Let's begin with the basic problem of estimating q from a sequence of Bernoulli trials

In this case, we have the probability model

$$S \sim \text{Binomial}(N, q)$$

We know that if we assume nothing about q, then $\hat{q} = S/N$

But what does this mean probabilistically?

try assuming

$$q \sim \text{Uniform}(0,1)$$

this distribution represents our prior knowledge of q. The key is that we are thinking of q as a random variable even before we see the data.

We can then think of our model for S as a conditional model

$$q \sim \text{Uniform}(0,1)$$
$$S|q \sim \text{Binomial}(N, q)$$

In Bayesian statistics, the goal is to compute $q|S$

Bayes' formula

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

allows us to "invert" conditional probability

proof

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{and} \quad P(A \cap B) = P(B|A)P(A)$$

Important: this holds for probability densities as well!

In the context of Bernoulli trials,

$$\underbrace{f(q \mid S)}_{\text{posterior}} = \frac{\overbrace{P(S \mid q)}^{\text{likelihood}}\ \overbrace{f(q)}^{\text{Prior}}}{\underbrace{P(S)}_{\text{evidence}}}$$

Note that in present context likelihood is only thing depending on q:

$$f(q \mid S) = \underbrace{A}_{\text{constant in } q} \times q^S (1-q)^{N-S}$$

Can calculate A by:

$$1 = \int_0^1 f(q \mid S)\, dq = A \int_0^1 q^S (1-q)^{N-S} = A \frac{(1+N-S)!\,(1+S)!}{(N+2)!}$$

$$\Longrightarrow A = \frac{(N+2)!}{(1+N-S)!\,(1+S)!}$$

Random Var w/ density

$$f(q) \propto q^{\overbrace{S}^{\alpha-1}} (1-q)^{\overbrace{N-S}^{\beta-1}}$$

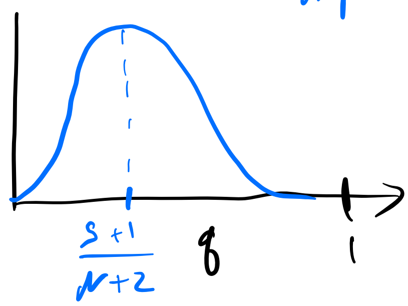
is called $\mathrm{Beta}(S+1, N-S+1)$

Laplace rule of succession

this is because w/ $q \sim \text{Uniform}(0,1)$

as priors $P(q=0)=P(q=1)=0$

$$\mathbb{E}[q|s] = \frac{s+1}{N+2} = \hat{q}_L$$



$$Var(q|s) = \frac{(1+s)(1-s+N)}{(2+N)^2(3+N)}$$

large n

$$\approx \frac{s}{N}\left(1-\frac{s}{N}\right)\frac{1}{N} = \frac{\hat{q}(1-\hat{q})}{N}$$

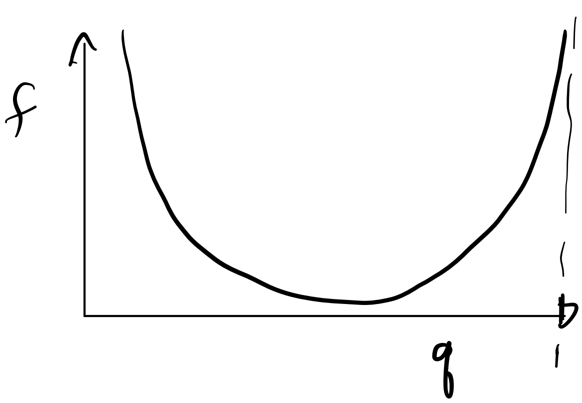
same as $se(\hat{q})$

$$\boxed{\text{the posterior is the Bayesian version of the sample distribution}}$$

Because it tells us about uncertainty in our inference, but unlike sample dist, it is interpreted as expression of our belief, not empirical distribution from repeating experiments

So where does $\hat{q} = s/N$ come from?

take improper prior $f(q) \propto \frac{1}{q(1-q)}$



But $\int_0^1 \frac{1}{q(1-q)}\,dq = \infty$

Weird!

Linear Regression

$\hat{\beta}$ from ridge regression

posterior mean of regression coefficients in Bayesian linear regression w/ normal priors