STAT 5444: Homework #4

Due on November 118 2016

Professor Scott Leman

Kevin Malhotra

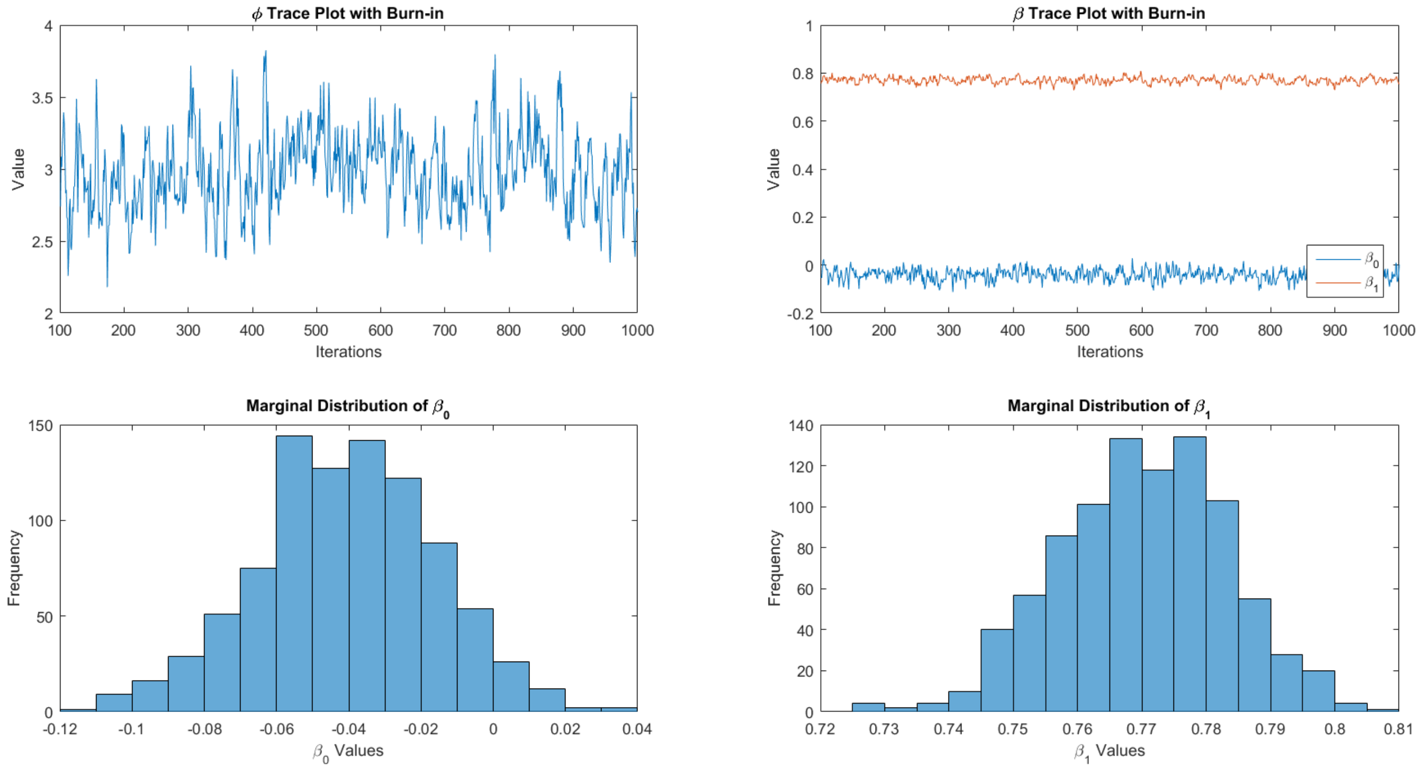
Problem 1

Part 1:

$$
\begin{aligned}
&y_i = \beta_0 + \beta_1 x_i + \epsilon_i (i = 1, ..., N) \rightarrow \epsilon_i = y_i - \beta_0 - \beta_1 x_i \rightarrow \epsilon_i = y_i - \beta x_i^T, \beta = [\beta_0, \beta_1] \\
&\gamma_i - Gamma(\frac{1}{2}, \frac{1}{2}) \\
&\epsilon - N(0, \sigma^2/\gamma_i) \\
&\phi = 1\sigma^2 \\
P(\beta, \phi, \vec{\gamma}|Y, X) &\propto L(\beta, \phi, \vec{\gamma}|Y, X) * P(\vec{\gamma}) * P(\phi) \\
P(\beta, \phi, \vec{\gamma}|Y, X) &\propto \left[\prod_{i=1}^{N} \gamma_i^{1/2} \phi^{1/2} e^{\frac{-\phi\gamma_i}{2}(y_i - x_i^T\beta)^2}\right] * \gamma_i^{\frac{1}{2}-1} e^{\frac{-1}{2}\gamma_i} * \phi^{-1} \\
P(\beta, \phi, \vec{\gamma}|Y, X) &\propto \left[\prod_{i=1}^{N} \gamma_i^{1/2} \phi^{1/2} e^{\frac{-\phi\gamma_i}{2}(y_i - x_i^T\beta)^2}\right] * \gamma_i^{\frac{1}{2}-1} e^{\frac{-1}{2}\gamma_i} * \phi^{-1} \\
P(\beta, \phi, \vec{\gamma}|Y, X) &\propto \phi^{N/2} \left[\prod_{i=1}^{N} \gamma_i^{1/2} e^{\frac{-\phi\gamma_i}{2}(y_i - x_i^T\beta)^2}\right] * \gamma_i^{\frac{1}{2}-1} e^{\frac{-1}{2}\gamma_i} * \phi^{-1} \\
&\Gamma = \begin{bmatrix} \gamma_1 & 0 & \dots & 0 \\ 0 & \gamma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \gamma_N \end{bmatrix} \\
P(\beta, \phi, \vec{\gamma}|Y, X) &\propto \phi^{N/2} \left[\prod_{i=1}^{N} \gamma_i^{1/2}\right] e^{\frac{-\phi}{2}\sum_{i=1}^{N}\gamma_i(y_i - x_i^T\beta)^2} * \gamma_i^{\frac{1}{2}-1} e^{\frac{-1}{2}\gamma_i} * \phi^{-1} \\
P(\beta, \phi, \vec{\gamma}|Y, X) &\propto \phi^{N/2}|\Gamma|^{1/2} e^{\frac{-\phi}{2}(Y - X^T\beta)^T\Gamma(Y - X^T\beta)} * \gamma_i^{\frac{1}{2}-1} e^{\frac{-1}{2}\gamma_i} * \phi^{-1} \\
P(\beta, \phi, \vec{\gamma}|Y, X) &\propto \phi^{N/2}|\Gamma|^{1/2} e^{\frac{-\phi}{2}\left[Y^T\Gamma Y - 2\beta^T X^T\Gamma Y + \beta^T X^T\Gamma X B\right]} * \gamma_i^{\frac{1}{2}-1} e^{\frac{-1}{2}\gamma_i} * \phi^{-1} \\
P(\beta|-) &\propto e^{\frac{-\phi}{2}\left[Y^T\Gamma Y - 2\beta^T X^T\Gamma Y + \beta^T X^T\Gamma X B\right]} \\
P(\beta|-) &\propto e^{\frac{-\phi}{2}\left[\beta^T X^T\Gamma X B - 2\beta^T X^T\Gamma Y\right]} \\
P(\beta|-) &- N(\frac{1}{\phi}(X^T\Gamma X)^{-1} * \phi X^T\Gamma Y, \frac{1}{\phi}(X^T\Gamma X)^{-1}) \\
P(\beta|-) &- N((X^T\Gamma X)^{-1} X^T\Gamma Y, \frac{1}{\phi}(X^T\Gamma X)^{-1}) \\
P(\phi|-) &\propto \phi^{N/2-1} e^{\frac{-\phi}{2}\left[(Y - X^T\beta)^T\Gamma(Y - X^T\beta)\right]} \\
P(\phi|-) &- Gamma(N/2, \frac{1}{2} * (Y - X^T\beta)^T\Gamma(Y - X^T\beta)) \\
P(\gamma_i|-) &\propto \left[\gamma_i^{1/2}\right] e^{\frac{-\phi}{2}\gamma_i(y_i - x_i^T\beta)^2} * \gamma_i^{\frac{1}{2}-1} e^{\frac{-1}{2}\gamma_i} \\
P(\gamma_i|-) &\propto \left[\gamma_i^{1-1}\right] e^{\frac{-\phi}{2}\gamma_i(y_i - x_i^T\beta)^2 + \frac{-1}{2}\gamma_i} \\
P(\gamma_i|-) &- Gamma(1, \frac{\phi}{2}(y_i - x_i^T\beta)^2 + \frac{1}{2})
\end{aligned}
$$

Problem 1 continued on next page. . .

Part 2:

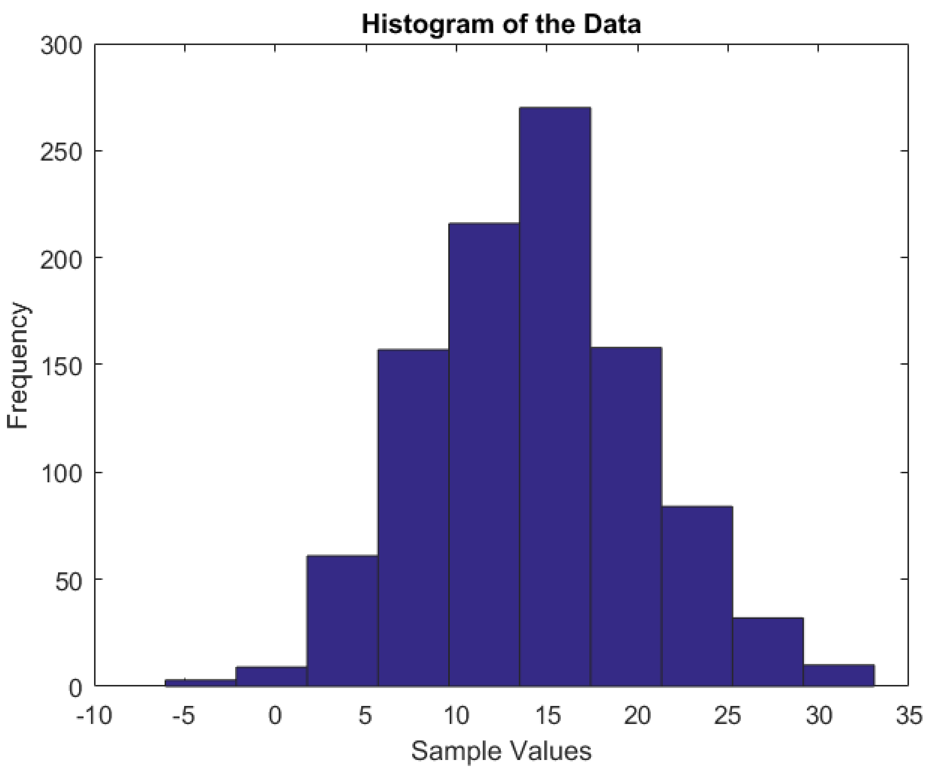



```
%PART 1 %%%%%%%%%%%%%%%%%%
%C = [1 0.8; 0.8 1;];
%cauchy = mvtrnd(C, 1, 1000);
%X = [ones(size(cauchy, 1), 1) cauchy(:, 1)]; Y = cauchy(:, 2);
%figure(1)
%scatter(cauchy(:, 1), cauchy(:, 2));
%title('1000 Pgoint Cauchy Simulation')
%xlabel('X Coordinate')
%ylabel('Y Coordinate')

% Hyperparameter Tuning
T = 1000;
Burnin = 100;
phi0 = 1;
beta0 = [1, 1];
N = size(Y, 1);

gamma = zeros(T, N);
beta = zeros(T+1, 2); beta(1, :) = beta0;
phi = zeros(T+1, 1); phi(1) = phi0;
% Note: Take Inverse of Gamma dist beta term because matlab sucks
for t=1:T
    gamma(t, :) = gamrnd(1,((phi(t)/2).*(Y - X * beta(t, :)').^2 + 0.5).^(-1));
    Gamma = diag(gamma(t, :));
    beta(t+1, :) = mvnrnd(inv(X'*Gamma*X)*X'*Gamma*Y, (1/phi(t))*inv(X'*Gamma*X));
    %x = X(:, 1);
```

```
    %beta(t+1,1) = normrnd((x'*Gamma*Y)*(inv((x'*Gamma*x))),sqrt(phi(t)*x'*Gamma*x)^(-1));
    %x = X(:, 2);
    %beta(t+1,2) = normrnd((x'*Gamma*Y)*(inv((x'*Gamma*x))),sqrt(phi(t)*x'*Gamma*x)^(-1));
    phi(t+1) = gamrnd(N/2, (0.5*((Y - X*beta(t+1, :)')' * Gamma * (Y - X*beta(t+1, :)')))^(-1));
end


figure(1);
clf;
subplot(2, 2, 1)
plot(Burnin+1:T, phi(Burnin+1:T));
xlabel('Iterations'); ylabel('Value'); title('\phi Trace Plot with Burn-in');
subplot(2, 2, 2)
plot(Burnin+1:T, beta(Burnin+1:T, :));
xlabel('Iterations'); ylabel('Value'); title('\beta Trace Plot with Burn-in');
legend('\beta_0', '\beta_1', 'Location', 'southeast');
subplot(2, 2, 3)
histogram(beta(Burnin+1:T, 1));
xlabel('\beta_0 Values'); ylabel('Frequency'); title('Marginal Distribution of \beta_0');
subplot(2, 2, 4)
histogram(beta(Burnin+1:T, 2));
xlabel('\beta_1 Values'); ylabel('Frequency'); title('Marginal Distribution of \beta_1');
```

Problem 2

Part A:



Part B:

$P(\theta) = \Pi_0 \delta(\theta = 10) + (1 - \Pi_0) P_{HA}(\theta)$

$P_{HA}(\theta) = \frac{1}{3}\left[\delta(\theta = 15) + \delta(\theta = 17) + \delta(\theta = 20)\right]$

$\Pi_0 = \frac{1}{4}$

Part C:

$$B = \frac{\int_{\theta \in H_0} L(\theta|X) P_{H_0}(\theta) d\theta}{\int_{\theta \in H_A} L(\theta|X) P_{H_A}(\theta) d\theta}$$

$$B = \frac{\sum_{\theta_i}^{\theta \in H_0} L(\theta_i|X)}{\frac{1}{3}\sum_{\theta_i}^{\theta \in H_A} L(\theta_i|X)}$$

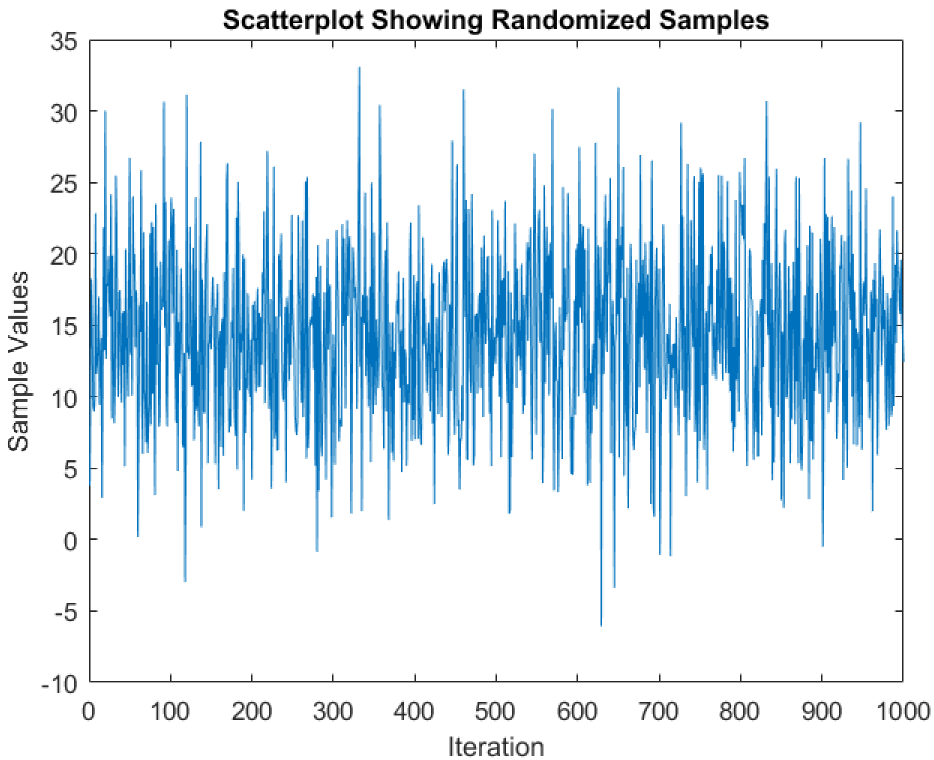
$$H_0 = \left[\theta = 10\right], H_A = \left[\theta = 15, 17, 20\right]$$

$$H_0 = \left[\theta = 15\right], H_A = \left[\theta = 10, 17, 20\right]$$

$$H_0 = \left[\theta = 17\right], H_A = \left[\theta = 15, 10, 20\right]$$

$$H_0 = \left[\theta = 20\right], H_A = \left[\theta = 15, 17, 10\right]$$

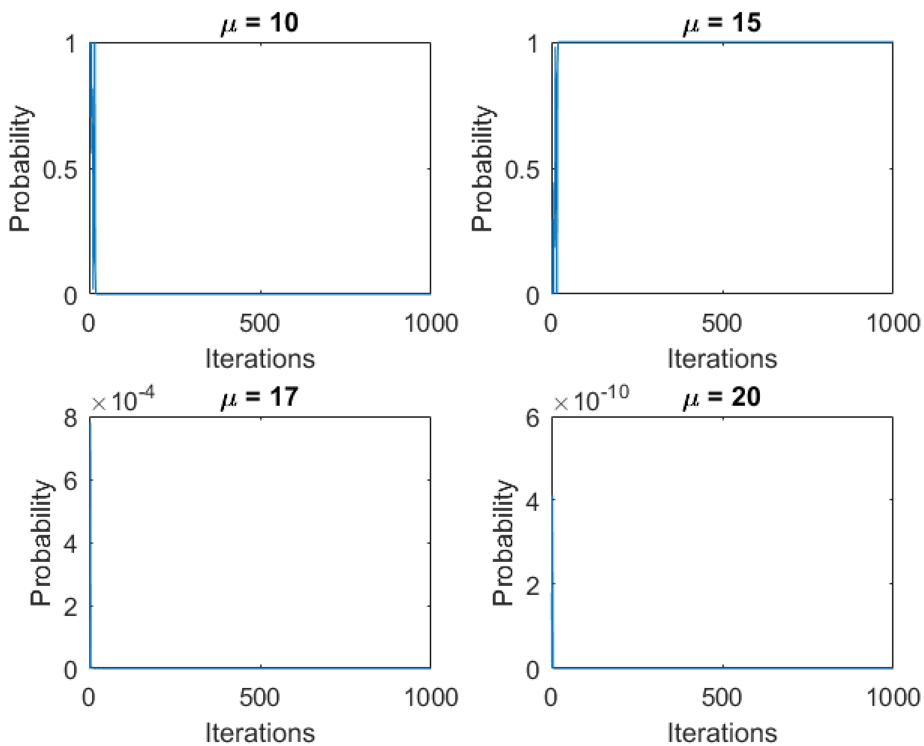
Part D:



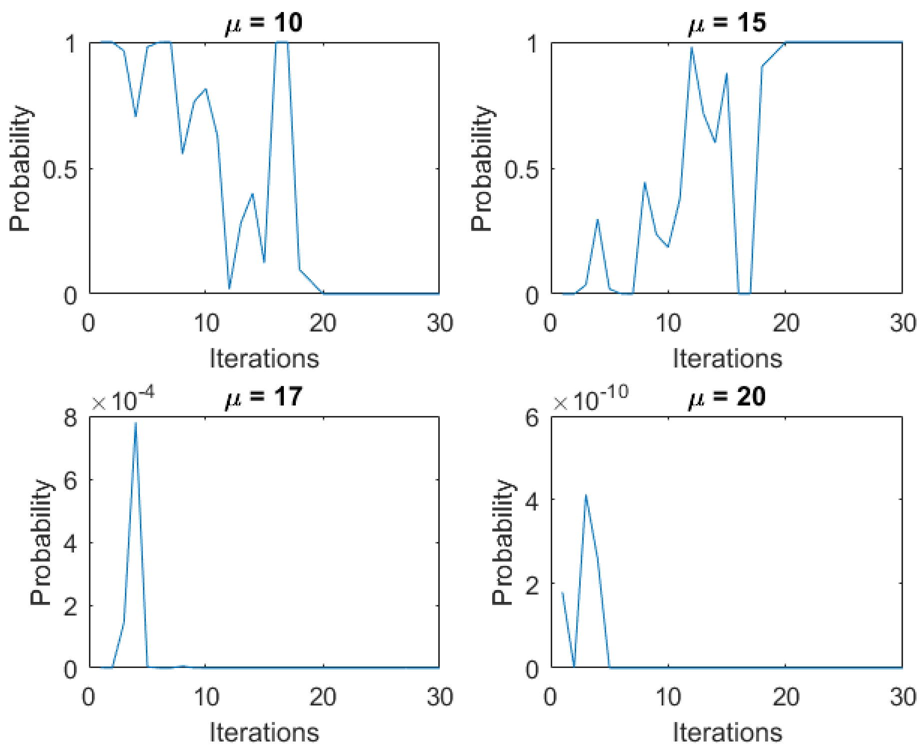
Part E:

$$P(\theta|X) \propto L(\theta|X)P_{H_0}(\theta)$$
$$P(\theta|X) \propto \prod_{i=1}^{N} e^{\frac{-1}{2\sigma^2}(x_i-\mu)^2}$$
$$P(\theta|X) \propto e^{\sum_{i=1}^{N} \frac{-1}{2\sigma^2}\left[(x_i-\bar{x})+(\bar{x}-\mu)\right]^2}$$
$$P(\theta|X) \propto e^{\sum_{i=1}^{N} \frac{-1}{2\sigma^2}(\bar{x}-\mu)^2-2(x_i-\bar{x})(\bar{x}-\mu)+(x_i-\bar{x})(x_i-\bar{x})}$$
$$P(\theta|X) \propto e^{\sum_{i=1}^{N} \frac{-1}{2\sigma^2}(\bar{x}-\mu)^2}$$
$$P(\theta|X) \propto e^{\frac{-N}{2\sigma^2}(\bar{x}-\mu)^2}$$

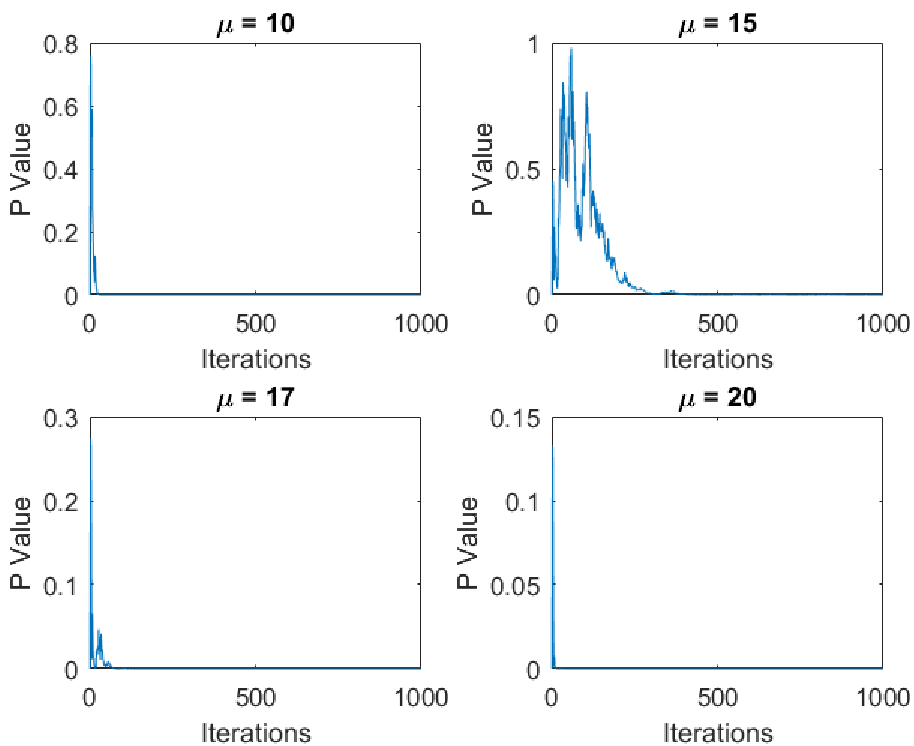
The following function was used to compute the sequenced posteriors. This is numerically stable algorithm.



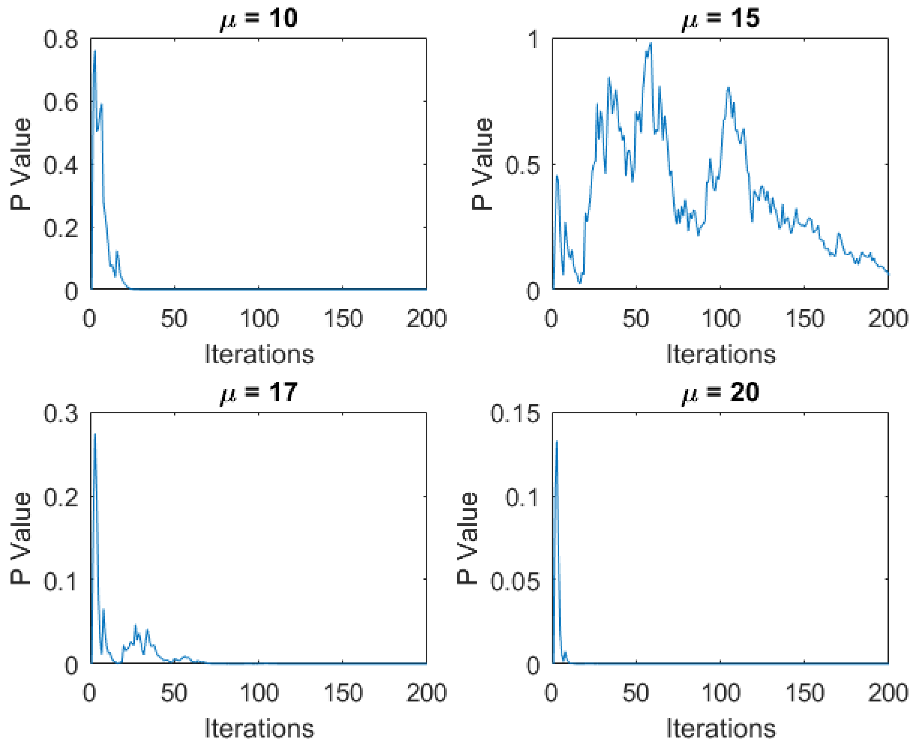
Part E Zoom:



Part F:



Part F Zoom:



Part H:
We can see that P-Values are not a useful metric for describing the true mean. As we increase sample size the we fail to reject the two tail test for all cases. The posterior probabilities kept getting updated with more data points and the mean would slowly approach the ground truth mean. The probability would eventually reach 1 for the mu that would be closest to ground truth.
Code:

```
% Problem 2

% Part A
theta = [10, 15, 17, 20];
sigmaSq = 5;
modelSize = 1000;

mixedModels = zeros(modelSize, 1);
mixedModels(1:400) = normrnd(theta(1), sigmaSq, [400, 1]);
mixedModels(401:600) = normrnd(theta(2), sigmaSq, [200, 1]);
mixedModels(601:900) = normrnd(theta(3), sigmaSq, [300, 1]);
mixedModels(901:end) = normrnd(theta(4), sigmaSq, [100, 1]);
figure(1);
hist(mixedModels); xlabel('Sample Values'); ylabel('Frequency'); title('Histogram of the Data');

% Part D
mixedModels2 = mixedModels(randperm(modelSize));
figure(2);
plot(1:modelSize, mixedModels2); xlabel('Iteration'); ylabel('Sample Values'); title('Scatterplot Showing Randomi

% Part E & F
H0 = theta(1);
prior = zeros(length(theta), modelSize);
```

```
prior(:, 1) = log(1/4);
pval = zeros(length(theta), modelSize);
cummean = cumsum(mixedModels2)./(1:1000)';
prior(1, :) = exp((cummean - theta(1)).^2 .* ((1:1000)'/sigmaSq) .* (-1/2));
prior(2, :) = exp((cummean - theta(2)).^2 .* ((1:1000)'/sigmaSq) .* (-1/2));
prior(3, :) = exp((cummean - theta(3)).^2 .* ((1:1000)'/sigmaSq) .* (-1/2));
prior(4, :) = exp((cummean - theta(4)).^2 .* ((1:1000)'/sigmaSq) .* (-1/2));

for i=1:modelSize
    pval(:, i) = 2*(tcdf(-abs((mean(mixedModels2(1:i)) - theta))/(std(mixedModels2(1:i))/sqrt(i)), i));
end
normalize = sum(prior, 1);
prior(1, :) = prior(1, :) ./ normalize;
prior(2, :) = prior(2, :) ./ normalize;
prior(3, :) = prior(3, :) ./ normalize;
prior(4, :) = prior(4, :) ./ normalize;

%Part E
% Plot Posteriors
figure(3);
tmp = modelSize;
for i=1:length(theta)
    subplot(2, 2, i)
    plot(1:tmp, prior(i, 1:tmp));
    title(['\mu = ', num2str(theta(i))]);
    xlabel('Iterations'); ylabel('Probability');
end
% Part F
% Plot P Values
figure(4);
for i=1:length(theta)
    subplot(2, 2, i)
    plot(1:tmp, pval(i, 1:tmp));
    title(['\mu = ', num2str(theta(i))]);
    xlabel('Iterations'); ylabel('P Value');
end
```

Problem 3

$$P(H_0|X) = \frac{\int_{p \in H_0} (X|p)P(p)dP}{\int_{p \in H_0, H_A} P(p)dP}$$

$$P(X) = \int_{p \in H_0, H_A} P(X|p)P(p)$$

$$P(X) = \int_{p \in H_0, H_A} \left[\binom{n}{k} p^x (1-p)^{n-x}\right]\left[\gamma(p=0.5) * \frac{1}{1} + p^{\frac{3}{2}-1}(1-p)^{\frac{3}{2}-1}\frac{1}{\beta(\frac{3}{2},\frac{3}{2})}\right] * 0.5dp$$

$$P(X) = \int_{p \in H_0, H_A} \left[\binom{n}{k} p^x (1-p)^{n-x} \gamma(p=0.5) * 0.5\right] + \left[\binom{n}{k} p^{x+\frac{3}{2}-1}(1-p)^{n-x+\frac{3}{2}-1}\frac{1}{\beta(\frac{3}{2},\frac{3}{2})} * 0.5\right] dp$$

$$P(X) = \int_{p \in H_0, H_A} \left[\binom{n}{k} p^x (1-p)^{n-x} \gamma(p=0.5) * 0.5\right] + \left[\binom{n}{k} \frac{\beta(x+\frac{3}{2}-1, n-x+\frac{3}{2}-1)}{\beta(\frac{3}{2},\frac{3}{2})} * 0.5\right] dp$$

$$P(X) = \int_{p \in H_0, H_A} \left[\binom{n}{k} p^x (1-p)^{n-x} \gamma(p=0.5) * 0.5\right] dp + 0.5$$

$$P(X) = \left[\binom{n}{k} (p=0.5)^x (1-(p=0.5))^{n-x} * 0.5\right] + 0.5$$

$$P(X) = \left[\binom{n}{k} (0.5)^x (0.5)^{n-x} * 0.5\right] + 0.5$$

$$P(X) = \left[\binom{n}{k} (0.5)^n * 0.5\right] + 0.5$$

$$P(X) = 0.5 * \left[\binom{n}{k} (0.5)^n + 1\right]$$

$$P(H_0|X) = \frac{\int_{p \in H_0} (X|p)P(p)dP}{\int_{p \in H_0, H_A} P(p)dP}$$

$$P(H_0|X) = \frac{\int_{p \in H_0} \left[\binom{n}{k} p^x (1-p)^{n-x}\right]\left[\gamma(p=0.5) * 0.5\right]}{0.5 * \left[\binom{n}{k}(0.5)^n + 1\right]}$$

Problem 4

$$p = P(X \geq t) = 1 - P(X < t) = 1 - F(t) = 1 - z$$
$$z = F(t)$$

The following information is describing the p-value as a continuous random variable and t an observed measurement. F(X) and F(t) are the cdfs since cdfs are continuous and monotonically increasing we can write the following.

$$
\begin{aligned}
P(X \geq t) &= P(F(X) \geq F(t)) = P(F(X) \geq z) = 1 - z \\
P(X \geq t) &= P(F(X) \geq F(t)) = 1 - P(F(X) < F(t)) = 1 - P(F(X) < z) = 1 - z \\
1 - P(F(X) < z) &= 1 - z \\
z &= P(F(X) < z)
\end{aligned}
$$

This states that no matter what sample t we choose the probability that F(t) is less than the continuous random variable cdf will always be the same. This value is the F(t). Thus, in the null hypothesis the P-Value has a uniform distribution since all F(t) values will be same probability regardless of preexisting distribution or the t value. This is the goal of the p-value in the null space.