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

\( P(t) = [p e^t + (1-p)]^n \) for all \(-\infty < t < \infty\)

\[ \text{SD}(X) = \sqrt{np(1-p)} \]

**Case Study**

*Castanedo v. Partida (1977)*

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

*Hidalgo County, Texas*

*Extremely southern border of TX with Mexico*

Eligible pool was 79.1% Mexican-American

2½ 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 21 yr period, let \( X \) be your prediction of # of Mexican Americans among the 220 people if no discrimination.

\[ X \sim \text{Binomial} \left( 220, 0.791 \right) \]

\[ (8.1 T_1) \]

\[ E(\bar{x} \mid T_1) = (220)(0.791) = 174.0 \]

\[ \sigma(\bar{x} \mid T_1) = \sqrt{np(1-p)} = 6.0 \]

Q: If you were expecting 174 give or take 6, would you be surprised to see 100? A: You'd be astonished.

Frequentist statistical answer:

\[ P(\bar{x} = 100 \mid T_1) = 8 \times 10^{-28} \]

So \( T_1 \) looks ridiculous.

Bayesian statistical answer:

Need to compute \( P(T_1 \mid \bar{x} = 100) \), not the other way around (later).
Can show that \( B(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha + \beta)} \).

\( \alpha, \beta \) jointly control the shape of the Beta(\( \alpha, \beta \)) dist.

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

\[ \gamma(x) = 1 + \sum_{k=1}^{\infty} \frac{(k-1)!}{(\alpha + k)(\beta + k)} x^k \]

\[ E(x) = \frac{\alpha}{\alpha + \beta} \]

\[ V(x) = \frac{\alpha}{(\alpha + \beta)(\alpha + \beta + 1)} \]

Case Study: 220 jurors chosen from the eligible population of Hidalgo County (Castaneda v. Partida, Texas, which was 79.1% Mexican-American). The jurors were Mexican-American. Since \( n = 220 \) is large, and only 52 jurors were Mexican-American, we may assume that the sample is drawn from a sufficiently large population. We are interested in summarizing the information in a Bayesian fashion about evidence of discrimination.
Data \( S \) = # Mexican-Americans in jury selection of \( n = 220 \) people unknown

\[ \Theta = \text{actual probability of an eligible Mexican-American person being chosen} \] \((0 < \Theta < 1)\)

\[ (S | \Theta) \sim \text{Binomial}(n, \Theta), \]

\[ f(S | \Theta) = P(S = s | \Theta) = \binom{n}{s} \Theta^s (1-\Theta)^{n-s} \]

Bayesian Information internal approach to dataset about \( \Theta \) summarized by the likelihood (un-normalized) density, defined to be \[ L(\Theta | S) = c \cdot P(S = s | \Theta), \]

\( c \) an arbitrary positive constant - think of \( P(S = s | \Theta) \) as a function of \( \Theta \) for fixed \( s \).
Here \( E(θ | s) \) can be absorbed into \( c \) since \( c \) does not depend on \( θ \). Here are some possibilities for the prior, depending on your knowledge base:

1. Neutral prior \( θ \sim \text{Uniform}(0, 1) \)

2. 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 and do a sensitivity analysis (use each prior one by one & see if answer is the same)

3. Combine internal & external information

With Bayes' Theorem

Here

\[ f(\theta | s) = c \cdot f(\theta) \cdot f(s | \theta) \]

Posterior (information) = (normality) \cdot (prior information) \cdot (likelihood) 

\[ f(\theta | s) = c \cdot f(\theta) \cdot \theta^s (1-\theta)^n-s \]

Rev. Bayes himself noticed back in 1760
that if you take \( f(\theta) = c \theta^\alpha (1-\theta)^{\beta-1} \)
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 calculation easier.

Moreover, we already know the name of densities that look like \( \theta^\alpha (1-\theta)^{\beta-1} \):

\( X \sim \text{Beta}(\alpha, \beta) \quad (\alpha > 0, \beta > 0) \)  

Beta distributions \( f(x) = c \theta^{\alpha-1} (1-\theta)^{\beta-1} \)

So let's take \( f(\theta) = c \theta^\alpha (1-\theta)^{\beta-1} \)
in the lawsuit case study; then

\[
f_{\theta|s}(\theta|s) = c \left[ \theta^{\alpha-1} (1-\theta)^{\beta-1} \right] \left[ \theta^5 (1-\theta)^{5-5} \right]
\]
So the prior-to-posterior updating looks like this:

\[ \theta \sim \text{Beta}(\alpha, \beta) \]
\[ s = 100, \quad h = 220 \]
\[ \text{Uniform}(0, 1) = \theta^{-1} (1-\theta)^{-1} \]
\[ \text{So } \theta \sim \text{Uniform}(0, 1) \leftrightarrow \theta \sim \text{Beta}(1, 1) \]

(b) cut slack prior

There's an extremely useful thing that happens with conjugate priors:
The prior distribution acts like a dataset with \( \alpha+5 \) and \( \beta+5 \) with the property that if you do a Bayesian analysis with the Beta(\( \alpha, \beta \)) prior and I do a Frequentist analysis on the dataset with (\( \alpha+5 \)) Is and (\( \beta+\nu-5 \)) as formed by merging the prior & sample datasets, we'll get the same results.
Suppose I want to put in information equivalent to a prior sample size $\frac{1}{10}$ as big as the data sample size (507); set $(d + \beta) = \frac{1}{10} n = 22$.

Solve: \[ \begin{cases} d = 17.4 \\ \beta = 4.6 \end{cases} \]

$n = 220$ \hspace{1cm} likelihood is \[ \frac{S^s (1-\theta)^{n-s}}{c \theta^{s+1} (1-\theta)^{n-s+1}} \]

Neutral prior: $Beta(1,1)$

posterior is $Beta(d + s, \beta + n - s)$

(same as likelihood)

(3) Cut the slack prior
The non-discrimination rate of 0.791 is

$\text{Posterior SD} = \sqrt{0.791 \times \text{Prior SD}}$

<table>
<thead>
<tr>
<th>Prior</th>
<th>C+DA</th>
<th>Neutral</th>
<th>Posterior Beta (d+s)</th>
<th>Posterior Beta (d+s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slack</td>
<td>0.485</td>
<td>0.455</td>
<td>0.0872</td>
<td>0.0333</td>
</tr>
<tr>
<td>Neutral</td>
<td>0.033</td>
<td>0.033</td>
<td>0.1230</td>
<td>0.0333</td>
</tr>
<tr>
<td>Mean</td>
<td>0.485</td>
<td>0.455</td>
<td>0.1230</td>
<td>0.0333</td>
</tr>
</tbody>
</table>

DA

Beta: (17.4, 4.8) Prior

$\frac{d+s}{d+\beta+n}$
under the neutral prior and

\[ 0.791 - 0.485 = 9.5 \text{ posterior S.Ds} \]

\[ 0.0321 \]

away from posterior expectations under the cut-off slack prior; \[ \textit{P.E.D.} \]

discrimination

Multinomial You're contemplating a distribution population that contains (back to discrete) elements of \( k \geq 2 \) types (e.g., \{Democrat, Republican, Libertarian, Independent, Green\}). Suppose \( \pi \) proportion of elements of type \( i \) is \( \pi_i = \pi_i(\theta) \) with \( \sum_{i=1}^{k} \pi_i = 1; \theta = (\pi_1, \ldots, \pi_k) \).