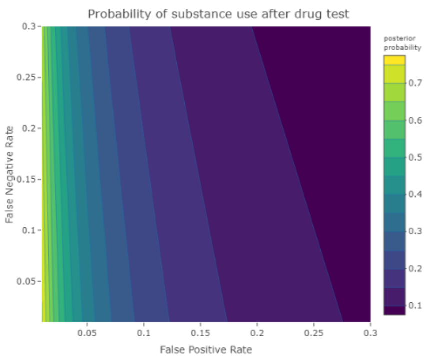
Problem Set 1

1.2 Consider the drug testing problem given in Example 1.1. Consider the false negative rate and the false positive rate of the drug test as two variables.

1. Construct a grid of hypothetical values for these two variables. At each point on the grid, compute the posterior probability of H_U, the hypothesis 'the subject uses the prohibited substance' given the prior on this hypothesis of P (Hu) = .03 and a positive test result. Use a graphical technique such as a contour plot or an image plot to summarize the results.



Obviously, at lower error rates, there is a higher posterior probability of actually using the substance after testing positive. Moreover, FPR tend to have more abrupt effect on the posterior probability of substance use than the FNR.

2. What values for the two error rates of the test give rise to a posterior probability on H_U that exceeds 0.5?

Solution:

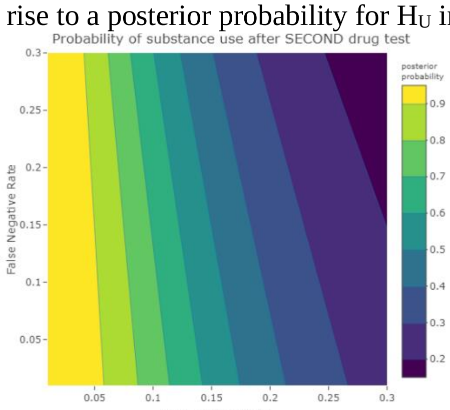
$$\frac{0.03(1-FNR)}{0.03(1-FN)+(1-0.03)(FP)} > 0.5$$

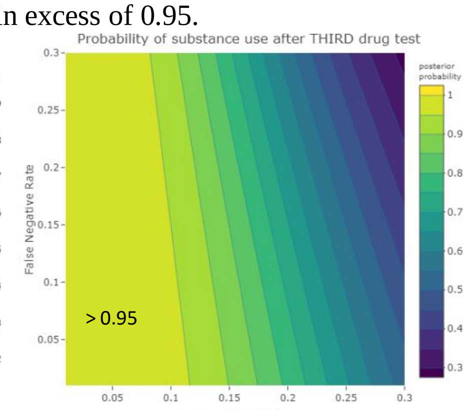
$$\rightarrow 0.06(1-FNR) > [0.03(1-FNR)+(0.97)(FPR)]$$

$$\rightarrow 0.03(1-FNR) > (0.97)(FPR) \rightarrow \boxed{(1-FNR) > \left(\frac{0.97}{0.03}\right)(FPR)}$$

Any **values for FNR and FPR that satisfy this inequality** will give a posterior probability > 0.5

3. Repeat this exercise, but now considering a run of 3 positive tests: what values of the test error rates give rise to a posterior probability for H_U in excess of 0.95.





After 3 positive tests, posterior probabilities have increased. Error rates corresponding to the **lightest yellow area** give > 0.95 posterior probability that the subject uses the substance.

Take note that the false positive rate must be approximately at most 0.12.

1.3 Suppose $p(\theta) \equiv \chi^2_2$. Compute a 50% highest density region for θ. Compare this region with the inter-quartile range of $p(\theta)$.

<u>*Solution:*</u>

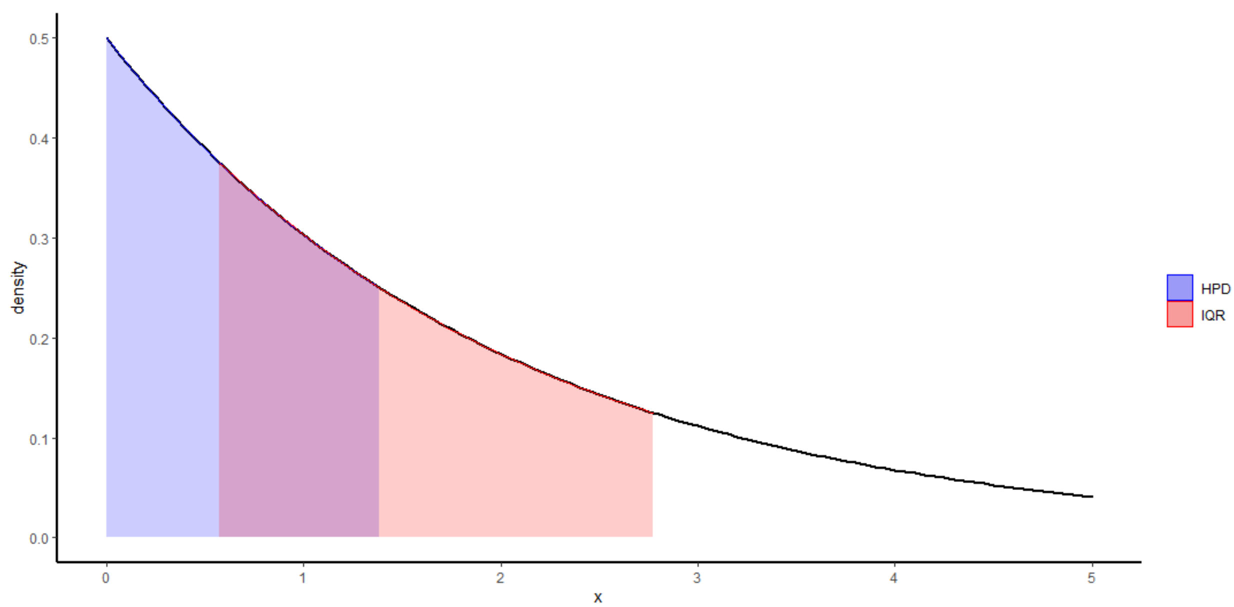
$p(\theta) \equiv \chi^2_2$ has a monotone decreasing function with mode = 0, hence, the region $[0, a]$ (blue shaded region in the figure below) will satisfy $P\,(\theta_1) \geq P\,(\theta_2), \forall\, \theta_1 \in C, \theta_2 \notin C.$

We already know that the lower bound is 0. Now, we just need to find the upper bound of the interval that will satisfy $P\,(\theta \in C) = 0.5$

$$P\,(\theta \leq a) = 0.5$$
$$\rightarrow 1 - e^{-\frac{a}{2}} = 0.5$$
$$\rightarrow -2\ln 0.5 = a$$
$$\rightarrow a \cong 1.3863$$

Hence, a <u>50% highest density region C for θ is **[0, 1.39]**</u>

On the other hand, <u>the interquartile range for χ^2_2 is given by **[0.5754, 2.7726].**</u>



While of course having the same area**,** <u>HPD interval has **shorter length** than the IQR.</u>

1.4 Consider a density $p(\theta)$. Under what conditions can a HDR for θ of content α be determined by simply noting the $(1-\alpha)/2$ and $1-(1-\alpha)/2$ quantiles of $p(\theta)$

Solution:

$\left(\frac{1-\alpha}{2}\right)$ and $\left(1-\frac{1-\alpha}{2}\right)$ quantiles are equal tails of a density function. For the following to hold:

$$p\left(\theta_{\frac{1-\alpha}{2}}\right) = p\left(\theta_{1-\frac{1-\alpha}{2}}\right)$$

the density must be **symmetric.** The following figure demonstrates the HPD for symmetric (Normal) and skewed (X^2).

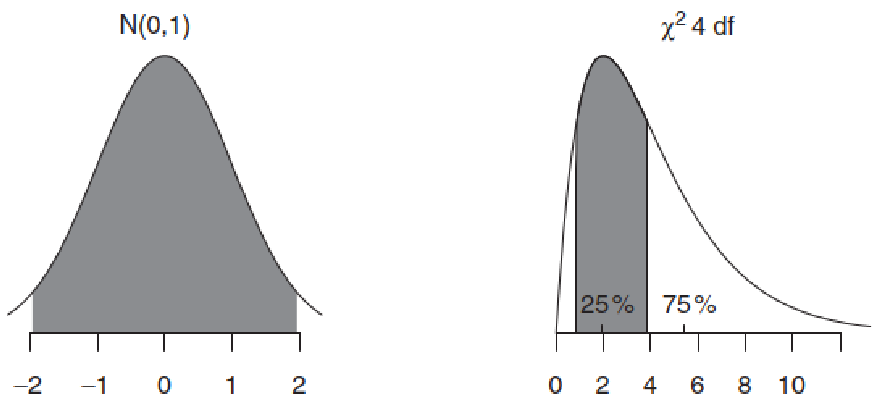


Figure 1.5 95 % HPD intervals for standard normal (left panel) and χ^2_4 densities.

Furthermore, if the distribution has multiple highest points (modes), the HDR will be a disjoint set or not unique for some cases, hence $\boldsymbol{p(\theta)}$ must be **unimodal** so that HDR is contained within $(1-\alpha)/2$ and $1-(1-\alpha)/2$ quantiles.

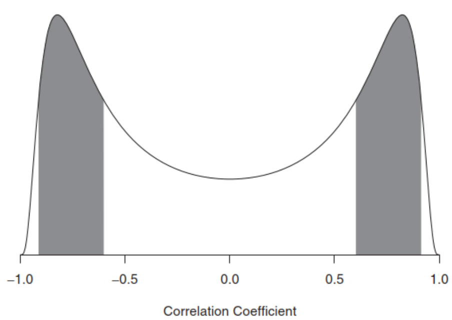


Figure 1.6 Bimodal posterior density for a correlation coefficient, and 50 % HPD.

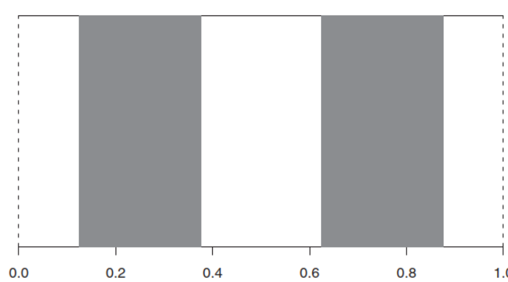


Figure 1.7 Uniform density and (non-unique) 25 % HPDs.

Therefore, to compute HDR by obtaining the $\frac{1-\alpha}{2}$ and $1-\frac{1-\alpha}{2}$ quantiles of $p(\theta)$, the density must be <u>both **symmetric** (i.e. mean = median = mode) and **unimodal.**</u>

1.6 A poll of 500 adults in the United States taken in the Spring of 2008 finds that just 29% of respondents approve of the way that George W. Bush is handling his job as president.

1.6.1 Report the posterior probabilities of H_0: $\theta > 0.33$ and H_1: $\theta < 0.33$. The threshold $\theta = 0.33$ has some politically interest, say, if we assume that (up to a rough approximation) the electorate is evenly partitioned into Democrat, Independent, and Republican identifiers.

Solution:

Let θ be the unknown population proportion of the respondents who approve George Bush as president. The question is if less than 1/3 of the population approves of him, assuming Democrats, Republican, and Independent supporters have equal number from the population.

The estimate for the proportion is $\hat{\theta} = 0.29$

Data is sampled from 500 independent Bernoulli trials, i.e. follows $Binomial(n = 500, p = \theta)$. However, with this large sample, the likelihood can be approximated by a Normal Distribution, with mean = $\hat{\theta}$ and standard deviation computed as the following:

$$se(\hat{\theta}) = \sqrt{\frac{\hat{\theta}\,(1-\hat{\theta}\,)}{n}} = \sqrt{\frac{0.29(1-0.29)}{500}} \approx 0.02$$

Therefore, the **likelihood of the data** has approximate distribution,

$$p(\boldsymbol{y}|\theta) \equiv Normal(0.29, 0.02^2)$$

We have no prior information to the analysis, so the posterior density has the same shape as the likelihood, i.e. $p(\theta|\boldsymbol{y}) \propto p(\boldsymbol{y}|\theta)$. Thus, $p(\theta|\boldsymbol{y}) \approx Normal(0.29, 0.02^2)$.

So, we have the posterior probabilities for H_0: $\theta > 0.33$ and H_1: $\theta < 0.33$

$$P(H_0|\boldsymbol{y}) = P(\theta > 0.33|\boldsymbol{y}) = \int_{0.33}^{\infty} p(\theta|\boldsymbol{y})\,d\theta = 1 - \Phi\left(\frac{0.33-0.29}{0.02}\right) \approx 0.0244$$

$$P(H_1|\boldsymbol{y}) = P(\theta < 0.33|\boldsymbol{y}) = \int_{-\infty}^{0.3} p(\theta|\boldsymbol{y})\,d\theta = \Phi\left(\frac{0.33-0.29}{0.02}\right) \approx 0.9756$$

<u>This shows that $\boldsymbol{H_1}$ **is more plausible than** $\boldsymbol{H_0}$ given the data.</u>

1.6.2 Report a Bayes factor for H_0 vs H_1. Comment briefly on your finding.

Solution

Using posterior and prior odds:

We have no prior knowledge about θ, we just assume $\theta \in [0,1]$. So, given the hypotheses $H_0: 0.33 < \theta < 1$ and $H_1: 0 < \theta < 0.33$, the corresponding prior probabilities of these hypotheses are 2/3 and 1/3.

Posterior probabilities were computed in part (1). Now plugging in the values:

$$B_{10} = \left\{\frac{P(H_1|\boldsymbol{y})}{P(H_0|\boldsymbol{y})}\right\} \Big/ \left\{\frac{P(H_1)}{P(H_0)}\right\} = \left\{\frac{0.9756}{0.0244}\right\} \Big/ \left\{\frac{1/3}{2/3}\right\} = \boxed{\mathbf{80.1216}}$$

Using marginal likelihood

The hypotheses $H_0: 0.33 < \theta < 1$ and $H_1: 0 < \theta < 0.33$ generate the prior distributions for θ.

$$p_0(\theta_0) \equiv Uniform(0.33, 1) = \begin{cases} 1.5 & \text{if } 0.33 \le \theta_0 \le 1 \\ 0 & \text{otherwise} \end{cases}$$

$$p_1(\theta_1) \equiv Uniform(0, 0.33) = \begin{cases} 3 & \text{if } 0 \le \theta_0 < 0.33 \\ 0 & \text{otherwise} \end{cases}$$

Obtaining the marginal likelihood for the two hypotheses:

$$p(\boldsymbol{y}|H_0) = \int_{0.33}^{1} p(\boldsymbol{y}|H_0, \theta_0)\ p_0(\theta_0) d\theta_0 = 1.5 \int_{0.33}^{1} p(\boldsymbol{y}|H_0, \theta_0)\ d\theta_0$$

$$= 1.5 \left[\Phi\left(\frac{1-0.29}{0.02}\right) - \Phi\left(\frac{1-0.29}{0.02}\right)\right] \approx 1.5(0.0244)$$

$$\approx 0.0365$$

$$p(\boldsymbol{y}|H_1) = \int_{0}^{0.33} p(\boldsymbol{y}|H_1, \theta_1)\ p_1(\theta_1) d\theta_1 = 3 \int_{0.33}^{1} p(\boldsymbol{y}|H_1, \theta_1)\ d\theta_1$$

$$= 3 \left[\Phi\left(\frac{0.33-0.29}{0.02}\right) - \Phi\left(\frac{0-0.29}{0.02}\right)\right] \approx 3(0.9756)$$

$$\approx 2.9269$$

Now plugging in the likelihood values for the Bayes factor:

$$B_{10} = \frac{p(\boldsymbol{y}|H_1)}{p(\boldsymbol{y}|H_0)} = \frac{2.9269}{0.0365} = \boxed{\mathbf{80.1216}}$$

This Bayes factor indicates that the **data strongly favor $\boldsymbol{H_1}$ over $\boldsymbol{H_0}$**

1.6.3 Contrast how a frequentist approach would distinguish between these two hypotheses.

<u>*Solution:*</u>

A question a frequentist would ask is *"How frequently would I observe a result at least as extreme as the one obtained if H_0 is true?"*

For the frequentist approach, **probabilities** of the two hypotheses **will not be computed**, but instead, we determine if we can gather the same (or more extreme away from H_0) data as the observed result (29% approval), if we assume that H_0 (approval > 33%) is true.

From this, if evidence shows that **observing the given dataset is an extremely rare event under the assumptions of H_0,** then we decide that H_0 must not be true, and we favor H_a.

So, we set the hypotheses H_0: $\theta = 0.33$ *and* H_1: $\theta < 0.33$ and compute the one-tailed p-value:

$$P(\hat{\theta} \leq 0.29 \,|\theta = 0.33) = \Phi\left(\frac{0.29 - 0.33}{0.021}\right) = 0.0284$$

This p-value suggests that observing the proportion 0.29 from the survey is an extremely rare event if we assume that the 33% of the population approve George Bush. Therefore, we reject this assumption in favor of the alternative that less than 33% approve of George Bush.

Problem Set 2

2.9 We are interested in the quantity $\boldsymbol{\delta = \theta_1 - \theta_0}$, where $\boldsymbol{\theta_j} \in [\mathbf{0}, \mathbf{1}]$, and so $\boldsymbol{\delta} \in [-\mathbf{1}, \mathbf{1}]$. Suppose we have a uniform prior on $\boldsymbol{\delta}$, i.e., $\boldsymbol{\delta} \sim \mathbf{Unif}(-\mathbf{1}, \mathbf{1})$.

1. Is this prior on δ sufficient to induce prior densities over both θ_j?

   No. Convolution requires at least one of the θ_0 and θ_1 to have a given distribution.

2. Suppose $\theta_0 \sim$ Unif(0, 1) and $\delta \sim$ Unif(−1, 1).
   What is $p(\theta_1)$, the density of $\theta_1 = \theta_0 + \delta$?
   Indeed, with this setup, is it the case that $p(\theta_1) = 0, \forall\, \theta_1 \notin [0, 1]$?

   Solution: Using convolution, we have the following:

$$p(\theta_1) = \int_0^1 p_{\Theta_0,\Delta}(\theta_0, \theta_1 - \theta_0)d\theta_0$$

assuming independence, we get

$$= \int_0^1 p_{\Theta_0}(\theta_0)p_{\Delta}(\theta_1 - \theta_0)d\theta_0$$

$$= \int_0^1 \theta_0\, I_{[0,1]}(\theta_0)\frac{(\theta_1 - \theta_0) + 1}{1 + 1} I_{[-1,1]}(\theta_1 - \theta_0)d\theta_0$$

$$= \int_0^1 \frac{1}{2}(\theta_0\,\theta_1 - \theta_0^2 + \theta_0\,)I_{[0,1]}(\theta_0)I_{[\theta_1 - 1, \theta_1 + 1]}(\theta_0)d\theta_0$$

$$= \int_0^1 \frac{1}{2}(\theta_0\,\theta_1 - \theta_0^2 + \theta_0\,)I_{[0,1]\cap[\theta_1 - 1, \theta_1 + 1]}(\theta_0)d\theta_0$$

$$= \begin{cases} 0 & \theta_1 \le -1 \\ \displaystyle\int_0^{\theta_1 + 1} \frac{1}{2}(\theta_0\,\theta_1 - \theta_0^2 + \theta_0\,)d\theta_0 & -1 < \theta_1 \le 0 \\ \displaystyle\int_0^1 \frac{1}{2}(\theta_0\,\theta_1 - \theta_0^2 + \theta_0\,)d\theta_0 & 0 < \theta_1 \le 1 \\ \displaystyle\int_{\theta_1 - 1}^1 \frac{1}{2}(\theta_0\,\theta_1 - \theta_0^2 + \theta_0\,)d\theta_0 & 1 < \theta_1 \le 2 \\ 0 & \theta_1 > 2 \end{cases}$$

Aside: solving the indefinite integral

$$\int \frac{1}{2}(\theta_0\,\theta_1 - \theta_0^2 + \theta_0\,)d\theta_0 = \frac{1}{2}\left(\frac{\theta_0^2\theta_1}{2} - \frac{\theta_0^3}{3} + \frac{\theta_0^2}{2}\right) + C = \theta_0^2\left(\frac{3\theta_1 - 2\theta_0 + 3}{12}\right) + C$$

Going back:

$$p(\theta_1) = \begin{cases} \theta_0^2 \left(\dfrac{3\theta_1 - 2\theta_0 + 3}{12}\right)\Bigg|_{\theta_0=0}^{\theta_0=\theta_1+1} & -1 < \theta_1 \leq 0 \\ \theta_0^2 \left(\dfrac{3\theta_1 - 2\theta_0 + 3}{12}\right)\Bigg|_{\theta_0=0}^{\theta_0=1} & 0 < \theta_1 \leq 1 \\ \theta_0^2 \left(\dfrac{3\theta_1 - 2\theta_0 + 3}{12}\right)\Bigg|_{\theta_0=\theta_1-1}^{\theta_0=1} & 1 < \theta_1 \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} (\theta_1 + 1)^2 \left(\dfrac{3\theta_1 - 2(\theta_1 + 1) + 3}{12}\right) & -1 < \theta_1 \leq 0 \\ 1 \left(\dfrac{3\theta_1 - 2 + 3}{12}\right) & 0 < \theta_1 \leq 1 \\ 1 \left(\dfrac{3\theta_1 - 2 + 3}{12}\right) - (\theta_1 - 1)^2 \left(\dfrac{3\theta_1 - 2(\theta_1 - 1) + 3}{12}\right) & 1 < \theta_1 \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \dfrac{(\theta_1 + 1)^3}{12} & -1 < \theta_1 \leq 0 \\ \left(\dfrac{3\theta_1 - 1}{12}\right) & 0 < \theta_1 \leq 1 \\ \left(\dfrac{3\theta_1 - 1}{12}\right) - (\theta_1 - 1)^2 \left(\dfrac{\theta_1 + 5}{12}\right) & 1 < \theta_1 \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$p(\theta_1) = \begin{cases} \dfrac{(\theta_1 + 1)^3}{12} & -1 < \theta_1 \leq 0 \\ \left(\dfrac{3\theta_1 + 1}{12}\right) & 0 < \theta_1 \leq 1 \\ \left(\dfrac{3\theta_1 + 1}{12}\right) - (\theta_1 - 1)^2 \left(\dfrac{\theta_1 + 5}{12}\right) & 1 < \theta_1 \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

■

Furthermore, the case that $p(\theta_1) = 0, \forall\ \theta_1 \notin [0,1]$ is not true because $p(\theta_1)$ is in fact defined within $(-1,2)$.

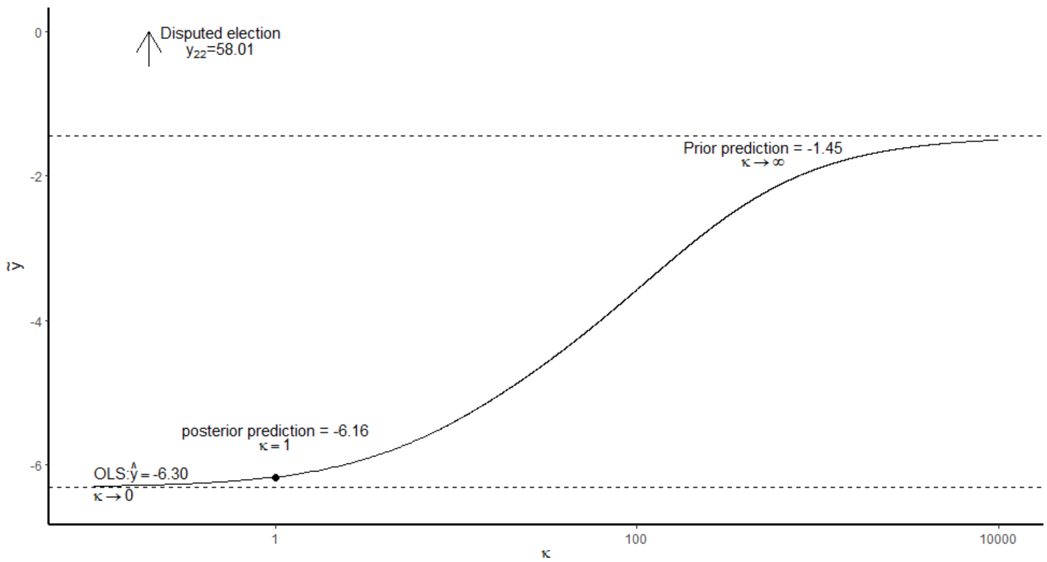
2.18 The data used to examine irregularities in Pennsylvania state senate elections (Example 2.15) appears as part of the author's R package, `pscl`. In the `pscl` package, the data frame is called `absentee`. The disputed election is the last election in the data set.

- Replicate the analysis in Example 2.15, over a range of priors.
- What kind of prior beliefs would one have to hold in order to find that the disputed election is not particularly unusual?
- Perform a sensitivity analysis of the sort presented graphically in Figures 2.16, but where the **output of interest is the posterior predictive density** for the disputed election result.

Solution:

For the sensitivity analysis of the posterior predictive density, we compute the following at different values of κ

$$\mathbf{b}_1 = \left(\kappa\mathbf{B}_0^{-1} + \mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\left(\kappa\mathbf{B}_0^{-1}\mathbf{b}_0 + \mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\boldsymbol{\beta}}\right)^{-1}$$
$$\tilde{y} = \mathbf{b}_1^{\mathrm{T}}\mathbf{X}_{22}$$

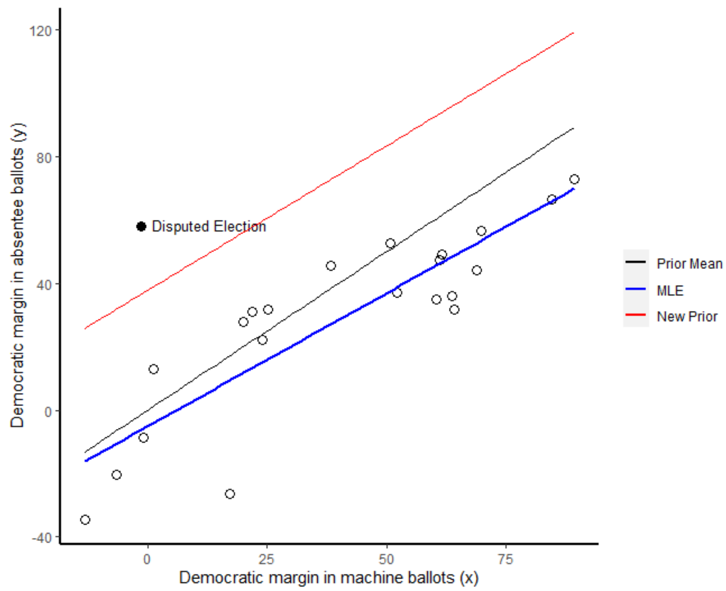


	Coefficients	Prediction (given $x_{22} = -1.45$)
Prior ($\kappa \to \infty$)	$\mathbf{b}_0 = [0 \quad 1]^{\mathrm{T}}$	$\tilde{y} = -1.45$
Posterior ($\kappa = 1$)	$\mathbf{b}_1 = [-4.95 \quad 0.84]^{\mathrm{T}}$	$\tilde{y} = -6.16$
OLS/Improper prior ($\kappa \to 0$)	$\hat{\boldsymbol{\beta}} = [-5.08 \quad 0.84]^{\mathrm{T}}$	$\hat{y} = -6.30$
Disputed election result		$y_{22} = 58.01$

For the disputed election to be determined as not unusual (i.e. no fraud), the **prior must be extremely biased towards the actual value of y**, and **κ must be extremely large** so that the new information will have little to no effect on the posterior.

Suppose the following custom prior

$$\begin{cases} \boldsymbol{\beta}|\sigma^2 \sim Normal\left(\mathbf{b}_0 = \begin{bmatrix} 40 \\ 1 \end{bmatrix}, \mathbf{B}_0 = \sigma^2 \kappa^{-1} \begin{bmatrix} 2.25 & 0 \\ 0 & 0.0022 \end{bmatrix}\right) \\ \sigma \sim InverseGamma\left(v_0 = \dfrac{6.2}{2}, \quad v_0 \sigma_0^2 = \dfrac{6.2}{2} 47.07\right) \end{cases}$$

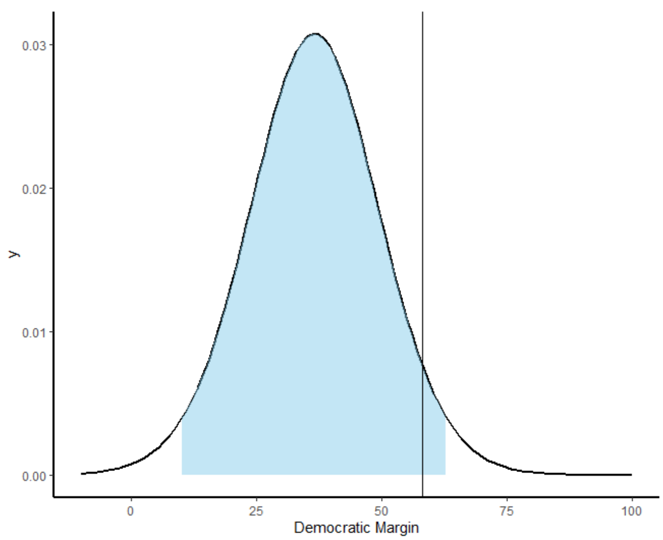


set $\kappa = 1000$ to obtain a precise posterior that overwhelms and almost disregards the new data

$$\boldsymbol{\beta}|\sigma^2, \mathbf{y}, \mathbf{X} \sim Normal\left(\mathbf{b}_1 = \begin{bmatrix} 57 \\ 0.87 \end{bmatrix}, \mathbf{B}_0 = \sigma^2 \begin{bmatrix} 0.0022 & -0.0000 \\ -0.0000 & 0.0000 \end{bmatrix}\right)$$

Then the posterior predictive density for $\tilde{y}|\, \tilde{x} = -1.45$ is:

$$\tilde{y}|\, \tilde{x} = -1.45 \sim t_{27.2}(location = \ 55.71, squared\ scale = 164.96\,)$$



With the new custom prior and small variance, the disputed election $y_{22} = 58.01$ will seem not unusual.

2.19 Prove Proposition 2.9. That is, suppose $y_i \overset{\text{iid}}{\sim} N(\mu,\sigma^2)$ with μ and σ^2 unknown, and $i = 1,\ldots,n$. If the (improper) prior density for (μ,σ^2) is $p(\mu,\sigma^2) \propto 1/\sigma^2$, prove that

1. $\mu|\sigma^2,\boldsymbol{y} \sim N(\bar{y},\sigma^2/n)$
2. $\sigma^2|\boldsymbol{y} \sim inverseGamma\left(\frac{n-1}{2},\frac{S}{2}\right)$, where $S = \Sigma_{i=1}^n (y_i-\bar{y})^2$

Proof:

$$\begin{aligned}
p(\mu,\sigma^2|\boldsymbol{y}) &\propto p(\mu,\sigma^2)p(\boldsymbol{y}|\mu,\sigma^2)\\
&\propto (\sigma^2)^{-1}\left\{(\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\left(S+n(\bar{y}-\mu)^2\right)\right]\right\}\\
&= \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}-\frac{1}{2}+1+\frac{1}{2}}\exp\left[-\frac{1}{2\sigma^2}S\right]\exp\left[-\frac{1}{2\sigma^2}n(\bar{y}-\mu)^2\right]\\
&\propto \underbrace{\left(\frac{1}{\sigma^2}\right)^{\frac{n-1}{2}+1}\exp\left[\frac{-S/2}{\sigma^2}\right]}_{p(\sigma^2|\boldsymbol{y})}\underbrace{\left(\frac{1}{\sqrt{\sigma^2/n}}\right)\exp\left[-\frac{1}{2}\frac{(\bar{y}-\mu)^2}{\sigma^2/n}\right]}_{p(\mu|\sigma^2,\boldsymbol{y})}
\end{aligned}$$

$p(\sigma^2|\boldsymbol{y})$ is the pdf of $inverseGamma\left(\frac{n-1}{2},\frac{S}{2}\right)$

$p(\mu|\sigma^2,\boldsymbol{y})$ is the pdf of $N(\bar{y},\sigma^2/n)$

■

2.22 Suppose we have two sets of data, labelled 1 and 2, and a normal linear regression model holds in each. Consider the model that results from "pooling" the two sets of data, i.e.

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \end{bmatrix}$$

where $\boldsymbol{\varepsilon}_j$ are stochastic disturbances with $E(\boldsymbol{\epsilon}_j|\boldsymbol{X}_j) = \mathbf{0}$ and $V(\boldsymbol{\epsilon}_j|\boldsymbol{X}_j) = \sigma_j^2 \boldsymbol{I}_{n_j}, j = 1,2$

Show that the least squares estimate of $\boldsymbol{\beta}$ formed by *pooling* data sets 1 and 2 is equivalent to a Bayesian analysis, in which the information about $\boldsymbol{\beta}$ in one data set is treated as prior information.

Solution:

We perform Bayesian analysis, treating $[\boldsymbol{y}_1 \quad \mathbf{X}_1]$ as the prior information and $[\boldsymbol{y}_2 \quad \mathbf{X}_2]$ as new information.

Prior density
Let f be the distribution of β with mean = $\mathbf{b}_0$, variance = $\sigma^2 \mathbf{B}_0$, σ^2 is also a random variable.

Our prior mean is the least squares estimate $\widehat{\boldsymbol{\beta}}_1$ from the data $[\boldsymbol{y}_1 \quad \mathbf{X}_1]$

$$\mathbf{b}_0 = \widehat{\boldsymbol{\beta}}_1 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y}_1$$

The prior variance is the covariance matrix $\sigma^2 \mathbf{B}_0$ where $\mathbf{B}_0 = (\mathbf{X}_1'\mathbf{X}_1)^{-1}$

Data and likelihood
New data: $[\boldsymbol{y}_2 \quad \mathbf{X}_2]$
MLE: $\widehat{\boldsymbol{\beta}}_2 = (\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{y}_2$

Posterior mean

$$\begin{aligned} \mathbf{b}_1 &= (\mathbf{B}_0^{-1} + \mathbf{X}_2'\mathbf{X}_2)^{-1}\left(\mathbf{B}_0^{-1}\widehat{\boldsymbol{\beta}}_1 + \mathbf{X}_2'\mathbf{X}_2\widehat{\boldsymbol{\beta}}_2\right) \\ &= (\mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_2'\mathbf{X}_2)^{-1}\left(\underbrace{\mathbf{X}_1'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}}_{\mathbf{I}}\mathbf{X}_1'\mathbf{y}_1 + \underbrace{\mathbf{X}_2'\mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}}_{\mathbf{I}}\mathbf{X}_2'\mathbf{y}_2\right) \\ &= [\mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_2'\mathbf{X}_2]^{-1}[\mathbf{X}_1'\boldsymbol{y}_1 + \mathbf{X}_2'\boldsymbol{y}_2] \end{aligned}$$

Least squares estimate for the pooled data
We have the following form of the estimate $\widehat{\boldsymbol{\beta}}$ using the pooled data

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \left[\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}' \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}\right]^{-1} \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}' \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}$$

$$= [\mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_2'\mathbf{X}_2]^{-1}[\mathbf{X}_1'\boldsymbol{y}_1 + \mathbf{X}_2'\boldsymbol{y}_2]$$

We have shown that the least squares estimate $\widehat{\boldsymbol{\beta}}$ formed by *pooling* data sets 1 and 2 is equivalent to the posterior mean $\mathbf{b}_1$ using Bayesian analysis ■

Problem Set 3

3.3 Refer to Example 3.3. Use Monte Carlo methods to compute the posterior density of the quantities

1. θ_1/θ_0
2. The odds ratio: $\frac{\theta_1(1-\theta_0)}{(1-\theta_1)\theta_0}$
3. The log odds ratio: $\log\frac{\theta_1(1-\theta_0)}{(1-\theta_1)\theta_0}$

Use histograms and numerical summaries to communicate interesting features of these posterior densities.

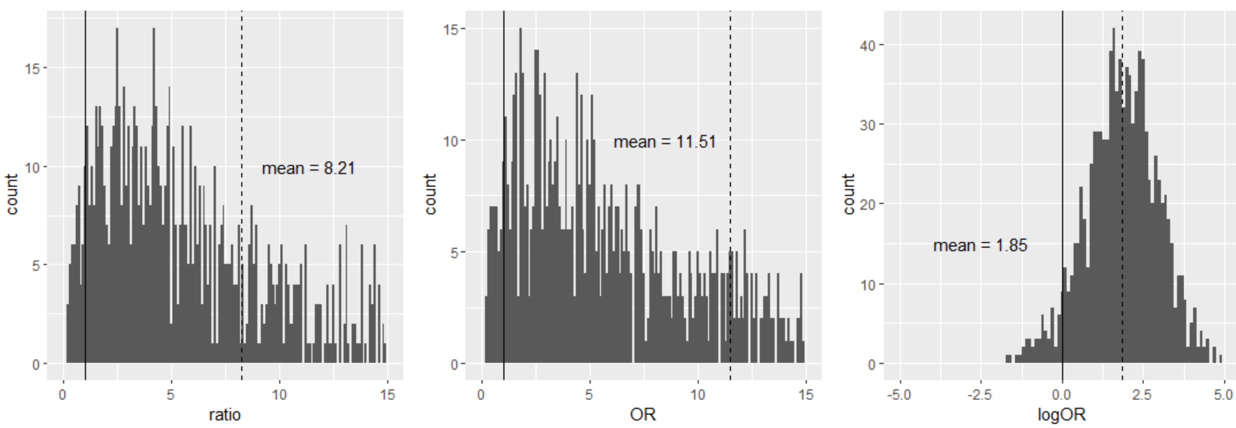
Solution:

```r
n <- 10000
theta_0 <- rbeta(n,2,8)
theta_1 <- rbeta(n,3,75)

#1 ----
ratio <- theta_1/theta_0

#2 ----
OR <- (theta_1/(1-theta_1))/(theta_0/(1-theta_0))

#3 ----
logOR <- log(OR)
```



The ratio θ_1/θ_0 and odds ratio $\frac{\theta_1(1-\theta_0)}{(1-\theta_1)\theta_0}$ both ranges from 0 to ∞ and obviously skewed to the right, with 1 as the null value. On the other hand, the log odds ratio can take any real value and symmetrically distributed around a certain number, with 0 as the null value.

4.9 Verify the claim in Example 4.6:
Suppose $\{\theta^{(t)}\}$ is a Markov chain with $p\left(\theta^{(t)}\right) \equiv N(\rho\theta^{(t-1)}, \sigma^2)$, with $|\rho| < 1$, i.e. a stationary, first order autoregressive Gaussian process. Show that

$$p(\theta) \equiv N\left(0, \frac{\sigma^2}{1-\rho^2}\right)$$

solves Equation 4.5 and so is the stationary distribution of the chain.

Solution:

Recall Equation 4.5:

$$p^{(t)} = \int_{\boldsymbol{\Theta}} K\left(\boldsymbol{\theta}^{(t-1)}, \cdot\right) p^{(t-1)} d\boldsymbol{\theta}^{(t-1)}$$

Need to show that this equation is true given the following:

- $p^{(t)} = p\left(\theta^{(t)}\right) \equiv N(\rho\theta^{(t-1)}, \sigma^2)$
- $K\left(\theta^{(t-1)}, \theta^{(t)}\right) \equiv N\left(\rho\theta^{(t-1)}, \sigma^2\right)$
- $p(\theta) \equiv N\left(0, \frac{\sigma^2}{1-\rho^2}\right)$

Now, simplifying the right-hand side of Equation 4.5 while assuming $p(\theta) \equiv N\left(0, \frac{\sigma^2}{1-\rho^2}\right)$

$$= \int_{\boldsymbol{\Theta}} \frac{1}{\sqrt{2\pi\sigma^2/(1-\rho^2)}} \exp\left(\frac{-\left(\theta^{(t)}\right)^2}{2\sigma^2/(1-\rho^2)}\right) p^{(t-1)} d\theta^{(t-1)}$$

$$p^{(t)} = \underbrace{\frac{1}{\sqrt{\frac{2\pi\sigma^2}{1-\rho^2}}} \exp\left(\frac{-\left(\theta^{(t)}\right)^2}{\frac{2\sigma^2}{1-\rho^2}}\right)}_{\equiv N\left(0, \frac{\sigma^2}{1-\rho^2}\right)} \underbrace{\int_{\boldsymbol{\Theta}} p^{(t-1)} d\theta^{(t-1)}}_{1}$$

■

4.12 Conduct a Monte Carlo experiment to investigate the result in Equation 4.12. That is, repeat the following steps a large number of times, indexed by $m = 1, \dots, M$, with M set to a large number (e.g. $M = 5000$):

a. Generate a large sample from the stationary, Gaussian, AR(1) process $z_t | z_{t-1} \sim N(\rho z_{t-1}, 1 - \rho^2)$, with ρ set to a relatively large value (e.g. $\rho = 0.95$) and $E(z_t) = 0$. Note that marginally, $Var(z_t) = 1$. Try $T = 10^5$ or so. Again, the `arima.sim` function in `R` is an easy way to do this. Compute and store the mean of the sampled $z_t, \bar{z}_T^{(m)}$.

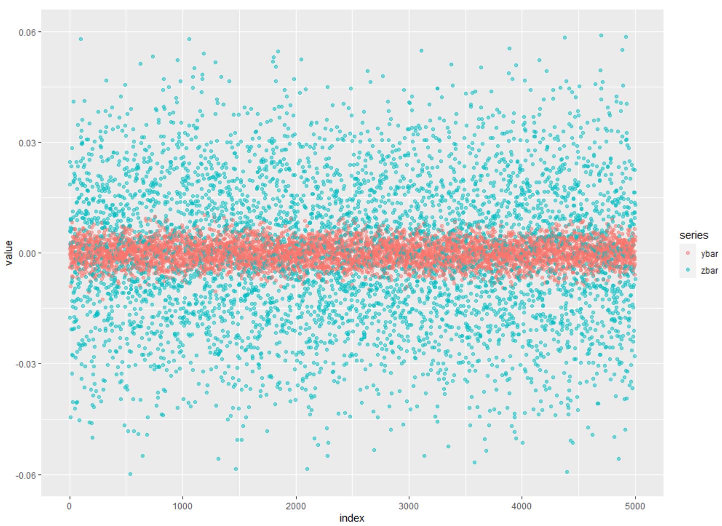
b. Use an independence sampler to generate T draws from the marginal distribution $y_t \sim N(0,1)$; use the `rnorm` function in `R`. Compute and store the mean of the sampled $y_t, \bar{y}_T^{(m)}$

Over the M replicates of this Monte Carlo experiment, compute the variance of the means $\bar{z}_T^{(m)}$ and $\bar{y}_T^{(m)}$. Up to Monte Carlo error, you should observe the result in Equation 4.10; the independence sampler generates $\bar{y}_T$ that are less dispersed around zero than the AR(1) sampler, with $\frac{var(\bar{z})}{var(\bar{y})} \approx \frac{1+\rho}{1-\rho}$.

Algorithm (R codes):

```
M = 5000; rho = 0.95; t = 10^5
zbar <- c(); ybar <- c()
for (m in 1:M){
    # a: AR(1) process
    z_t <- arima.sim(n = t, list(ar = rho), sd = sqrt(1-rho^2))
    zbar[m] <- mean(z_t)
    # b: independence sampler
    y_t <- rnorm(t)
    ybar[m] <- mean(y_t)
}
var(zbar) ; var(ybar)
```

The scatterplot shows both series centered around 0, but $\bar{z}_T^{(m)}$ is more dispersed than $\bar{y}_T^{(m)}$.



Furthermore, computing the variance of these series, the following equality proves to be true:

$$\frac{var(\bar{z})}{var(\bar{y})} = \frac{0.0003957366}{0.000009934624} = 39.83408$$

$$\approx \frac{1+\rho}{1-\rho} = \frac{1+0.95}{1-0.95} = 39$$