\[ p(\theta) = U(0, 1) \]

Improper prior \( U(0, \infty) \)

Sometimes incoherent
$p(y_1 | y_{-1})$

$y_{-1} = (y_2, \ldots, y_n)$ (cross-validation)

$P(y_1 | y_{-1})$

$log$ score criterion
\[ \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 1 \end{bmatrix} \beta \\
\begin{bmatrix} 1 \\ 1 \end{bmatrix} \alpha \]

\[ \Rightarrow (\alpha + \beta) = 30 \]

prior like a data set

\[ \Rightarrow \text{prior sample size} \]

mean \( \frac{\alpha}{\alpha + \beta} = 0.15 \)

\[ \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 1 \end{bmatrix} \frac{(n-s)}{n} \]

\[ \Rightarrow n = 400 \text{ sample size} \]

data

mean \( \bar{y} = \frac{5}{n} = 0.18 \)

the prior is like a data set: imagine merging prior, sample data set, I feel merged data to the likelihood

\[ \begin{bmatrix} 0 \\ \vdots \\ (n-s) \end{bmatrix} (\beta + \frac{1}{n-s}) \]

machinery: equivalent to Bayesian analysis with this prior, this is sample.
\[ p(\theta | y) \]

Bayesian posterior

\[ \text{post} \]

\[ S^* = \sqrt{\frac{\alpha^* \beta^*}{(\alpha^* + \beta^*)^2 (\alpha^* + \beta^* + 1)}} \]

\[ \alpha^* = d + s \]
\[ \beta^* = n - s - \alpha^* - \beta^* \]

\[ = \sqrt{\frac{\alpha^*}{(\alpha^* + \beta^*)}} \left( \frac{\beta^*}{(\alpha^* + \beta^*)} \right) \left( \frac{1}{(\alpha^* + \beta^*)} \right) \]

\[ \hat{\theta}_B (1 - \hat{\theta}_B) \]

\[ n_B \leq \text{posterior sample size} \]

\[ \frac{\alpha^*}{\alpha^* + \beta^*} = \text{posterior mean} = \hat{\theta}_B = \text{Bayesian point estimate} \]

\[ \frac{1}{\alpha^* + \beta^* + \beta^*} = \text{Frequentist point estimate} \]

\[ \hat{\theta}_{\text{MLE}} = \frac{1}{\hat{\theta}} \]

\[ \text{SE}(\hat{\theta}_{\text{MLE}}) = \sqrt{\frac{\hat{\theta} (1 - \hat{\theta})}{n}} \]
\[ \text{repeated sampling distribution of } \bar{Y} \]

\[ \text{posterior distribution of } \theta \text{ given } \bar{Y} \]

\[ Y \text{ fixed} \]

\[ \text{random} \]

\[ F \]

\[ \text{squared if } n \text{ large and prior info small} \]

\[ \text{(Bernstein - Von Mises Theorem)} \]
\[ p(\theta) \]

An I case study

\[ \text{not possible in beta} \]

\[ 0 \quad 0.4 \quad \text{family} \]

\[ \begin{bmatrix} Y \\ X \end{bmatrix} \]

standard logistic regression model:

\[ n \quad (Y_i | p_i) \sim B(p_i) \]

\[ \log \left( \frac{p_i}{1 - p_i} \right) = \beta_0 + \beta_1 X_i \]

- natural parameterisation in Bernoulli likelihood
steps in Bayesian conjugate machinery:

1. identify joint sampling distribution for observables

\[ P(z_1, \ldots, z_n \mid \gamma_0, \ldots, \gamma_n) \]

2. think of 1 as f(\theta) of \theta for fixed \gamma = (\gamma_1, \ldots, \gamma_n) \Rightarrow \text{like } f(\theta) \mid \gamma_0 \]

3. think of \( l(\theta \mid y) \) as density in \( \theta \) & try to find another density \( p(\theta) \) st. \( l(\theta \mid y) \cdot p(\theta) \) has same \text{pdf} form as \( p(\theta) \); this is conj. prior for \( \theta \) (choose a member of this family)
4 use boys' theorem to do conj. updating
5 create prediction list for future data
\[ L(Y) = \sum_{k=0}^{\infty} \frac{(\gamma^k e^{-\gamma})}{k!} \left( \frac{e^{-\lambda}}{\lambda^k} \right)^{n_k} \]

\[ = \sum_{k=1}^{\infty} \frac{\gamma^k e^{-\gamma}}{k!} \left( \frac{e^{-\lambda}}{\lambda^k} \right)^{n_k} \]

S is evidently sufficient for \( J \)

in this model lots of suff. stat.
in any given problem, e.g. \( Y \) is itself suff. (but not helpful to notice this)
Another such stat is \((\eta, \sigma^2)\) from \(n \) dim. to \(2\) but \(s^2\) goes down from \(n \) to \(2\) \& is therefore a "better" suff. stat \((s^2 \in \mathbb{Q}\) minimal suff. stat because can't get any lower than \(1\) dimension with \(2\) parameter)

**Empirical Rule**

For almost any dist., if you start at mean \(\mu\) and

1. \(\mu\) either way, you'll capture about \(\frac{2}{3}\) (68%)
2. \(\pm \sigma\) most (95%)
3. \(\pm 2\sigma\) almost all (99.7%)
E.R. exactly followed by all Gaussian dist. & approx. true for almost all non-Gaussian dist. as well.
\[ V(y_{n+1}, y) = \left( \frac{x+5}{\beta + n} \right) \left( 1 + \frac{1}{\beta + n} \right) \]

\[
\begin{bmatrix}
  y_2 \\
  \vdots \\
  y_n
\end{bmatrix} = y_{n-1}
\]

\[ p(y, y_{n-1}) \]

Out each \( y_i \) one at a time in turn

\[ \text{histogram of } z \]