The Fisher Information measures the information in X about θ.

Let $X \sim \text{Binom}(n;\theta)$ Then

$$
\begin{aligned}
X \sim \text{Binom}(n;\theta) &= \binom{n}{x}\theta^x(1-\theta)^{n-x} \\
l(\theta;x) &= ln\frac{n}{x} + x\ln\theta + (n-x)\ln(1-\theta) \\
l'(\theta;x) &= \frac{x}{\theta} - \frac{n-x}{1-\theta} \\
l''(\theta;x) &= \frac{-x}{\theta^2} - \frac{n-x}{(1-\theta)^2} \\
I(\theta) &= \text{E}_x[-l''(\theta;x)] \\
&= \text{E}[\frac{x}{\theta^2} + \frac{n-x}{(1-\theta)^2}] \\
&= \frac{\text{E}[X]}{\theta^2} + \frac{n-\text{E}[X]}{(1-\theta)^2} \\
&= \frac{n\theta}{\theta^2} + \frac{n-n\theta}{(1-\theta)^2} \\
&= n(\frac{1}{\theta} + \frac{1}{1-\theta}) \\
&= n\frac{1}{\theta(1-\theta)}
\end{aligned}
$$

The Fisher information for the Binomial distribution is $n\frac{1}{\theta(1-\theta)}$.
For example, if $X \sim \text{Binom}(1, 0.5)$, $I(\theta) = 4$; if $X \sim \text{Binom}(1, 0.01)$, $I(\theta) = 101.01$.

Given $\mathcal{F} = \mathbb{P}(X \mid \theta)$, pick $\mathbb{P}(\phi)$ where $\phi = t(\theta)$ and t is 1-1 and smooth.

$$\mathbb{P}(X \mid \theta) \overset{\text{pick}}{\to} \mathbb{P}(\theta) \text{ and } \mathbb{P}(X \mid \phi) \overset{\text{pick}}{\to} \mathbb{P}(\phi)$$

But we want $\mathbb{P}(\theta)$ and $\mathbb{P}(\phi)$ to be related via change of variables.
Jeffrey's Prior: $\mathbb{P}(\theta) \propto \sqrt{I(\theta)}$

Let $X \sim \text{Binom}(n, \theta)$ Then

$$\begin{aligned}
\mathbb{P}(\theta) &\propto \sqrt{n(\frac{1}{\theta(1-\theta)})} \\
&\propto \frac{1}{\theta(1-\theta)} \\
&= \theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \\
&\propto \text{Beta}(\frac{1}{2}, \frac{1}{2}) \\
&= \underbrace{\frac{1}{B(\frac{1}{2}, \frac{1}{2})}}\pi\theta^{-\frac{1}{2}}(1-\theta)^{-\frac{1}{2}} \\
&= \frac{1}{\pi\sqrt{\theta(1-\theta)}}
\end{aligned}$$

This is the arcsin distribution. It is equidistant from Beta(0,0) and Beta(1,1). It is also called Jeffrey's prior (uninformative).

$$\mathbb{P}(X \mid \theta) \to \mathbb{P}(\theta) = \frac{1}{\pi\sqrt{\theta(1-\theta)}}$$

Recall that $R = t(\theta) = \frac{\theta}{1-\theta}$ and $\theta = t^{-1}(R) = \frac{R}{R+1}$.
Let $X \sim \text{Binom}(n, \theta)$. Then

$$\begin{aligned}
\mathbb{P}(X \mid R) &= \binom{n}{x}(\frac{R}{R+1})^x(\underbrace{1 - \frac{R}{R+1}}_{\frac{1}{R}+1})^{n-x} \\
&= \binom{n}{x}\frac{R^x}{(R+1)^n} \\
l(X; R) &= \ln\binom{n}{x} + x\ln R - n\ln(R+1) \\
l'(X; R) &= \frac{X}{R} - \frac{n}{R+1} \\
l''(X; R) &= -\frac{X}{R^2} + \frac{n}{(R+1)^2} \\
I(R) &= \text{E}[-l''(X; R)] = \text{E}[\frac{X}{R^2} - \frac{n}{(R+1)^2}] \\
&= \frac{\text{E}[X]}{R^2} - \frac{n}{(R+1)^2} \\
&= \frac{n\frac{R}{R+1}}{R^2} - \frac{n}{(R+1)^2} \\
&= n\Big(\frac{1}{R(R+1)} + \frac{1}{(R+1)^2}\Big) \\
&= n\frac{1}{R(R+1)^2}
\end{aligned}$$

Therefore
$$\mathbb{P}\left(R\right) \propto \sqrt{n}R(R+1)^2 \propto \frac{1}{\sqrt{R}}\frac{1}{R+1} \propto \frac{1}{\pi}\frac{1}{\sqrt{R}}\frac{1}{R+1} = \mathbb{P}\left(\phi\right)$$
By change of variables,
$$\begin{aligned}\mathbb{P}_R(R) &= \mathbb{P}_\theta((t^{-1}(R)))\Big|\frac{d}{dr}[t^{-1}(R)]\Big| \\ &= \frac{1}{\pi}(\frac{R}{R+1})^{-\frac{1}{2}}(\frac{1}{R+1})^{-\frac{1}{2}} \cdot \frac{1}{(R+1)^2} \\ &= \frac{1}{\pi}R^{-\frac{1}{2}}(R+1)\frac{1}{(R+1)^2} \\ &= \frac{1}{\pi}\frac{1}{\sqrt{R}}\frac{1}{R+1}\end{aligned}$$
General Case: Given $\mathbb{P}\left(X \mid \theta\right)$, $\mathbb{P}\left(X \mid \phi\right)$, and that
$$\begin{aligned}\mathbb{P}\left(\theta\right) &\propto \sqrt{I(\theta)} \\ \mathbb{P}\left(\phi\right) &\propto \sqrt{I(\phi)}\end{aligned}$$
Then
$$\begin{aligned}\mathbb{P}\left(\phi\right) &= \mathbb{P}_\theta(\underbrace{t^{-1}(\phi)}_{\theta})\Big|\frac{d}{d\phi}t^{-1}(\phi)\Big| \propto \sqrt{I(\phi)} \\ &= \mathbb{P}_\theta(\theta)\Big|\frac{d\theta}{d\phi} \\ &\propto \sqrt{I(\theta)}\Big|\frac{d\theta}{d\phi}\Big| \\ &= \sqrt{I(\theta)\frac{d\theta}{d\phi}\frac{d\theta}{d\phi}} \\ &= \sqrt{\mathrm{E}[s(\theta;X)^2]\frac{d\theta}{d\phi}\frac{d\theta}{d\phi}} \\ &= \sqrt{\mathrm{E}[\frac{dl}{d\theta}\frac{dl}{d\theta}\frac{d\theta}{d\phi}\frac{d\theta}{d\phi}]} \\ &= \sqrt{\mathrm{E}[(\frac{dl}{dt})^2]} \\ &= \sqrt{\mathrm{E}[s(\phi;X)^2]} \\ &= \sqrt{I(\phi)}\end{aligned}$$
A baseball player's true batting average is given as follows:
$$\hat{\theta} = BA := \frac{\#\text{ hits}}{\#\text{ at bats}} = \frac{x}{n} = \hat{\theta}_{\text{MLE}}$$
Say # of hits $\propto$ Binom(# bats, θ). For $n = 2$, if $x = 0$, then BA $= 0$. If $x = 1$, BA $= \frac{1}{2}$. If $x = 2$, BA $= 1$. This is absurd. Thus let's use $\theta \sim \text{Beta}(\alpha, \beta)$ to shrink. Fix a beta to the

prior data. Let's say $\hat{\alpha}_{\text{MLE}} = 78.7$ and $\hat{\beta}_{\text{MLE}} = 224.8$. Then $\hat{\alpha} + \hat{\beta} = 303.5$ which is strong. It also follows that $\hat{\theta}_{\text{MMSE}} = \frac{x+\alpha}{n+\alpha+\beta} = \frac{x+78.7}{n+303.5}$. For n large, use this estimation. This is called Empirical Bayes.
Steps

1. Get all data.
2. Fit prior to all data using MLE.
3. Use this fit's hyperparameters for inference.

Let $\mathcal{F}$ = Geometric. Then $X|\theta \sim (1-\theta)^x\theta$ where X is number of failures. $\text{Supp}[X] = \{0, 1, \dots\}$. $\Theta = (0,1)$ and $\text{E}[X] = \frac{1}{\theta} - 1$. If θ is large, then x is small; if θ is small, then x is large. Let's say $X_1 \sim \theta_1, \dots, X_n \sim \theta_n \overset{iid}{\sim} \text{Geom}(\theta)$. Then

$$\mathbb{P}(X \mid \theta) = \prod_{i=1}^{n} (1-\theta_i)^n \theta_i = (1-\theta)^{\sum x_i}\theta^n$$

Furthermore,

$$\begin{aligned}
\mathbb{P}(\theta \mid X) &\propto \mathbb{P}(X \mid \theta)\,\mathbb{P}(\theta) \\
&= \underbrace{(1-\theta)^{\sum x_i}\theta^n}_{\text{kernel of beta}}\,\mathbb{P}(\theta) \\
&\propto \theta^n (1-\theta)^{\sum x_i}\theta^{\alpha-1}(1-\theta)^{\beta-1} \\
&= \theta^{n+\alpha-1}(1-\theta)^{\sum x_i + \beta - 1} \\
&= \text{Beta}(n+\alpha, \sum x_i + \beta)
\end{aligned}$$

This is done using $\mathbb{P}(\theta) = \text{Beta}(\alpha, \beta)$. What we found here is that beta is also the conjugate prior for the geometric random variable.

If $X_1|\theta, \dots, X_n|\theta \overset{iid}{\sim} \text{Geom}(\theta)$ and $\theta \sim \text{Beta}(\overbrace{\alpha, \beta}^{\text{hyperparameters}})$, then

$$\theta | X_1, \dots, X_n \sim \text{Beta}(\underbrace{n+\alpha}_{\alpha'}, \underbrace{\sum x_i + \beta}_{\beta'})$$

Furthermore

$$\begin{aligned}
\hat{\theta}_{\text{MMSE}} &= \frac{n+\alpha}{n+\alpha+\sum x_i + \beta} \\
\hat{\theta}_{\text{MAE}} &= \text{qbeta}(0.5, n+\alpha, \sum x_i + \beta) \\
\hat{\theta}_{\text{MAP}} &= \frac{n+\alpha-1}{n+\alpha+\sum x_i + \beta - 2}
\end{aligned}$$

α = pseudo number of trials, β = seen total number of failures. If $\theta \sim \text{Beta}(0,0)$, Haldone, where $\alpha = 0$ and $\beta = 0$, this is complete ignorance. If $\theta \sim U(0,1) = \text{Beta}(1,1)$, Laplace, where $\alpha = 1$ and $\beta = 1$, this is indifference prior which gives no special preference. What's Jeffrey's prior?

$$\begin{aligned}
\mathcal{L}(\theta; X) &= (1-\theta)^{\sum x_i}\theta^n \\
l(\theta; X) &= \sum x_i \ln(1-\theta) + n\ln\theta \\
l'(\theta; X) &= -\frac{\sum x_i}{1-\theta} + \frac{n}{\theta} \\
l''(\theta; X) &= -\frac{\sum x_i}{(1-\theta)^2} - \frac{n}{\theta^2} \\
I(\theta) &= \mathrm{E}[-l''(\theta; X)] = \mathrm{E}\Big[\frac{\sum x_i}{(1-\theta)^2} + \frac{n}{\theta^2}\Big] \\
&= \frac{\mathrm{E}[x_i]}{(1-\theta)^2} + \frac{n}{\theta^2} \\
&= \frac{n\mathrm{E}[X]}{(1-\theta)^2} + \frac{n}{\theta^2} \\
&= n\Big(\frac{\frac{1}{\theta} - 1}{(1-\theta)^2} + \frac{1}{\theta^2}\Big) \\
&= n\Big(\frac{\frac{1-\theta}{\theta}}{(1-\theta)^2} + \frac{1}{\theta^2}\Big) \\
&= n(\frac{1}{\theta(1-\theta)} + \frac{1}{\theta^2}) \\
&= n(\frac{1}{\theta^2(1-\theta)})
\end{aligned}$$

Therefore

$$\mathbb{P}(\theta) \propto \sqrt{I(\theta)} = \sqrt{n\frac{1}{\theta^2(1-\theta)}} \propto \theta^{-1}(1-\theta)^{-\frac{1}{2}} \propto \mathrm{Beta}(0, \frac{1}{2})$$

Jeffrey's prior is $\theta \sim \mathrm{Beta}(0, \frac{1}{2})$, with $\alpha = 0$ and $\beta = \frac{1}{2}$. This is an improper prior and similar to Wilson's estimate.