Homework 2

Q1

Suppose data is generated from the model $y_i \mid \mu \overset{\text{iid}}{\sim} N(\mu, 1)$ for $i = 1, \ldots, n$. Consider a mixture normal prior:

$$\mu \sim .5N(-1, 1) + .5N(1, 1)$$

that is,

$$p(\mu) = \frac{.5}{\sqrt{2\pi}} e^{-\frac{1}{2}(\mu+1)^2} + \frac{.5}{\sqrt{2\pi}} e^{-\frac{1}{2}(\mu-1)^2}$$

Suppose we have observed $y = 1$. Find the posterior distribution of μ.

1. Start by using Bayes theorem: $p(\mu \mid y) \propto p(y \mid \mu)p(\mu) = p(y \mid \mu)\{.5\phi(\mu; -1, 1) + .5\phi(\mu; 1, 1)\}$ and simplify the two components using the following result.

$$\phi\left(x; \mu_1, \sigma_1^2\right) \phi\left(x; \mu_2, \sigma_2^2\right) = \phi\left(x; \frac{\mu_1/\sigma_1^2 + \mu_2/\sigma_2^2}{1/\sigma_1^2 + 1/\sigma_2^2}, \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}\right) \phi\left(\mu_1 - \mu_2; 0, \sigma_1^2 + \sigma_2^2\right)$$

   where $\phi\left(x; \mu, \sigma^2\right)$ is the density of a normal distribution with mean μ and variance σ^2.

2. The posterior is going to be a mixture of two normal distributions. So you will only need to identify the two mean and variance parameters as well as the weights for the two normal distributions.

Since the likelihood function is based on one single observation:

$$\begin{aligned}\mathcal{L}(Y = 1 | \mu, \sigma^2 = 1) &= \frac{1}{\sqrt{2\pi}} \cdot \exp\left\{\frac{-1}{2} \times \left(\frac{1-\mu}{1}\right)^2\right\} = \frac{1}{\sqrt{2\pi}} \cdot \exp\left\{\frac{-1}{2} \times \left(\frac{\mu-1}{1}\right)^2\right\} \\ &= \phi(\mu, 1, 1)\end{aligned}$$

$$\begin{aligned}\text{Posterior} &\propto \text{Likelihood} \times \text{Prior} \\ &= \phi(\mu, 1, 1) \times \left[\left(\frac{1}{2} \times \phi(\mu; -1, 1)\right) + \left(\frac{1}{2} \times \phi(\mu; 1, 1)\right)\right] \\ &= \left[\phi(\mu, 1, 1) \times \frac{1}{2} \times \phi(\mu; -1, 1)\right] + \left[\phi(\mu, 1, 1) \times \frac{1}{2} \times \phi(\mu; 1, 1)\right]\end{aligned}$$

■ Blue Portion: We calculate $\frac{\mu_1/\sigma_1^2+\mu_2/\sigma_2^2}{1/\sigma_1^2+1/\sigma_2^2} = 0$, $\frac{1}{1/\sigma_1^2+1/\sigma_2^2} = 1/2$, $(\mu_1 - \mu_2) = 2$, and $\sigma_1^2 + \sigma_2^2 = 2$

■ Orange Portion: We calculate $\frac{\mu_1/\sigma_1^2+\mu_2/\sigma_2^2}{1/\sigma_1^2+1/\sigma_2^2} = 1$, $\frac{1}{1/\sigma_1^2+1/\sigma_2^2} = 1/2$, $(\mu_1 - \mu_2) = 0$, and $\sigma_1^2 + \sigma_2^2 = 2$

Then, using the hints, $\left[\phi(\mu, 1, 1) \times \frac{1}{2} \times \phi(\mu; -1, 1)\right] + \left[\phi(\mu, 1, 1) \times \frac{1}{2} \times \phi(\mu; 1, 1)\right]$ becomes:

$$= \left[\phi(\mu, 0, \frac{1}{2}) \times \frac{1}{2} \times \phi(2; 0, 2)\right] + \left[\phi(\mu, 1, \frac{1}{2}) \times \frac{1}{2} \times \phi(0; 0, 2)\right]$$
$$= \left[\phi(\mu, 0, \frac{1}{2}) \times \frac{1}{2} \times \frac{1}{\sqrt{4\pi}} \exp\{-1\}\right] + \left[\phi(\mu, 1, \frac{1}{2}) \times \frac{1}{2} \times \frac{1}{\sqrt{4\pi}}\right]$$

Let ε' represent the new weight for the blue portion above. Normalizing gives us:

$$\varepsilon' = \frac{\cancel{\frac{1}{2}}\cancel{\frac{1}{\sqrt{4\pi}}} \exp\{-1\}}{\cancel{\frac{1}{2}}\cancel{\frac{1}{\sqrt{4\pi}}} \exp\{-1\} + \cancel{\frac{1}{2}}\cancel{\frac{1}{\sqrt{4\pi}}}} = \frac{\exp\{-1\}}{\exp\{-1\} + 1} \approx 0.268941$$
$$(1-\varepsilon)' = 1 - \frac{\exp\{-1\}}{\exp\{-1\} + 1} \approx 0.73106$$

Thus the resulting distribution is a normal mixture posterior of the form:

$$\boxed{\pi(\mu|Y=1) = \left[\left(\frac{\exp\{-1\}}{1+\exp\{-1\}}\right) \times \mathcal{N}(0, \frac{1}{2})\right] + \left[\left(1 - \frac{\exp\{-1\}}{1+\exp\{-1\}}\right) \times \mathcal{N}(1, \frac{1}{2})\right]}$$

or equivalently:

$$\boxed{\pi(\mu|Y=1) = \left[\left(\frac{1}{1+e}\right) \times \mathcal{N}(0, \frac{1}{2})\right] + \left[\left(\frac{e}{e+1}\right) \times \mathcal{N}(1, \frac{1}{2})\right]}$$

Q2

Engineering system of type k-out-of-n is operational if at least k out of n components are operational. Otherwise, the system fails. Suppose that a k-out-of-n system consists of n identical and independent elements for which the lifetime has Weibull distribution with parameters r and λ. More precisely, if T is a lifetime of a component,

$$P(T \geq t) = e^{-\lambda t^r},\ t \geq 0.$$

Time t is in units of months, and consequently, rate parameter λ is in units $(\text{month})^{-1}$. Parameter r is dimensionless.

Assume that $n = 10, k = 7,\ r = 1.3$ and $\lambda = 1/20$.

1. Find the probability that a k-out-of-n system is still operational when checked at time $t = 6$.

   In our situation, the probability of the system working at time t is $p = \exp\{-\frac{1}{20} \cdot 6^{1.3}\}$. Then X, the number of components working is Binomial($n = 10$,$p = \exp\{-\frac{1}{20} \cdot 6^{1.3}\}$) with probability mass function:

$$P(X = x) = \binom{10}{x} \exp\Big\{-\frac{1}{20} \cdot 6^{1.3}\Big\}^x \left(1 - \exp\Big\{-\frac{1}{20} \cdot 6^{1.3}\Big\}\right)^{10-x}$$

We calculate the probability as:

$$
\begin{aligned}
P(X \geq 7) &= P(X \leq 10) - P(X \leq 6) \\
&= 1 - \sum_{x=0}^{6} P(X = x) \\
\boxed{P(X \geq 7) \approx 0.3782}
\end{aligned}
$$

2. At the check up at time $t = 6$ the system was found to be operational. What is the probability that at that time exactly 7 components were operational?

   Let θ be the event that exactly 7 components were operational and let D be the event that the system is operational after a 6-month check. By Bayes' theorem:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

   Since the system was found to be operational at the 6-month check with 7 components working, $P(D|\theta) = 1$. The probability that exactly 7 components are working is given by:

$$P(X = 7) = \binom{10}{7} \exp\Big\{-\frac{1}{20} \cdot 6^{1.3}\Big\}^7 \Big(1 - \exp\Big\{-\frac{1}{20} \cdot 6^{1.3}\Big\}\Big)^{10-7} \approx 0.2135$$

   Returning to Bayes' theorem above:

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)} = \frac{(1) \cdot (0.2135)}{0.3782} \boxed{= 0.5645}$$

3. At the check up at time $t = 6$, the system was found operational. What is the probability that the system would still be operational at the time $t = 9$?

   Let A be the event that the system was operational at time $t = 6$, and B will denote the event that the system was operational at time $t = 9$. We calculated $P(A) = P(X \geq 7) \approx 0.3782$ above in Q1.1. To calculate $P(B)$, we note that that the probability of an individual component failing at time $t = 9$ is:

$$P(T \geq 9) = \exp\{\frac{-1}{20} \times 9^{1.3}\}$$

   The number of components that are functional at time $t = 9$ is another Binomial($n = 7$, $p = \exp\{\frac{-1}{20} \times 9^{1.3}\}$) random variable.

$$P(Y = y) = \binom{10}{y} \exp\Big\{-\frac{1}{20} \cdot 9^{1.3}\Big\}^y \Big(1 - \exp\Big\{-\frac{1}{20} \cdot 9^{1.3}\Big\}\Big)^{10-y}$$

   in another binomial distribution to get $P(B)$:

$$
\begin{aligned}
P(B) &= P(Y \geq 7) = P(Y \leq 10) - P(Y \leq 6) \\
&= 1 - \sum_{x=0}^{6} P(Y = y) \\
P(B) &\approx 0.0703
\end{aligned}
$$

   Then:

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} = \frac{P(B)}{P(A)} = \frac{0.0703}{0.3782} \boxed{= 0.1859}$$

Hint: The probability that a k-out-of-n system is operational corresponds to the tail probability of binomial distribution: $P(X \geq k)$, where X is the number of components working. You can do exact binomial calculations or use `binocdf` in Octave/MATLAB (or `dbinom` in R, or `scipy.stats.binom.cdf` in Python). Be careful with $\leq$ and $<$, because of the discrete nature of the binomial distribution. Part 2 is straightforward Bayes formula. Part 3 is the total probability with hypotheses whose probabilities are obtained as in (b).

Q3

From the first page of Rand's book *A Million Random Digits with 100,000 Normal Deviates.*

31060	10805	45571	82406	35303	42614	86799	07439	23403	09732
85269	77602	02051	65692	68665	74818	73053	85247	18623	88579
63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
73796	45753	03529	64778	35808	34282	60935	20344	35273	88435
98520	17767	14905	68607	22109	40558	60970	93433	50500	73998

The second 50 five-digit numbers form the Rand's "A Million Random Digits with 100,000 Normal Deviates" book (shown above) are rescaled to $[0, 1]$ (by dividing by 100,000) and then all numbers < 0.7 are retained. We can consider the $n = 35$ retained numbers as a random sample from uniform $\mathcal{U}(0, 0.7)$ distribution.

0.3106	0.10805	0.45571	0.35303	0.42614	0.07439	0.23403
0.09732	0.02051	0.65692	0.68665	0.18623	0.63573	0.32135
0.05325	0.47048	0.57548	0.28468	0.28709	0.25624	0.45753
0.03529	0.64778	0.35808	0.34282	0.60935	0.20344	0.35273
0.17767	0.14905	0.68607	0.22109	0.40558	0.60970	0.50500

Pretend now that the threshold 0.7 is not known to us, that is, we are told that the sample is from uniform $\mathcal{U}(0, \theta)$ distribution, with θ to be estimated.

Let M be the maximum of the retained sample $u_1, \ldots, u_{35}$, in our case $M = 0.68665$. The likelihood is

$$f\left(u_1, \ldots, u_{35} \mid \theta\right) = \prod_{i=1}^{35} \frac{1}{\theta} \mathbf{1}\left(\theta > u_i\right) = \theta^{-35} \mathbf{1}(\theta > M)$$

where $\mathbf{1}(A)$ is 1 if A is true, and 0 if A is false.

Assume noninformative (Jeffreys') prior on θ,

$$\pi(\theta) = \frac{1}{\theta} \mathbf{1}(\theta > 0)$$

Posterior depends on data via the maximum M and belongs to the Pareto family, $\mathcal{P}a(c, \alpha)$, with a density

$$\frac{\alpha c^{\alpha}}{\theta^{\alpha+1}} \mathbf{1}(\theta > c)$$

TABLE OF RANDOM DIGITS 1

00000	10097	32533	76520	13586	34673	54876	80959	09117	39292	74945
00001	37542	04805	64894	74296	24805	24037	20636	10402	00822	91665
00002	08422	68953	19645	09303	23209	02560	15953	34764	35080	33606
00003	99019	02529	09376	70715	38311	31165	88676	74397	04436	27659
00004	12807	99970	80157	36147	64032	36653	98951	16877	12171	76833
00005	66065	74717	34072	76850	36697	36170	65813	39885	11199	29170
00006	31060	10805	45571	82406	35303	42614	86799	07439	23403	09732
00007	85269	77602	02051	65692	68665	74818	73053	85247	18623	88579
00008	63573	32135	05325	47048	90553	57548	28468	28709	83491	25624
00009	73796	45753	03529	64778	35808	34282	60935	20344	35273	88435
00010	98520	17767	14905	68607	22109	40558	60970	93433	50500	73998
00011	11805	05431	39808	27732	50725	68248	29405	24201	52775	67851
00012	83452	99634	06288	98083	13746	70078	18475	40610	68711	77817
00013	88685	40200	86507	58401	36766	67951	90364	76493	29609	11062
00014	99594	67348	87517	64969	91826	08928	93785	61368	23478	34113
00015	65481	17674	17468	50950	58047	76974	73039	57186	40218	16544
00016	80124	35635	17727	08015	45318	22374	21115	78253	14385	53763
00017	74350	99817	77402	77214	43236	00210	45521	64237	96286	02655
00018	69916	26803	66252	29148	36936	87203	76621	13990	94400	56418
00019	09893	20505	14225	68514	46427	56788	96297	78822	54382	14598
00020	91499	14523	68479	27686	46162	83554	94750	89923	37089	20048
00021	80336	94598	26940	36858	70297	34135	53140	33340	42050	82341
00022	44104	81949	85157	47954	32979	26575	57600	40881	22222	06413
00023	12550	73742	11100	02040	12860	74697	96644	89439	28707	25815

Figure 1: First page of RAND's book.

1. What are α and c?

   Start with the likelihood times prior.

$$\pi(\theta \mid u_1, \ldots, u_{35}) \propto \theta^{-35} \frac{1}{\theta} \mathbf{1}(\theta > M) \mathbf{1}(\theta > 0) = \theta^{-36} \mathbf{1}(\theta > M)$$

   Compare it with the given posterior PDF:

$$\pi(\theta \mid \text{data}) = \frac{\alpha c^{\alpha}}{\theta^{\alpha+1}} \mathbf{1}(\theta > c)$$

   We need to ensure normalization, meaning that:

$$\int_{\Theta} \frac{\alpha c^{\alpha}}{\theta^{\alpha+1}} \mathbf{1}(\theta > c) d\theta = 1$$

   But we don't want to calculate this integral. Instead, we recognize that

$$\frac{1}{\theta^{35+1}} \mathbf{1}(\theta > M) \quad \text{and} \quad \frac{1}{\theta^{\alpha+1}} \mathbf{1}(\theta > c)$$

   are of the same form. Our posterior is Pareto$(c = 0.68665, \alpha = 35)$.

2. Estimate θ in Bayesian fashion. Then calculate the 95% equi-tailed credible set. Is the true value of parameter (0.7) in the credible set?

   We first want a single point estimate from our posterior distribution. Common estimators are the mean, median, or MAP (posterior mode).

$$\hat{\theta}_{\text{mode}} = c = 0.68665, \quad \hat{\theta}_{\text{mean}} = \frac{\alpha\, c}{\alpha - 1} = 0.70685, \quad \hat{\theta}_{\text{median}} = c\, 2^{1/\alpha} = 0.70038$$

The equi-tailed credible set can be calculated using the inverse CDF. If the Pareto(c, α) distribution has CDF (for $x \geq c$)

$$F(x \mid c, \alpha) \;=\; 1 \;-\; \left(\frac{c}{x}\right)^{\alpha},$$

then the inverse CDF (quantile function) is found by setting $u = F(x)$ and solving for x:

$$u \;=\; 1 - \left(\tfrac{c}{x}\right)^{\alpha} \quad \Longrightarrow \quad 1 - u \;=\; \left(\tfrac{c}{x}\right)^{\alpha} \quad \Longrightarrow \quad \left(\tfrac{c}{x}\right) \;=\; (1-u)^{1/\alpha} \quad \Longrightarrow \quad x \;=\; \frac{c}{(1-u)^{1/\alpha}}.$$

Hence,

$$F^{-1}(u \mid c, \alpha) \;=\; c\,(1-u)^{-\frac{1}{\alpha}}, \quad 0 < u < 1.$$

$$F^{-1}(0.025) = 0.68715, F^{-1}(0.975) = 0.76297$$

Our credible set is $(0.68715, 0.76297)$, which does contain the true value of θ.

3. Plot the posterior PDF, adding marks for the regions bound by the above credible set, along with your point estimate, for each plot.

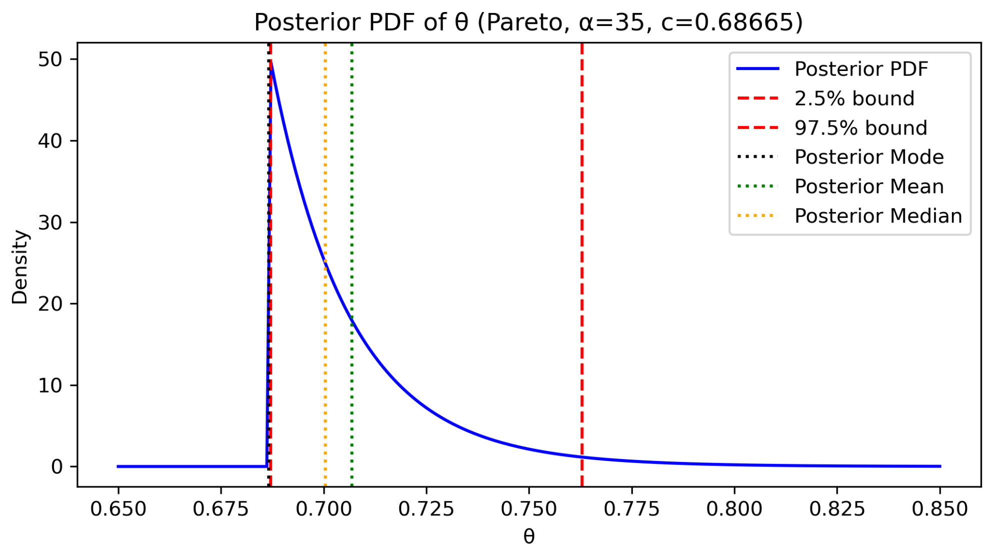

Figure 2: Posterior distributions under different Pareto priors (blue) compared with the Jeffreys prior (dashed red).

4. Experiment by replacing the Jeffrey's prior with increasingly informative Pareto priors. Start with $c < M$ and very small α. Report what happens to the posterior when varying the Pareto prior parameters and compare them to the Jeffrey's prior model.

   Find some different posterior distributions using the conjugate pair from U4L5. Notice that if $c < M$ the MAP will always be at M. If $c > M$ the MAP will be at c. Very small α will be equivalent to the Jeffreys prior version from parts 1-3. Very large α will concentrate the density near c.

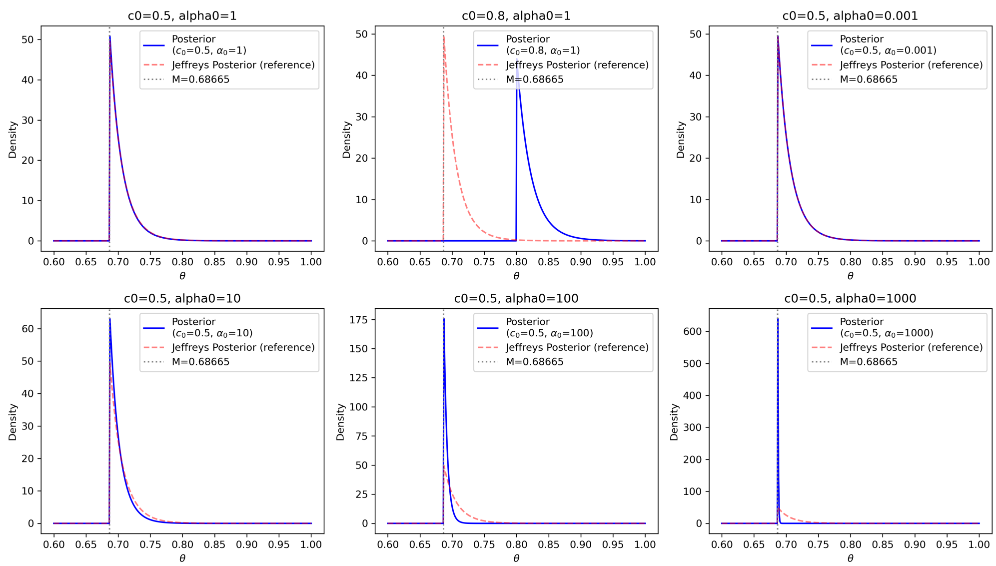

Figure 3: Posterior distributions under different Pareto priors (blue) compared with the Jeffreys prior (dashed red).