$X \sim \text{Binomial}(n, p)$ | $E(X) = n \cdot p$ | $V(X) = np(1-p)$

$$\psi_X(t) = [pe^t + (1-p)]^n \text{ for all } -\infty < t < \infty$$

$$SD(X) = \sqrt{np(1-p)}$$

Case Study: Castaneda v. Partida (1977)

Grand juries in the U.S. judicial system have catchment areas: everybody 18 & over living in the judicial district for that grand jury (& a few other minor restrictions)

Hidalgo County, Texas — extreme southern border of TX with Mexico

eligible pool was 79.1% Mexican-American

$2\frac{1}{2}$ yr period at issue in Supreme Court case: 220 people called to serve on grand juries, but only 100 of them were Mexican-American

Q: Prima facie case of discrimination?

Before this $2\frac{1}{2}$ yr period, let X be your prediction of # of Mexican-Americans among the 220 people

If no discrimination,

$X \sim \text{Binomial}(220, 0.791)$
$(X \mid T_1)$

T_1 = theory 1 = no discrimination

$E(X \mid T_1) = \overset{n}{(220)} \cdot \overset{p}{(0.791)} \doteq 174.0$

$SD(X \mid T_1) = \sqrt{n\,p\,(1-p)} \doteq 6.0$

$\boxed{Q:}$ If you were expecting 174 give or take ~~8~~ 6, would you be surprised to see 100?

$\boxed{A:}$ You'd be astonished

Frequentist statistical answer

$P(X \le 100 \mid T_1) = 8.0 \cdot 10^{-28}$ so T_1 looks ridiculous

Bayesian statistical answer

Need to compute $P(T_1 \mid X = 100)$, not the other way around (later)

Can show that $B(\alpha, \beta) = \dfrac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

(α, β) jointly control the shape of the Beta(α, β) dist.

(yuck)

$$\boxed{X \sim \text{Beta}(\alpha, \beta)} \qquad \psi_X(t) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha + r}{\alpha + \beta + r} \right) \frac{t^k}{k!}$$

$$E(X) = \frac{\alpha}{\alpha + \beta} \qquad V(X) = \left(\frac{\alpha}{\alpha + \beta} \right) \left(\frac{\beta}{\alpha + \beta} \right) \left(\frac{1}{\alpha + \beta + 1} \right)$$

~~Case Study~~ (Castaneda v. Partida continued)

$n = 220$ grand jurors chosen from eligible population of Hidalgo County, Texas, which was 79.1% Mexican-American, but only $s = 100$ selected grand jurors were Mexican-American; summarize the information in a Bayesian fashion about evidence of discrimination.

283

Data: $\mathcal{S}$ = # Mexican-Americans chosen in grand jury selection of $n = 220$ people

Unknown: θ = actual probability of an eligible Mexican-American person being chosen ($0 < \theta < 1$)

Sampling model: $(\mathcal{S} \mid \theta) \sim \text{Binomial}(n, \theta)$,

i.e., $f_{\mathcal{S}|\theta}(s \mid \theta) = P(\mathcal{S} = s \mid \theta) = \binom{n}{s} \theta^{s} (1-\theta)^{n-s} \cdot I(s = 0, 1, \dots, n)$

Bayesian approach: ① Information internal to data set about θ summarized by the likelihood (un-normalized) density, defined to be $\boxed{\ell(\theta \mid s) = c\, P(\mathcal{S} = s \mid \theta)}$,

c an arbitrary positive constant — just think of $P(\mathcal{S} = s \mid \theta)$ as a function of θ for fixed s.

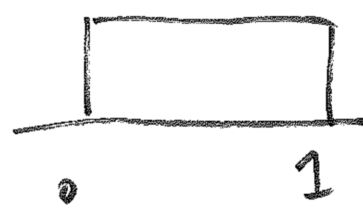
Here $\ell(\theta \mid s) = c \binom{n}{s} \theta^s (1-\theta)^{n-s}$ can be absorbed into c since does not depend on θ

$= c\, \theta^s (1-\theta)^{n-s}$

284

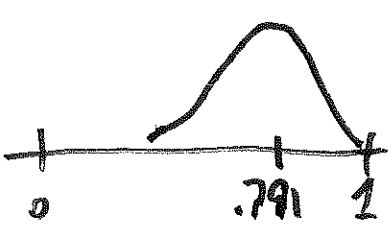
(2) Information external to dataset about θ summarized by the prior density $f_{\Theta}(\theta)$. Here are some possibilities for the prior, depending on your knowledge (information) base:

(a) neutral prior $\theta \sim \text{Uniform}(0,1)$



this Dist. embodies the information {θ could be anywhere between 0 and 1, with no value favored}

(b) Cut the district attorney some slack prior



this prior gives the DA the benefit of the doubt when you're uncertain about what prior to use, write down all the reasonable priors & do a sensitivity analysis (use each prior one by one & see if posterior answer is the same)

③ Combine internal & external information with Bayes' Theorem

$$f_{\Theta|S}(\theta|s) = c \cdot f_{\Theta}(\theta) \cdot \ell(\theta|s)$$

$$\begin{pmatrix}\text{posterior}\\\text{information}\end{pmatrix} = \begin{pmatrix}\text{normalizing}\\\text{constant}\end{pmatrix} \cdot \begin{pmatrix}\text{prior}\\\text{information}\end{pmatrix} \cdot \begin{pmatrix}\text{likelihood}\\\text{information}\end{pmatrix}$$

Here

$$f_{\Theta|S}(\theta|s) = c\, f_{\Theta}(\theta)\, \theta^{s}(1-\theta)^{n-s}$$

Rev. Bayes himself noticed back in 1760

that if you take $f_{\Theta}(\theta) = c\, \theta^{\text{power}} (1-\theta)^{\text{power}}$ then the product of 2 such densities is another such density, meaning that the posterior would have the same form as the prior & likelihood, making calculations easier. Moreover, we already know the name of densities that look like $\theta^{\text{power}} (1-\theta)^{\text{power}}$:

The Beta distributions: $X \sim \text{Beta}(\alpha, \beta)$ $(\alpha > 0, \beta > 0) \rightarrow$

$$f_X(x) = c\, \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

So let's take $f_{\Theta}(\theta) = c\, \theta^{\alpha-1} (1-\theta)^{\beta-1}$ in the law suit case study; then

$$f_{\Theta|S}(\theta|s) = c \left[\theta^{\alpha-1} (1-\theta)^{\beta-1} \right] \left[\theta^{s} (1-\theta)^{n-s} \right]$$

$$= c\ \theta^{(\alpha+s)-1}\,(1-\theta)^{(\beta+n-s)-1} = \text{Beta}(\alpha+s,\ \beta+n-s)$$

So the prior-to-posterior <u>updating</u> looks like this:

"Beta dist. is <u>conjugate</u> to the Binomial <u>likelihood</u>"

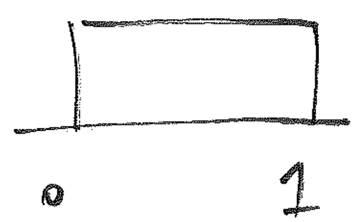
$$\left.\begin{array}{l}\theta \sim \text{Beta}(\alpha, \beta) \\ (S \mid \theta) \sim \text{Binomial}(n, \theta)\end{array}\right\} \rightarrow (\theta \mid s) \sim \text{Beta}(\alpha+s,\ \beta+n-s)$$

$s = 100$

$n = 220$

How choose (α, β)?

(a) Neutral prior



but $\text{Uniform}(0,1) = \theta^{1-1}(1-\theta)^{1-1}$

so $\theta \sim \text{Uniform}(0,1) \leftrightarrow \theta \sim \text{Beta}(1,1)$

(b) CHF DA slack prior

There's an extremely useful thing that happens with conjugate priors:

$$\begin{array}{cl} \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hline 1 \\ \vdots \\ 1 \end{array}\right] & \begin{array}{l} \beta \\ \\ \\ \alpha \end{array} \end{array} \Bigg\} \text{ prior effective sample size } (\alpha+\beta)$$

mean $\dfrac{\alpha}{\alpha+\beta}$

$$\begin{array}{cl} y & \\ \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ \hline 1 \\ \vdots \\ 1 \end{array}\right] & \begin{array}{l} n-s \\ \\ \\ s \end{array} \end{array} \Bigg\} \text{ data set sample size } n$$

mean $\bar{y} = \dfrac{s}{n}$

Beta prior distribution acts like a dataset with α 1s & β 0s with the property that if you do a Bayesian analysis with the Beta(α, β) prior and I do a frequentist analysis on the dataset with $(\alpha+s)$ 1s and $(\beta+n-s)$ 0s formed by merging the prior & sample data sets, we'll get the same results.

(b) cut the DA slack prior

mean of Beta(α, β) dist. is $\frac{\alpha}{\alpha+\beta}$; set this equal to 0.791

28

Suppose I want to put in prior information equivalent to a prior sample size $\frac{1}{10}$ as big as the data sample size (say); set

$(\alpha+\beta) = \frac{1}{10} n = 22$;

Solve: set $\begin{cases} \alpha = 17.4 \\ \beta = 4.6 \end{cases}$

$n = 220$

$s = 100$

likelihood is

$c\, \theta^{s} (1-\theta)^{n-s} = c\, \theta^{(s+1)-1} (1-\theta)^{(n-s+1)-1}$

$=$ Beta$(\underset{101}{s+1}, \underset{121}{n-s+1})$ dist

(a) Neutral prior: Beta$(1,1)$

prior sample size 2

posterior is Beta$(\underset{101}{\alpha+s}, \underset{121}{\beta+n-s})$

(same as likelihood)

250

(b) Cut DA slack prior

Beta(17.4, 4.6) prior
$\alpha \quad \beta$

posterior is $\text{Beta}(\underbrace{\alpha + s}_{117.4}, \underbrace{\beta + \overset{220}{n} - \overset{100}{s}}_{124.6})$

prior	posterior mean	posterior SD
neutral	0.455	0.0333
Cut DA slack	0.485	0.0321

posterior mean of θ is

$$\frac{\alpha + s}{\alpha + \beta + n}$$

Posterior SD is $\sqrt{\left(\frac{\alpha + s}{\alpha + \beta + n}\right)\left(\frac{\beta + n - s}{\alpha + \beta + n}\right)\left(\frac{1}{\alpha + \beta + n + 1}\right)}$

The no-discrimination rate of 0.791 is

$$\frac{0.791 - 0.455}{0.0333} = 10.1$$

posterior SDs away from posterior expectation under the neutral prior and

$$\frac{0.791 - 0.485}{0.0321} \doteq 9.5 \text{ posterior SDs}$$

away from posterior expectation under the cut-DA-slack prior; there was Q.E.D. discrimination

Multinomial Distributions (back to discrete)

You're contemplating a population that contains elements of $k \geq 2$ types (e.g., {Democrat, Republican, Libertarian, Independent, Green}). Suppose the proportion of elements of type i is $0 < p_i < 1$ with $\sum_{i=1}^{k} p_i = 1$; $\boldsymbol{p} = (p_1, \dots, p_k)$.