Clayton Bass	**Final Project**
Due 8/31/20	**Prof. Parno's MATH 76 20X**

A Bayesian Regression Analysis of U.S. Suicide Rate Trends from 1985-2015

Introduction

As online news sources and reporting become ever more ingrained into our daily lives, so too do those horrifying stories and reports that remind us all of the serious troubles in our world, whether social, political, cultural, environmental, or otherwise. With such remarkable ease of access to news sources—just a few clicks away on a computer or a phone—it might at times seem as though everything negative or harmful in the world is only growing worse. But it might just be the case that our world is *not* entirely growing worse, but rather that we simply have exponentially increasing access to all of the painful and upsetting news stories that were once either not publicized or simply swept under the rug.

In this project, I aim to test whether suicide rates in the U.S. have been decreasing over the past 30 years. While suicide is a remarkably prevalent issue, and day by day we hear of more and more suicides in the news and from other sources, it is important to consider the facts when informing our opinions about the state of suicide in this country. It is incredibly important to trace suicide trends over the past 30 or so years to see just how much worse suicide rates have gotten in light of the certain uptick in news media stories covering suicides (unless in fact the broad trend is that they have not gotten worse). Is it the case that suicides were once more prevalent than they are now but were simply not as heavily reported? Is it the case that suicides are both more heavily reported now *and* more prevalent? These are questions that might be answered by examining the trend in suicide rates since 1985 in the U.S.—recent enough data for even the oldest data points to be relevant. It would be helpful to understand these trends to help decrease suicides going forward and, if alarming or otherwise concerning trends are noticeable from this analysis, perhaps speculate on why this might be the case and how we may go forward to solve such a pressing problem.

More precisely, I aim to test the following situation. Consider that we have the suicide rate per 100,000 people in the U.S. for every year since 1985, and we calculate the suicide rate anomaly—that is, the suicide rate per 100,000 people minus the average suicide rate per 100,000 people from 1985 through 2000 (where we notice there may be greater anomalies after 2000 based on the data that we will examine). In this project, I ascertain the probability that the regression line fit to the data (via Bayesian linear regression with a Gaussian prior and likelihood) has a negative slope. This is essentially asking the following question: What is the probability that the suicide rate per 100,000 people in the U.S. has been decreasing from 1985 through 2015?

In particular, the dataset that I will use (see Kaggle in my Works Cited section for the link to the dataset published by user Rusty, or russellyates88) contains the number of suicides in any given year, broken down by country, then by age band, then by sex, with data coming from various credible sources listed in the Kaggle dataset description (Rusty). I will be performing empirical Bayes estimation (see the Variance Explained post by David Robinson in my Works Cited section, from the very end of the Conjugate Inference problem set) over the global data by aggregating over country, sex, and age band, calculating the sum of the suicides divided by the sum of the disjoint populations (since each band, separated by age, sex, and country, represents a disjoint population) and multiplying by 100,000 to achieve the suicide rate per 100,000 people (Robinson). (As will be explained later, empirical Bayes estimation is valid here since we have many more individuals/data points considered on the global scale than just in the U.S., and this serves to inform my selection of prior mean and covariance (Robinson).) With regard to my specific case study of the U.S., I simply focus only on the U.S. suicide data, aggregating the total number of suicides per year (over the sex and age bands given for that year) and dividing by the number of people in those groups, multiplying by 100,000. It is also very easy to calculate the suicide rate anomaly, as aforementioned, from this data.

Formulation

In this section, I describe my statistical model for the data (i.e., my likelihood function), what parameters are in my model (i.e., what I am trying to infer with Bayes rule), what my prior distribution is and how that choice reflects information known a priori about the parameters, and finally how I can use the posterior density to answer the question defined in my introduction. I will define my posterior density $f(x|y)$ and a function $h(x)$ such that $E_{x|y}[h(x)]$ helps me ascertain the probability that the suicide rate in the U.S. has been decreasing from 1985 through 2015 (i.e., the probability that the slope of the regression line fit to the data is less than 0).

Note that the posterior I define (i.e., the likelihood, prior, and evidence I define) is the same (in terms of its form) for both the global data (in the empirical Bayes estimation portion of my work) and the U.S. data. So, the following formulation of my Bayesian inference problem will apply to both global and U.S. suicide data. There is explicit delineation between the global and U.S. cases in the accompanying Python code.

Before defining our likelihood, note that in the joint density for $f(c, y)$ we have $f(c, y) = N\left(\begin{bmatrix}\mu_c \\ \mu_y\end{bmatrix}, \begin{bmatrix}\Sigma_{cc} & \Sigma_{cy} \\ \Sigma_{yc} & \Sigma_{yy}\end{bmatrix}\right)$. We define $\Sigma_{yy} = Cov[Vc + \epsilon] = V\Sigma_{cc}V^T + \Sigma_{\epsilon\epsilon}$, and $\Sigma_{yc} = E[(Y - \mu_y)(C - \mu_c)^T] = V\Sigma_{cc}$, and $\Sigma_{cy} = \Sigma_{yc}^T = \Sigma_{cc}V^T$, so that

$$f(c, y) = N\left(\begin{bmatrix}\mu_c \\ V\mu_c\end{bmatrix}, \begin{bmatrix}\Sigma_{cc} & \Sigma_{cc}V^T \\ V\Sigma_{cc} & V\Sigma_{cc}V^T + \Sigma_{\epsilon\epsilon}\end{bmatrix}\right).$$

We use some of the elements of our mean and covariance matrix in the functions we define as follows.

First and foremost, we define our likelihood function for the data. Our likelihood function should be a statistical model of our data. We will assume that we have a Gaussian likelihood on our anomaly data. This is because, when we define our suicide rate anomalies, which are essentially just the suicide rate per 100k people minus the average suicide rate per 100k people from 1985 through 2000, we should approximately see a distribution centered around 0 that can take both positive and negative values. We define a linear model for our observations of suicide rate anomalies as $y = Vc + \epsilon$, where we have the Vandermonde matrix $V := \begin{bmatrix}1 & t_1 \\ 1 & t_2 \\ ... & ... \\ 1 & t_N\end{bmatrix}$, and the vector of coefficients $c = \begin{bmatrix}c_0 \\ c_1\end{bmatrix}$, and our error term $\epsilon \sim N(0, \sigma_{\epsilon\epsilon})$. Let $y = \begin{bmatrix}y_1 \\ y_2 \\ ... \\ y_N\end{bmatrix}$ represent our vector of anomaly observations in each year (where $N = 31$, since there are 31 years from 1985 through 2015, inclusive). Then, in order to define our likelihood, we know we must use the conditional Gaussian identity that states, for some observed or fixed vector of coefficients $\bar{c}$, that $f(y|c = \bar{c}) = N(\mu_y + \Sigma_{yc}\Sigma_{cc}^{-1}(\bar{c} - \mu_c), \Sigma_{yy} - \Sigma_{yc}\Sigma_{cc}^{-1}\Sigma_{yc}^T)$. This simplifies, therefore, to the following likelihood via substitution:

$$f(y|c=\bar{c}) = N\left(V\mu_c + V\Sigma_{cc}\Sigma_{cc}^{-1}(\bar{c}-\mu_c), (V\Sigma_{cc}V^T + \Sigma_{\epsilon\epsilon}) - V\Sigma_{cc}\Sigma_{cc}^{-1}\Sigma_{cc}V^T\right).$$

Cancelling the inverse matrix multiplications and other terms, we get

$$f(y|c=\bar{c}) = N\left(V\bar{c}, \Sigma_{\epsilon\epsilon}\right) = (2\pi)^{-\frac{N}{2}}|\Sigma_{\epsilon\epsilon}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(y-V\bar{c})^T\Sigma_{\epsilon\epsilon}^{-1}(y-V\bar{c})\right],$$

where once again we have $N = 31$. We use this likelihood function both when we are moving from hyperprior to prior (i.e., when we are examining the global data, so we start with a mean $\mu_c = \begin{bmatrix} 0.1 \\ 0.001 \end{bmatrix}$ and diagonal variances of 1 in Σ_{cc}) and when we are moving from prior to posterior (i.e., when we use the posterior distribution from the global analysis as our prior for performing Bayesian regression on U.S. data, where our prior distribution will use the mean and variance of the coefficient vector from the analysis we just described of moving from prior to hyperprior). Note significantly that the Gaussian distribution may be justified as a likelihood function for our data because we may view each occurrence of a suicide as a Bernoulli random variable with some probability. Assuming that each suicide occurs independently of any other suicide, we know that the sum of some set of independent Bernoulli random variables follows a Binomial distribution. With enough observations (in this case, we have hundreds of thousands even just within a country in one year), we know that by the central limit theorem, the sample mean of a set of independent and identically distributed random variables approximates the Normal distribution with enough observations. Here, we have millions of observations, and we assume that the suicide rate anomalies follow a Gaussian distribution (approximately) as a result. (Dividing our suicide counts by the population count and then multiplying by 100,000 to get a rate out of 100,000 should not change our distribution assumption, since in any case we are dividing by constants throughout, and the suicide count is the only variable that actually matters. However, note that one may also view the aggregate suicide rate anomaly in any given year as a single observation since we are normalizing based on the population and we are performing a regression on 31 data points, for each year from 1985 through 2015, inclusive. We may still assume a normal distribution over the suicide rate anomalies since we have 31 data points, which may actually be enough for sufficient convergence to a Gaussian distribution. The rates themselves are not as well approximated by a Binomial distribution, since there is no sense of "yes" or "no" here—there is a floating-point suicide rate that could be count data, like a Poisson distribution, but the Poisson distribution itself approximates the Gaussian with enough data points, so we stick with our Gaussian likelihood assumption.)

Now, recall that we are trying to infer the parameters c_0 and c_1 in the coefficient vector c in this problem via Bayes' rule. We therefore must establish a prior distribution over our parameters c_0 and c_1 that reflects some information known a priori about the parameters. Note that per the article by Robinson on empirical Bayes estimation, it can often be helpful to establish a priori information known about some parameters in the prior of the distribution at hand by examining a much larger data set (that contains, as a much smaller subset, the data points of interest on which we must ultimately establish a final posterior density) and developing a hyperprior on that dataset and a posterior to that hyperprior (which will be the prior for our smaller subset of data) (Robinson). In this case, I am running my regression analysis on the entire set of global data, for which there are many more data points than there are in just the U.S. data set. Therefore, I start with some hyperprior assumption on the mean vector μ_c and covariance vector Σ_{cc}, based on my theory that the anomalies should be slightly above 0 in mean due to pull upwards in recent years, and with perhaps standard deviation 1 (and nothing huge like 10 or 100, since the rates themselves are quite small, and should not have changed drastically in the last 30 years). We start with the assumption on our hyperprior that $\mu_{c,hp} = \begin{bmatrix} 0.1 \\ 0.001 \end{bmatrix}$ and with diagonal variances of 1 in $\Sigma_{cc,hp}$. The *hp* designation is for hyperprior. We establish a Gaussian hyperprior on our coefficient vector here, since we assume some symmetric distribution around somewhere just above 0 for our vector parameters, and some small standard deviation 1. It is also convenient that our Gaussian prior is essentially conjugate to our Gaussian likelihood that we defined in the previous paragraph (which we employ exactly for both the global and U.S. data, simply with differing parameters). Similarly, we establish a Gaussian distribution on our prior coefficients with mean $\mu_{c,p}$ and covariance $\Sigma_{cc,p}$, derived explicitly as the *posterior* of the analysis of the global data. (In essence, these parameters here—$\mu_{c,p}$ and $\Sigma_{cc,p}$—are found through the form of our posterior density function, which we define in our next paragraph, but note that they are determined through a priori information gathered about *global* suicide rates, which will be discussed a bit more in the next section and which is covered quite extensively in the accompanying Jupyter Notebook.) So, our hyperprior is given by

$$f_{hp}(c) = N(\mu_{c,hp}, \Sigma_{cc,hp}) = (2\pi)^{-\frac{d}{2}} |\Sigma_{cc,hp}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(c - \mu_{c,hp})^T \Sigma_{cc,hp}^{-1} (c - \mu_{c,hp})\right],$$

and our prior is given by

$$f_p(c) = N(\mu_{c,p}, \Sigma_{cc,p}) = (2\pi)^{-\frac{d}{2}} |\Sigma_{cc,p}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(c - \mu_{c,p})^T \Sigma_{cc,p}^{-1} (c - \mu_{c,p})\right],$$

where in both cases we have $d = 2$, since that represents the dimension of our mean vector.

Finally, we define the posterior density and ultimately answer our question using an expectation over the posterior density. Note that we define two posterior densities here implicitly, since first, we will need to use a posterior density when we are moving from our hyperprior parameters to our prior parameters (since the prior parameters are posterior to the hyperprior parameters when we are examining the global data), and second, we will need to use a posterior density when we are moving from our prior parameters to our posterior parameters (in the U.S. data specifically). The regression analysis is performed identically in both cases, and the posterior is exactly the same in form (albeit with different values, which are explicitly delineated in the accompanying Jupyter Notebook code). To that end, note critically that the closed form of our posterior distribution is known in terms of Gaussian identities (and is itself Gaussian) since we chose a conjugate Gaussian prior for our Gaussian likelihood. We define our posterior distribution as $f(c|y) = N(\mu_{post}, \Sigma_{post})$, where c is our vector of regression coefficients $\begin{bmatrix} c_0 \\ c_1 \end{bmatrix}$ to be estimated, and y is our vector of suicide anomaly rate observations $y = Vc + \epsilon$, with $\epsilon \sim N(0, \Sigma_{\epsilon\epsilon})$. (Note that the error terms are defined as independent random variables in the manner $\epsilon \sim N(0, \Sigma_{\epsilon\epsilon}$, and are tuned in our software to fit the observed variation in the data on the plots of the posterior predictive distribution (i.e., the mean regression line) and cover all data points in case some data points are not covered.) We know by the Gaussian identity of conditional distributions that $\mu_{post} = \mu_c + \Sigma_{cc} V^T (V \Sigma_{cc} V^T + \Sigma_{\epsilon\epsilon})^{-1} (y - V\mu_c)$, and $\Sigma_{post} = \Sigma_{cc} - \Sigma_{cc} V^T (V \Sigma_{cc} V^T + \Sigma_{\epsilon\epsilon})^{-1} V \Sigma_{cc}$, where these terms are all as explicitly defined in our joint density $f(c, y)$ in the earlier paragraph in this section. Therefore,

$$f(c|y) = N\left(\mu_c + \Sigma_{cc} V^T (V \Sigma_{cc} V^T + \Sigma_{\epsilon\epsilon})^{-1} (y - V\mu_c), \Sigma_{cc} - \Sigma_{cc} V^T (V \Sigma_{cc} V^T + \Sigma_{\epsilon\epsilon})^{-1} V \Sigma_{cc}\right).$$

In other words, our posterior is given by

$$f(c|y) = N(\mu_{post}, \Sigma_{post}) = (2\pi)^{-\frac{d'}{2}} |\Sigma_{post}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(c - \mu_{post})^T \Sigma_{post}^{-1} (c - \mu_{post})\right],$$

where $d' = 2$ since again we have our mean vector defined over the regression coefficients, of which there are only two—that is, c_0 and c_1. Note that this is the posterior over our parameters c. We have that

$$y_{pred} \sim N(V\mu_{post}, V\Sigma_{post}V^T + \Sigma_{\epsilon\epsilon}),$$

and we will plot these values.

We also now define our function $h(c)$ such that we can utilize $h(c)$ in an expectation over $f(c|y)$ in order to determine whether the slope of the regression line fit to the data in the U.S. is negative (which would indicate that the suicide rate is decreasing over time). (We also apply this formula to the global data just for further info.) That is, we aim to turn our question into a problem of finding $E_{c|y}[h(c)]$. Note that if we define $h(c) := \mathbb{1}(c_1 < 0)$, where $\mathbb{1}$ is the indicator function for whether our slope coefficient c_1 is negative, then we can define the posterior expectation

$$E_{c|y}[h(c)] = \int_{c_1=-\infty}^{\infty} \int_{c_0=-\infty}^{\infty} \mathbb{1}(c_1 < 0) \cdot f(c|y) dc_0 dc_1,$$

where we recall that c is the ***vector*** of coefficients c_0 and c_1, and c_1 is a scalar representing the slope in our linear model. This directly yields the probability that the slope of our linear model is negative, since the indicator function is 1 if and only if $c_1 < 0$, but is 0 otherwise (and thus the entire integral is zero). We discuss the exact computational method of this integral in the next section (on solution strategy), although a fair amount of what will be discussed in that section is directly inferred (or in some cases essentially covered already) in this section by the nature of our discussion of likelihood, prior, and posterior definition.

Solution Strategy

In this section, we describe the details of how specifically we compute the posterior density and posterior expectations defined in the formulation section. We explicitly must state how the parameters in the posterior distribution are related to the parameters in the prior distribution (so that we may accurately define the posterior distribution in both the global case and the U.S. case), and then we must discuss how we will use these parameters to compute $E_{c|y}[h(c)]$. Finally, we conclude with a brief discussion of the software developed in order to answer these questions.

In the previous formulation section, we explicitly defined that

$$\mu_{post} = \mu_c + \Sigma_{cc}V^T(V\Sigma_{cc}V^T + \Sigma_{\epsilon\epsilon})^{-1}(y - V\mu_c),$$

and

$$\Sigma_{post} = \Sigma_{cc} - \Sigma_{cc}V^T(V\Sigma_{cc}V^T + \Sigma_{\epsilon\epsilon})^{-1}V\Sigma_{cc},$$

where these terms are all as explicitly defined in our joint density $f(c, y)$ in the earlier paragraph in this section. We know that $f(c|y) = N(\mu_{post}, \Sigma_{post})$ by the Gaussian identity of conditional distributions. Therefore, it is straightforward to explicitly calculate μ_{post} and Σ_{post} via standard matrix multiplication methods and linear algebra tools (particularly for inverting matrices) via NumPy (which we explain a bit more in-depth later in this section). Note that this is the posterior over our parameters c. We have that

$$y_{pred} \sim N(V\mu_{post}, V\Sigma_{post}V^T + \Sigma_{\epsilon\epsilon}),$$

and we will plot these values. Again, this calculation is straightforward given the values of μ_{post} and Σ_{post}. As long as we have explicitly defined μ_c and Σ_{cc} as we have done in our accompanying Python code (when generating hyperprior parameters), and we have defined V (which we do explicitly in the code), and $\Sigma_{\epsilon\epsilon}$ (which we also tune explicitly in the code based on how wide we think the error bars should be in our model, based on our plot of the data), and y (which is simply the observed suicide anomalies), then we can calculate μ_{post} and Σ_{post} in straightforward fashion. This is our statement of how the parameters in the posterior distribution are explicitly related to the parameters in the prior distribution. This methodology works when exactly the same when moving from the hyperprior to the prior (i.e., when computing the posterior on the global data), and when moving from the U.S. prior to the U.S. posterior (where the U.S. prior was essentially the posterior on the global data).

Now, we know that we have access to parameters μ_{post} and Σ_{post} in both the global and U.S. regression analyses. We also have that

$$E_{c|y}[h(c)] = \int_{c_1=-\infty}^{\infty} \int_{c_0=-\infty}^{\infty} \mathbb{1}(c_1 < 0) \cdot f(c|y) dc_0 dc_1,$$

where we recall that c is the ***vector*** of coefficients c_0 and c_1, and c_1 is a scalar representing the slope in our linear model. Note critically, however, that since our posterior distribution $f(c|y)$ is a joint distribution over both c_0 and c_1, and that our posterior distribution is Gaussian, we can use the Gaussian cumulative distribution function (implemented in the Python code in our accompanying Jupyter Notebook), taking the integral from $-\infty$ to our desired value 0 over our marginal density function over c_1. Fortunately, we know from our Gaussian identities that we can essentially just pick out the mean and the variance for c_1 from our joint density definition, so that we simply extract (in Python) the scalar μ_{c_1} from the vector μ_{post}, and extract the entry in row 2, column 2 of the covariance matrix Σ_{post} to get $\Sigma_{c_1,c_1} = Var[c_1] = \sigma_{c_1}^2$. We then have that

$$f(c_1|y) = N(\mu_{c_1}, \sigma_{c_1}^2) = \frac{1}{\sqrt{2\pi\sigma_{c_1}^2}} \exp\left[-\frac{1}{2}\frac{(c_1 - \mu_{c_1})^2}{\sigma_{c_1}^2}\right],$$

where we can easily calculate the integral

$$E_{c|y}[h(c)] = \int_{c_1=-\infty}^{\infty} \int_{c_0=-\infty}^{\infty} \mathbb{1}(c_1 < 0) \cdot f(c|y) dc_0 dc_1 = Pr[c_1 < 0] = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma_{c_1}^2}} \exp\left[-\frac{1}{2}\frac{(c_1 - \mu_{c_1})^2}{\sigma_{c_1}^2}\right] dc_1$$

via built-in methods in our accompanying Python code.

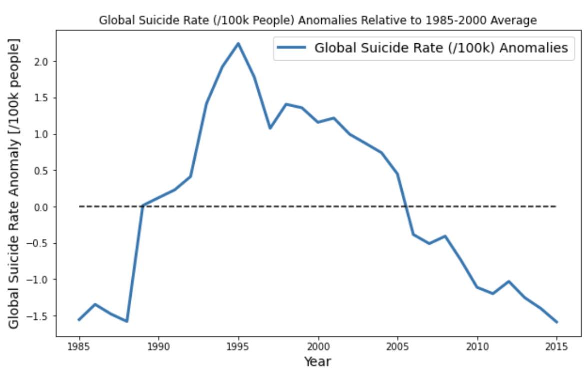
I developed a fair amount of software, drawing extensively on the code used in lecture 14 in the lecture14-code repository in the Jupyter Notebook file ExtentRegression.ipynb, modifying the code sufficiently enough to match my specific problem. I relied heavily on pandas, NumPy, matplotlib, pymuqModeling, pymuqApproximationWrappers, and scipy.stats in order to carry out my inference problem and make visualizations (via matplotlib in particular). I used identical software concepts in the empirical Bayes estimation portion of the project and in the main project (where I actually estimated the posterior parameters over the U.S. data), so I will discuss this all as one (Robinson). First and foremost, I cleaned the data so that it was aggregated over sex and age band, and such that I could generate a suicide rate per 100k people in a given year, both on a global scale and on a U.S.-specific scale. I then plotted the global suicide rate anomalies (by subtracting the mean suicide rate per 100k people in a given year from the suicide rate per 100k people for that year, for every year from 1985 through 2015) to get a sense of where my data lie.

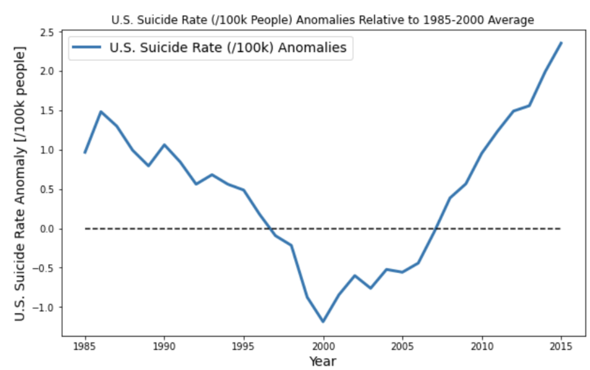
I defined functions CreateVandermonde, ConstructPrior, and ConstructPosterior (as in lecture 14) in order to do exactly what the functions state; these were critical in applying the formulas defined in the previous section and in this section to carry out the Gaussian regression analysis. I developed a Vandermonde matrix over the years over which to be predicted, and then created the predicted means and covariances (and standard deviations) over the hyperprior, the prior, and the posterior, utilizing the formulas relating the mean and covariance of y to the mean and covariance of c (via the linear model $y = Vc + \epsilon$). I evaluated the hyperprior and prior on global data in order to illustrate that bivariate density of slope (c_1) vs. intercept (c_0) and to visualize the distribution of the resulting coefficients. I then developed plots of the posterior predictive distributions for the hyperprior, and also plotted $\mu \pm 2\sigma$ to visualize the error bars. This illustrated the negative slope directly. Also, just for further info, I calculated the probability that the prior slope would be negative using the norm.cdf library-built-in function in Python, which allowed me to input the prior slope mean and prior slope covariance into the function's arguments. This process was repeated essentially identically for the main problem where we attempt to estimate the posterior regression coefficients over the U.S. data specifically, and the probabilities were ultimately computed using the same software.

Results

In this section, I describe what I found after solving the inference problem and computing the posterior expectations. I include some relevant plots of my posterior density and samples, and list the values of my posterior expectations.

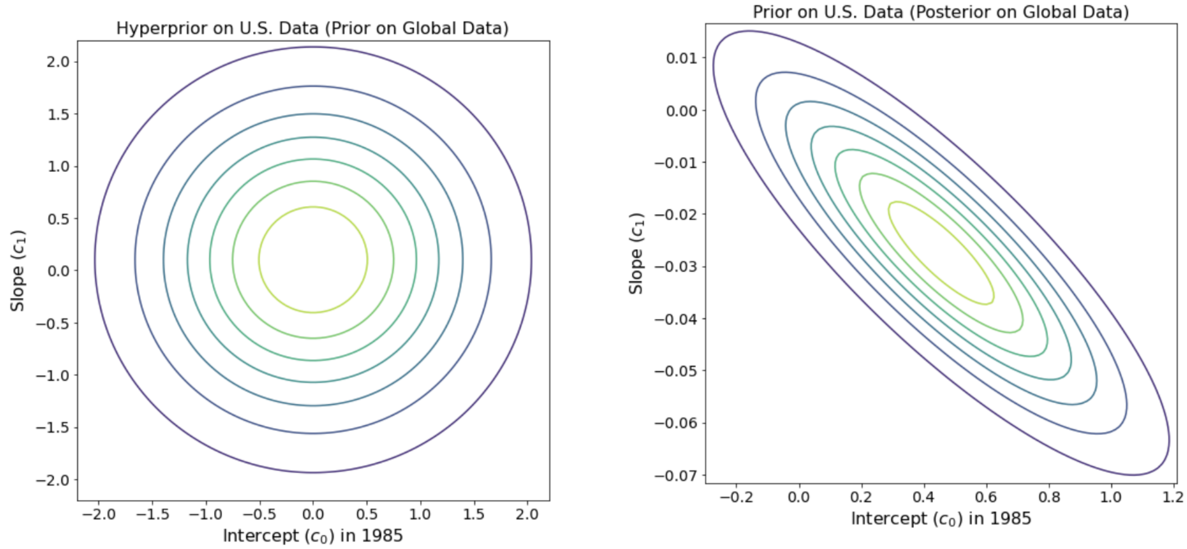
First, we illustrate some of our posterior density plots. Here, we have plots of the global and U.S. suicide rate anomalies per 100k people. Note that the shapes of these plots are actually quite different; there is a minimum around 2000 for the U.S., but there is a much higher value (relatively speaking) around that time on a global scale. Note that in all of the following side-by-side plots, we use the subcaption LaTeX package per the article by Yifan Peng (see the Works Cited section at the end), where an effective method of placing plots side-by-side is implemented (Peng).

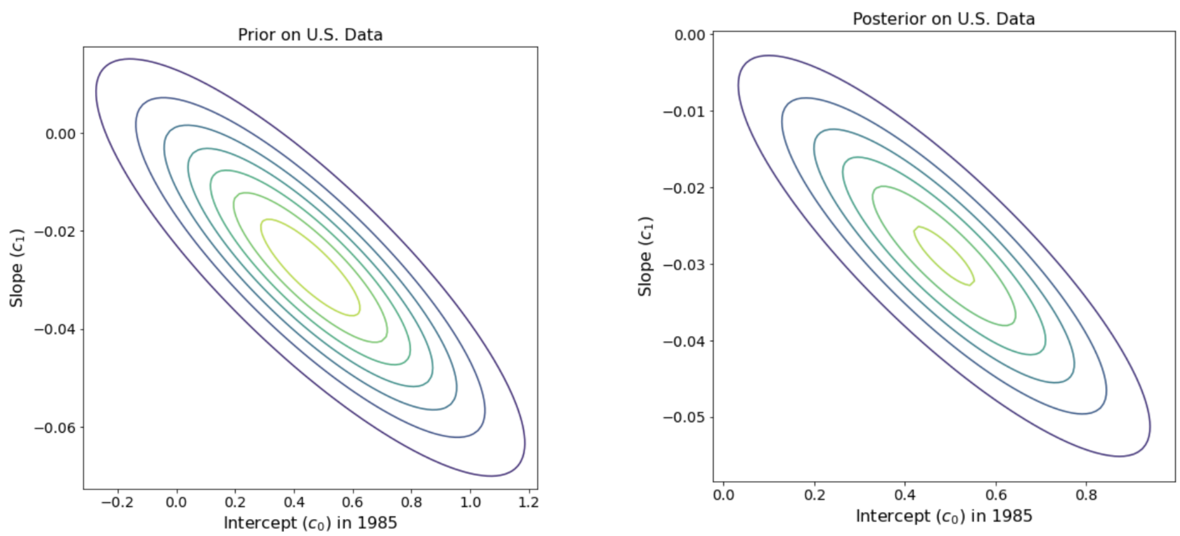




Next, we plot the hyperprior and prior densities, noting that we clearly have a majority of the density lying below 0.00 for the prior on the U.S. data (i.e., the posterior on the global data), indicating a high probability of a negative slope, which we will cover soon. Our hyperprior does not tell us a good deal about the location of the slope just yet. We implement this display with the subcaption package (Peng).

Next, we plot the prior and posterior densities on the U.S. data, noting again that we clearly have a majority of the density lying below 0.00 for the posterior on the U.S. data, indicating a high probability of a negative slope, which we will cover soon. We implement this display with the subcaption package (Peng). (These plots lie on the next page. The global posterior and U.S. prior are identical.)

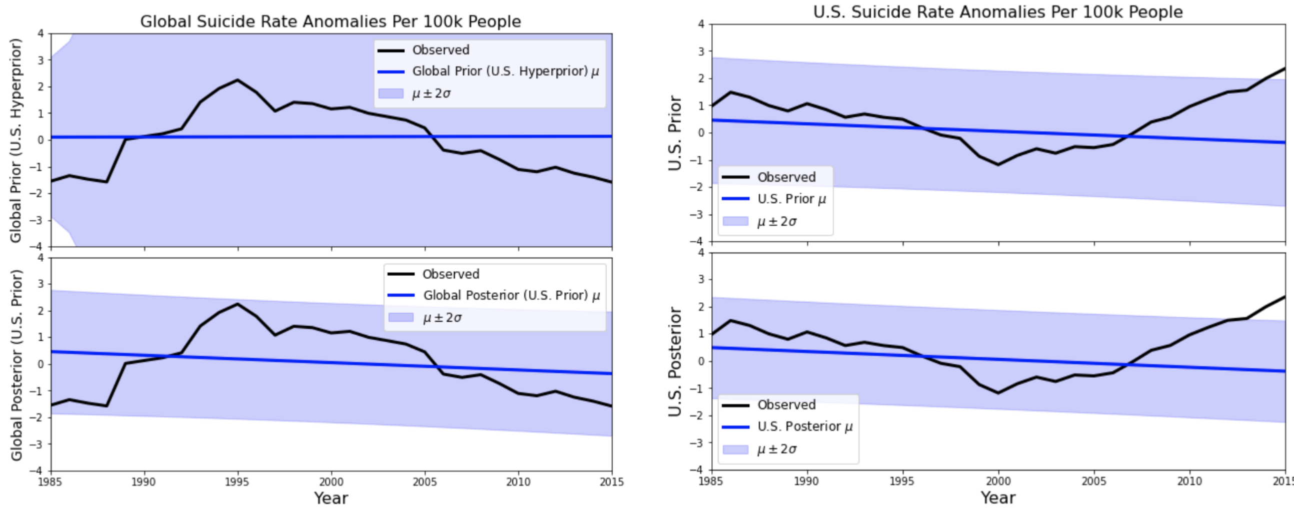




Note that our prior (in the second set of density plots) is the posterior from the previous plot (i.e., the posterior on the global data is our prior for the U.S. data). The U.S. posterior density tightens a noticeable amount but the location is quite similar to that of the prior. It is critical that we notice that the tightening of our posterior density since that directly indicates that we are more confident in our posterior parameters than we were about our prior parameters, so this gives us some sense that our empirical Bayes estimation was effective. Now, we plot the predictive distributions. We implement this display with the subcaption package (Peng).

Note that in Figure (a) at the top of the next page (where our posterior predictive distributions lie), we clearly see a transition from a very large variance around our data to a much tighter bound based on our tuning of the parameter for the covariance of ϵ (with itself) to fit all the data points. The error bars do tighten when moving from the prior to the posterior in the right plot of the U.S. data specifically, reflecting our greater confidence in our posterior parameter distribution.

Note that our global posterior expectation (so our U.S. prior expectation)—i.e., the probability that the slope of the regression line is negative—is given by 0.905, implemented cleanly in the



(a) Global Data

(b) U.S. Data

accompanying Python code, indicating a large probability that we have a decreasing number of suicides over time across the globe (on the whole) from 1985 through 2015.

Also, note that our U.S. posterior expectation—i.e., the probability that the slope of the regression line is negative—is given by 0.983, implemented cleanly in the accompanying Python code, indicating a large probability that we have a decreasing number of suicides over time in the U.S. from 1985 through 2015. In the next section, we reflect more on these values and also provide some more insight into the posterior distribution itself (i.e., means, variances, probabilities, etc.) to gain some more insight into the data and the problem.

Discussion

In this section, we provide some final analysis of the main suicide trends question, along with a discussion of the validity of the conclusions, whether or not they would change with more relaxed assumptions, and whether the posterior distribution's mean, variance, and probabilities may give some additional insight into the data or the problem at hand.

First, the answer to my question is yes—suicide rates have been decreasing from 1985 through 2015 in the U.S. (The answer is also yes on the global scale, but that is not the specific question we are seeking to answer, and we simply used an analysis of the global scale in order to better inform our prior distribution on the U.S. scale.) Since our U.S. posterior expectation—i.e., the probability that the slope of the regression line is negative—is given by 0.983, implemented cleanly in the accompanying Python code, we see that we have a large probability that we have a decreasing number of suicides over time in the U.S. from 1985 through 2015. We can say with with relative confidence based on our data that there is a downtrend in the suicide rate per 100k people in the United States over the 30 years from 1985 through 2015. This information is at first alarming and made me wonder how valid my conclusions were, or perhaps how valid the original data source is, or perhaps how serious the gaps in the data are. Consider the CDC webpage by Hedegaard et al. in my Works Cited—we see that there is a claimed 35% growth in suicides since 1999 in the U.S. (Hedegaard et al.). While this may at first seem to contradict my conclusion that there has been a relative downtrend in suicides from 1985 through 2015, note that if one examines closely the posterior predictive plot of U.S. suicide rates per 100k people, there is only a clear uptick in suicide rate starting at around the year 1999, which is conveniently when the CDC study I just mentioned begins recording suicides. It is therefore important to consider that there may very well be a serious uptick in suicides in the past 20 or so years, but it is also important to be cognizant of the general, overarching history of suicide rates in the United States and to realize that, considering cases as far back as 1985, there is a general downward slope, so all in all the suicides fall within some known range that is not at the moment wildly out of control. However, the presence of other studies that might indicate a recent growth in suicide rate is an important trend to keep an eye on in coming years, as human lives are so directly at stake here.

I would argue thus that my conclusions are valid considering data from 1985 through 2015 (which we have done in this project). While there is some structure in the data at hand that is not entirely captured by the linear model, it is important to consider that the linear regression model by definition must smooth out some of the bumps in the data. There is a relatively noticeable downtrend in suicide rates *as a whole* despite a recent apparent upsurge in suicide rates in the past 20 years, so by the nature of the question posed—covering a relatively long span of time—there will be some bumps or trends in between that are themselves smoothed out. But sometimes it is important to examine broader trends throughout history rather than focus exclusively on a narrow time interval, although of course both are tremendously important when addressing as critical of a question as suicide rates. If I were to relax some of my assumptions, then I am not quite sure my conclusions would be the same anymore. First of all, applying a Gaussian distribution assumption on my prior and likelihood functions was entirely necessary in order for me to obtain my Gaussian posterior, since we are essentially using our Gaussian prior as a conjugate prior for our likelihood. It is reasonable to assume that the suicide rate anomalies (for our likelihood) are themselves Gaussian with some mean close to zero and some standard deviation close to 1, but if we were not even to constrain these values to a distribution and instead leave them as a Poisson distribution, then we would not have quite as clean of a Bayesian analysis problem, since the behavior of a Gaussian prior coupled with a Gaussian likelihood is seemingly unparalleled in terms of efficiency of calculation and niceness of properties. The conditional distribution for the multivariate Gaussian, which is used explicitly in deriving our likelihood and posterior functions, does not actually depend on any independence assumptions or anything else that would simplify our calculations immensely; this simply stems from the Gaussian prior and likelihood assumptions we have made. The variance of the posterior predictions seem to be quite in line with the data, and the mean clearly reflects the slight downward trend in the data; although in the most recent years there are some outliers (beyond the 2σ deviation), it is not necessarily requisite that every single data point falls completely within our error bars, as long as much of our data is clearly represented within our uncertainty estimates (which we see is the case here). Note that $y_{pred} = Vc + \epsilon \Rightarrow Cov[y_{pred}] = V\Sigma_{post}V^T + \Sigma_{\epsilon\epsilon}$, where repeated testing in this problem has illustrated that the parameter variability is much less than the variability in the error term, so we are able to capture most of the data we see by adjusting $\Sigma_{\epsilon\epsilon}$ to appropriately reflect our posterior observations (a process similar to what we learned in Lecture 22). (We may be able to approximate the covariance of y_{pred} then by just $\Sigma_{\epsilon\epsilon}$, but to be precise and accurate in our problem formulation we simply keep the entire expression, although the approximation does not actually change too much graphically.)

However, we consider that all of the errors in our problem are independent and identically distributed such that $\epsilon \sim N(0, \Sigma_{\epsilon\epsilon}$. If the errors were not assumed to have zero mean, then calculating the mean of observations y would become more complicated, and otherwise we would see that the mean of the predictive distribution in our posterior would be shifted elsewhere (either farther up or farther down, depending on the choice of the mean of ϵ). We assume that our errors have mean zero so that they are essentially centered on our regression line. Also, if our errors were not independent, and perhaps depended on some other feature like the year at hand, then we may see either decreasing or increasing error bars depending on the year, which would certainly alter my conclusions about the level of confidence (i.e., the distribution) surrounding the posterior regression coefficients. The constant covariance matrix also could have been adjusted otherwise in the code manually, as I did somewhat in the code anyway, and altering this gives a much different picture of the spread we expect to see in our regression coefficient distribution when plotting the posterior predictive distribution, which again might alter our confidence about the conclusions we made regarding whether the suicide rate is uptrending or downtrending. Note that the properties of Gaussian random variables allow us to make so many conclusions about extracting marginal coefficients (like c_1 in order to compute probabilities over it) and about the conditional distributions involved. There is not much to assume aside from a Gaussian distribution, which is why such an assumption is so effective.

Finally, if we consider that we'd chosen another prior density, then again our data and results may have been different. Our prior clearly expresses the distributional assumptions we make about the regression coefficients c_0 and c_1 based on our a priori knowledge about their distribution on a global scale. Interestingly, although this was not performed too often in class, we adopted an empirical Bayes estimation strategy in order to estimate the prior parameters a priori as well as possible before feeding them into our U.S. regression model. If we had simply chosen our parameters by eye, or simply by some other informal knowledge about suicide rates in the U.S., then there is a chance that our posterior predictive variances would be different (and we would see different error bars than we see now), and also there is a chance that the mean would not have been quite as accurate given that we would be quite far off in the first attempt. However, we do have a fairly significant amount of data to be used from the U.S., if we consider each individual suicide an observation, and so the prior may simply get washed out anyway, in which case our likelihood dominates and, regardless of whether we adopt an empirical Bayes estimation approach, our posterior predictive distribution on the U.S. data, for all intents and purposes, is the same, along with the conclusions we make. Empirical Bayes estimation is an interesting idea because one is essentially re-using data in the actual main solution (finding the posterior on the U.S. data) since that data is incorporated into the estimation of the prior on the U.S. data (i.e., when we find the prior and posterior on the global data). That process may be counter-intuitive for some, but at the same time it is quite powerful seeing as our global dataset is much larger than that in the U.S., and the U.S. trends do not necessarily match those of the world more broadly, so using global data is at the very least an informed guess—perhaps more informed than the naked human eye would be able to accomplish by examining a plot of suicide rate anomalies—for our U.S. prior.

Finally, we consider whether expectations and variances on our posterior predictive distribution may give us any further insight into answer to our main question. We have a few interesting quantities to report. First, note that our posterior mean vector is given by

$$c_{post} = \begin{bmatrix} 0.4864 \\ -0.02898 \end{bmatrix}.$$

One can clearly see that there is a relatively positive intercept in our regression coefficient vector, which might indicate that we actually started out with a higher-than-normal suicide rate around 1985. But one also can note that the slope coefficient is given by -0.02898, which, while seemingly small, actually corresponds to a quite significant deviation over the course of 30 years since we are looking at deviations with units of "per-100,000" people here. Otherwise, note that our covariance matrix is given by

$$\Sigma_{post} = \begin{bmatrix} 0.0563 & -0.0028 \\ -0.0028 & 0.0002 \end{bmatrix}.$$

So, we see that the intercept coefficient c_0 has quite high variance (0.0563) relative to the slope coefficient c_1 (with variance 0.0002). This illustrates that we may have some significant remaining uncertainty in where exactly our intercept term should lie, so perhaps the starting point for our mean values in our posterior predictive distribution is not entirely clear and may thus be generating bias in our results. However, fortunately, we have very low variance in our slope coefficient c_1, and so the general trajectory of our suicide rate anomalies over time is a metric we can be more confident in when assessing U.S. suicide rate trends from 1985 through 2015.

Finally, note that we find (in our accompanying Python code, toward the end of the document) that the probability that the posterior slope is less than -0.025 moves all the way down to 0.614. This makes sense in the context of the mean and variance for the slope coefficient that we calculated in the previous paragraph, but it also reminds us that the slope here is not remarkably negative—we see some sort of averaging out over the past 30 years since there was a clear period of decrease followed by increase. But it is still significant to consider that there is a considerable chance our true regression slope is less than -0.025, which is a clear downward trend considering our units of suicide rate anomalies on the scale of "per-100,000" people. We also note that the probability that our posterior intercept is greater than 0 is given by 0.980, so despite our earlier concerns about variance, we can be very confident that we do have a positive intercept, and so it may very well be true that there were relatively elevated suicide levels in the U.S. around 1985.

We therefore conclude that there has been a general downtrend in suicide rate from 1985 through 2015. This has been bolstered by ample Bayesian regression analyses and makes sense in the context of graphical analyses and what is generally known about suicide rate trends in recent years. However, an interesting note for future exploration (and certainly to keep an eye on) is the relatively recent spike in suicide rate that of course increases our regression slope a lot more than if the downtrend from 1985 to 1999 had kept going, but unfortunately we see that there is an uptick in recent years. This must be addressed. Despite the general downward trend we see smooth out from 1985 through 2015, it is also important to consider immediate and recent trends that may be explained by rapidly changing global conditions. This is a critical issue that must be addressed across fields and disciplines; whether among social workers, statisticians, psychologists and psychiatrists, business owners, or pediatric doctors.

Works Cited

Hedegaard, Holly, M.D., et al. "Increase in Suicide Mortality in the United States, 1999-2018." Centers for Disease Control and Prevention, Apr. 2020, https://www.cdc.gov/nchs/products/databriefs/db362.htm. Accessed 13 Aug. 2020.

Peng, Yifan. "The Best Way to Place Figures Side-by-Side in LaTeX." P Guru by Yifan Peng, edited by Yifan Peng, 13 Sept. 2013, http://blog.pengyifan.com/the-best-way-to-place-figures-side-by-side-in-latex/. Accessed 13 Aug. 2020.

Robinson, David. "Understanding Empirical Bayes Estimation (Using Baseball Statistics)." Variance Explained, 1 Oct. 2015, http://varianceexplained.org/r/empirical_bayes_baseball/. Accessed 13 Aug. 2020.

Rusty. "Suicide Rates Overview 1985 to 2016." Kaggle, https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/data#. Accessed 13 Aug. 2020.