drop dead date for THT 2:

Thu 11 Feb in class

\[ \text{prior distribution parameter } \theta \]

\[ \text{if } (\theta | B) \sim \text{Beta}(\theta | \alpha, \beta) \]

\[ \text{then } p(\theta | x, B) \propto \theta^{d-1}(1-\theta)^{\beta-1} \]

\[ (d = 1, \beta = 1) \rightarrow c \]

primary mode

secondary mode

bimodal

multimodal
\[ p(\theta | y, \beta) = \frac{c \cdot p(\theta | y, \beta) \cdot L(\theta | y, \beta)}{\text{post.}} \]

\[ \text{post.} = c \cdot \text{prior \cdot likelihood} \]

\[ (\theta \alpha \beta \beta) \sim \text{Beta}(\alpha, \beta) \]

\[ L(\theta | y, \beta) \sim \text{Bernoulli}(\theta) \]

\[ i = 1, \ldots, n \]

\[ p(y_i | \theta, \beta) = \theta^{y_i} (1 - \theta)^{1 - y_i} \]

\[ L(\theta | y, \beta) = c \cdot p(y | \theta, \beta) \]

\[ \text{Fisher's def. of lik. f'n.} \]

\[ L(\theta | y, \beta) = c \cdot \prod_{i=1}^{n} p(y_i | \theta, \beta) \]

\[ = c \cdot \prod_{i=1}^{n} \theta^{y_i} (1 - \theta)^{1 - y_i} \]
\[ z = \sum_{i=1}^{n} \theta_i \]

\[ s = \sum_{i=1}^{n} y_i \]

\[ f(y | \theta, B) = c \theta^s (1 - \theta)^{n - s} \]

in this case \( f(y | \theta, B) \) depends on \( y \) only through \( s = \sum_{i=1}^{n} y_i \).

Fisher (1924)

Here \( s \) is said to be sufficient for \( \theta \) in the IID Bernoulli(\( \theta \)) sampling model.
Fisher's claim: Given your likelihood, if you've calculated a sufficient statistic, you can throw away the data vector y with no loss of useful information. In general, this is false, because you will generally have uncertainty about the "correct" sampling distribution \( p(y | \theta) \).

Total info in y diagnostic info about which \( \theta \) is correct. Yet \( \gamma = (1010101010... \) is also a sufficient statistic.
Fisher: "the (sampling dist.) 5
walks in the door with the
client" (this is false)

\[ P \rightarrow (0, 1, 1) \]

\[ M = \{ p(0 | B), p(1 | B) \} \]

\[ \ell (\theta | y | B) = \mathcal{C} \theta (1 - \theta) \]

\[ n = 400 \]
\[ s = 72 \]
as s, n t layer of underflow, so take logs: log likelihood function

\[
\ell(\theta \mid y, B) = \log p(y \mid \theta, B)
\]

\[
= \log p(y \mid \theta, B)
\]

\[
\ell(\theta \mid y, B) = c + \log \theta + (n-s) \log(1-\theta)
\]

What does \( \theta \) look like if \( \theta \sim N(\mu, \sigma^2) \)

\[
p(\theta) = c, \exp(-c_2(\theta - \mu)^2)
\]
whyBeta? post ~ c.prior lik

\[ p(\theta|y,B) \propto p(y|\theta,B) p(\theta|B) \]

\[ \alpha \sim p + n - 1 \]

\[ \beta \sim q + n - 1 \]

\[ T \sim \theta \]

post with Beta choice in Bernoulli sampling model, post & prior ( & likelihood ) have some mathematical form : people say Beta prior is conjugate to Bernoulli sampling dist.

Rao / F. (1955) Bayes knew this in 1766
simulate $f_{\theta}$

CDF

$\text{CDF}^{-1} = \text{quantile}$

density (or PMF)

\[
\begin{align*}
\theta | \gamma \beta & \sim \text{Beta}(\alpha, \beta) \\
\theta | \gamma \beta & \sim \text{Beta}(d + s, \beta + n - s) \\
\end{align*}
\]