}|\theta) = \ln \binom{n}{y} + y \ln(\theta) + (n-y) \ln(1-\theta)$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta} \ln f(\mathbf{x}|\theta) = \frac{y}{\theta} - \frac{n-y}{1-\theta}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \ln f(\mathbf{x}|\theta) = -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}$$

Jeffreys Prior

$$\Rightarrow -E\left[\frac{d^2}{d\theta^2}\ln f(\mathbf{x}|\theta)\right] = \frac{n\theta}{\theta^2} + \frac{n - n\theta}{(1-\theta)^2} = \frac{n}{\theta} + \frac{n}{1-\theta}$$

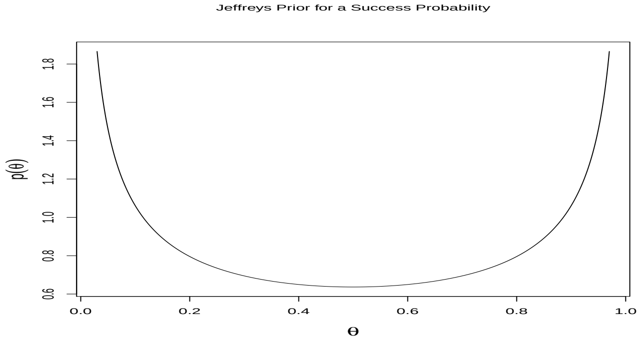
$$= \frac{n(1-\theta) + n\theta}{\theta(1-\theta)} = \frac{n}{\theta(1-\theta)}$$

$$\Rightarrow p_J(\theta) \propto \left[\frac{n}{\theta(1-\theta)}\right]^{1/2}$$

$$\Rightarrow p_J(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2} = \theta^{1/2-1}(1-\theta)^{1/2-1}$$

Jeffreys Prior

$\Rightarrow$ Jeffreys prior for θ is a Beta$(1/2, 1/2)$:
Picture:



Jeffreys Prior

- **Invariance**: If $p_J(\theta)$ is the Jeffreys prior for θ, for any transformation $\phi = g(\theta)$,

$$p_J(\theta) = p_J(\phi)\left|\frac{\mathrm{d}\phi}{\mathrm{d}\theta}\right|.$$

Other Noninformative Priors

- Other methods for noninformative priors include
  - Bernardo's reference prior, which seeks a prior that will maximize the discrepancy between the prior and the posterior and minimize the discrepancy between the likelihood and the posterior (a "dominant likelihood prior").
  - An improper prior, in which $\int_{\Theta} p(\theta) = \infty$.
  - A highly **diffuse** proper prior, e.g., for normal data with μ unknown, a $N(0, 1000000)$ prior for μ. (This is very close to the improper prior $p(\mu) \propto 1$.)

Informative Prior Forms

- Informative prior information is usually based on expert opinion or previous research about the parameter(s) of interest.

Power Priors

- Suppose we have access to **previous data $\mathbf{x}_0$** that is analogous to the data we will gather.
- Then our "power prior" could be

$$p(\theta|\mathbf{x}_0, a_0) \propto p(\theta)[L(\theta|\mathbf{x}_0)]^{a_0}$$

where $p(\theta)$ is an ordinary prior and $a_0 \in [0, 1]$ is an exponent measuring the influence of the previous data.

Power Priors

- As $a_0 \to 0$, the influence of the previous data is lessened.
- As $a_0 \to 1$, the influence of the previous data is strengthened.
- The posterior, given **our actual** data $\mathbf{x}$, is then

$$\pi(\theta|\mathbf{x}, \mathbf{x}_0, a_0) \propto p(\theta|\mathbf{x}_0, a_0) L(\theta|\mathbf{x})$$

- To avoid specifying a single a_0 value: We could put a, say, beta distribution $p(a_0)$ on a_0 and average over values of a_0 in $[0, 1]$:

$$p(\theta|\mathbf{x}_0) = \int_0^1 p(\theta)[L(\theta|\mathbf{x}_0)]^{a_0} p(a_0)\, \mathrm{d}a_0$$

Prior Elicitation

- A challenge is putting "expert opinion" into a form where it can be used as a prior distribution.
  Strategies:
- Requesting guesses for several quantiles (maybe $\{0.1, 0.25, 0.5, 0.75, 0.9\}$?) from a few experts.
- For a normal prior, note that a quantile $q(\alpha)$ is related to the z-value $\Phi^{-1}(\alpha)$ by:

$$q(\alpha) = \text{mean} + \Phi^{-1}(\alpha) \times (\text{std. dev.})$$

- Via regression on the provided $[q(\alpha), \Phi^{-1}(\alpha)]$ values, we can get estimates for the mean and standard deviation of the normal prior.

Prior Elicitation

- Another strategy asks the expert to provide a “predictive modal value” (most “likely” value) for the parameter.
- Then a rough 67% interval is requested from the expert.
- With a normal prior, the length of this interval is twice the prior standard deviation.
- For a prior on a Bernoulli probability, the “most likely” probability of success is often “clear”.