Statistical Inference 2 Assignment 1

Master of Statistics (M.Stat.) Year II, Session 2020-21

Name: Subhrajyoty Roy
Roll: MB1911

December 16, 2020

Problem 1. Consider the problem of estimation of a real parameter θ with the loss function

$$L(\theta, a) = \begin{cases} K_0(\theta - a) & \text{if } \theta - a \geq 0 \\ K_1(a - \theta) & \text{if } \theta - a < 0 \end{cases}$$

Show that the Bayes estimate is given by the quantile of order $K_0/(K_0 + K_1)$ of the posterior distribution (assume, for simplicity, uniqueness of the quantile).

Solution. Let $a \leq b$ be two real numbers, then,

$$\begin{aligned} L(\theta, a) - L(\theta, b) &= \begin{cases} K_0(\theta - a) - K_0(\theta - b) & \text{if } a \leq b \leq \theta \\ K_0(\theta - a) - K_1(b - \theta) & \text{if } a \leq \theta < b \\ K_1(a - \theta) - K_1(b - \theta) & \text{if } \theta < a \leq b \end{cases} \\ &= \begin{cases} K_0(b - a) & \text{if } a \leq b \leq \theta \\ (K_0 + K_1)\theta - (K_0 a + K_1 b) & \text{if } a \leq \theta < b \\ K_1(a - b) & \text{if } \theta < a \leq b \end{cases} \end{aligned}$$

Now note that, when $a \leq \theta < b$, $(K_0 + K_1)\theta \geq (K_0 + K_1)a$. Therefore,

$$L(\theta, a) - L(\theta, b) \geq K_1(a - b)\mathbf{1}_{\{\theta < b\}} + K_0(b - a)\mathbf{1}_{\{\theta \geq b\}} \qquad \forall a \leq b$$

On the other hand, if $a > b$, then also

$$\begin{aligned} L(\theta, a) - L(\theta, b) &= \begin{cases} K_0(\theta - a) - K_0(\theta - b) & \text{if } b \leq a \leq \theta \\ K_1(a - \theta) - K_0(\theta - b) & \text{if } b \leq \theta < a \\ K_1(a - \theta) - K_1(b - \theta) & \text{if } \theta < b \leq a \end{cases} \\ &= \begin{cases} K_0(b - a) & \text{if } b \leq a \leq \theta \\ (K_1 a + K_0 b) - (K_0 + K_1)\theta & \text{if } b \leq \theta < a \\ K_1(a - b) & \text{if } \theta < b \leq a \end{cases} \\ &\geq K_1(a - b)\mathbf{1}_{\{\theta < b\}} + K_0(b - a)\mathbf{1}_{\{\theta \geq b\}} \end{aligned}$$

Together, we have for any $a, b \in \mathbb{R}$,

$$L(\theta, a) - L(\theta, b) \geq K_1(a - b)\mathbf{1}_{\{\theta < b\}} + K_0(b - a)\mathbf{1}_{\{\theta \geq b\}} \tag{1}$$

Taking expectation of both sides of Eq. (1) with respect to the posterior distribution of θ, we obtain

$$
\begin{aligned}
R_{\pi|x}(\theta,a) - R_{\pi|x}(\theta,b) &= \mathbb{E}_{\pi|x}\left[L(\theta,a) - L(\theta,b)\right] \\
&\geq \mathbb{E}_{\pi|x}\left[K_1(a-b)\mathbf{1}_{\{\theta<b\}} + K_0(b-a)\mathbf{1}_{\{\theta\geq b\}}\right] \\
&= K_1(a-b)\mathbb{P}_{\pi|x}(\theta<b) + K_0(b-a)\mathbb{P}_{\pi|x}(\theta\geq b) \\
&= (a-b)\left[K_1\mathbb{P}_{\pi|x}(\theta<b) - K_0(1-\mathbb{P}_{\pi|x}(\theta<b))\right] \\
&= (a-b)\left[(K_0+K_1)\mathbb{P}_{\pi|x}(\theta<b) - K_0\right]
\end{aligned}
$$

Clearly, if b is chosen to be the $K_0/(K_0+K_1)$-th quantile of the posterior distribution, say $Q_{K_0/(K_0+K_1)}$, then this lower bound turns out to be equal to 0, which means,

$$
R_{\pi|x}(\theta,a) - R_{\pi|x}(\theta, Q_{K_0/(K_0+K_1)}) \geq 0,\ \forall a \in \mathbb{R}
$$

where the equality holds if and only if $a = Q_{K_0/(K_0+K_1)}$. This shows that the Bayes estimate with respect to the given loss function is given by $Q_{K_0/(K_0+K_1)}$, the quantile of order $K_0/(K_0+K_1)$ of the posterior distribution.

Problem 2. Given $0 < \theta < 1$, let $X_1, \cdots, X_n$ be i.i.d. $\text{Bin}(1, \theta)$. Consider the Jeffreys prior for θ. Find by simulation the frequentist coverage of θ by the two-tailed 95% credible interval for $\theta = \frac{1}{8}, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, \frac{7}{8}$. Do the same for the usual frequentist interval $\hat{\theta} \pm z_{0.025}\sqrt{\hat{\theta}(1-\hat{\theta})/n}$ where $\hat{\theta} = \sum_i X_i/n$.

Solution. We know that for the binomial setup, Jeffrey's prior is given by

$$\pi(\theta) = \frac{1}{\sqrt{\theta(1-\theta)}}, \qquad 0 < \theta < 1$$

Therefore, the posterior is

$$\begin{aligned}
\pi(\theta \mid X_1, \dots X_n) &\propto \pi(\theta) \prod_{i=1}^{n} f(X_i \mid \theta) \\
&= \frac{1}{\sqrt{\theta(1-\theta)}} \theta^{\sum_i X_i} (1-\theta)^{(n - \sum_i X_i)} \\
&= \theta^{(\sum_i X_i - 1/2)} (1-\theta)^{(n + \sum_i X_i - 1/2)}
\end{aligned}$$

which is proportional to a beta density with parameters $(\sum_i X_i + 1/2)$ and $(n - \sum_i X_i + 1/2)$. To find the HPD credible interval of the posterior distribution of θ, a numerical algorithm is employed. We know that HPD credible interval with confidence coefficient α is a set C of the form

$$C = \{\theta : \pi(\theta \mid X_1, \dots X_n) \geq K(\alpha)\}$$

where $K(\alpha)$ is the largest constant such that $\mathbb{P}(\theta \in C \mid X) \geq \alpha$. To obtain this, given a value of the right endpoint of the credible interval say r, we can obtain the left endpoint l as the $(\mathbb{P}(\theta \leq r \mid X_1, \dots X_n) - \alpha)$-th quantile of the posterior beta distribution (say denoted by $G(r)$). Then, a Newton-Raphson algorithm is employed to obtain r such that the posterior density at r equals the density at $G(r)$. The following function in `R` performs this algorithm.

```r
HPD.Beta <- function(a, b, alpha = 0.95) {
    # function to find HPD credible set with shape parameters a and b

    credint <- function(right_endpoint) {
        left_endpoint <- qbeta(pbeta(right_endpoint, a, b) - alpha, a, b)
        diff <- dbeta(left_endpoint, a, b) - dbeta(right_endpoint, a, b)
        return(diff^2)
    }

    root <- optimise(credint,
                     interval = c(qbeta(alpha + 1e-5, a, b), 1))$minimum
    int <- c(qbeta(pbeta(root, a, b) - alpha, a, b), root)  # the credible interval
    return(int)
}

```

Finally, to find the frequentist coverage of θ by 95% credible interval and the usual 95% frequentist confidence interval, we perform $B = 1000$ simulation. In each simulation, $n = 20$ datapoints are generated from fixed θ, then the confidence interval CI_b and credible set CR_b are created for $b = 1, 2, \ldots B$. The coverage probabilities are then determined as

$$\text{Coverage probability of Bayesian credible set} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\{\theta \in CR_b\}$$

$$\text{Coverage probability of usual frequentist confidence interval} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{1}\{\theta \in CI_b\}$$

```r
simQ2 <- function(theta) {
    set.seed(1911)  # set a seed for reproducibility
    B <- 1000   # perform 1000 simulations
    n <- 20   # sample size is 20
    pb <- txtProgressBar(min = 0, max = B, style = 3)  # set a progress bar

    bayesCoverage <- logical(B)
    freqCoverage <- logical(B)   # holds the indicator for coverage of the true theta by 95% CI

    for (b in 1:B) {
        data <- rbinom(n, size = 1, prob = theta)   # simulate the data
        Xsum <- sum(data)

        credible_int <- HPD.Beta(Xsum + 0.5, n - Xsum + 0.5)
        if (credible_int[1] <= theta & credible_int[2] >= theta) {
            bayesCoverage[b] <- TRUE
        }

        theta_hat <- Xsum / n

        if ((theta_hat + qnorm(0.025) * sqrt(theta_hat * (1 - theta_hat)/n) <= theta )
            & (theta_hat - qnorm(0.025) * sqrt(theta_hat * (1 - theta_hat)/n) >= theta )) {
            freqCoverage[b] <- TRUE
        }

        setTxtProgressBar(pb, value = b)
    }
    close(pb)

    cat("Frequentist CI Coverage", sum(freqCoverage)/B, "\n")
    cat("Bayesian CR Coverage", sum(bayesCoverage)/B, "\n")
}

simQ2(theta = 1/8)  # pass theta = 1/8, 1/4, 1/2, 3/4, 7/8

```

θ	**CP (Bayesian CR)**	**CP (Frequentist CI)**
1/8	0.962	0.92
1/4	0.968	0.897
1/2	0.957	0.957
3/4	0.966	0.897
7/8	0.962	0.92

Table 1: Frequentist coverage probabilities of Bayesian HPD credible set and the usual frequentist confidence interval for different values of θ

The obtained coverage probability are summarized in the following table. It can be clearly seen that Bayesian credible region has a consistent coverage probability at least as large as 95%, while the frequentist CI does not ensure enough coverage probability if the true θ is close to 0 or 1.

Problem 3. Let $X_1, \cdots, X_n$ be i.i.d. $\sim f(x|\theta),\ \theta \in \mathbb{R}$ and $\pi(\theta)$ be a prior density of θ. Use a result proved in the class (see proof of asymptotic normality of posterior distribution) to rigorously prove that

$$\log \int_{\mathbb{R}} \prod_{1}^{n} f(X_i|\theta)\pi(\theta)d\theta = \sum_{i=1}^{n} \log f(X_i|\hat{\theta}_n) - \frac{\log n}{2} + \frac{\log 2\pi}{2} - \frac{\log I(\theta_0)}{2} + \log \pi(\theta_0) + o_p(1)$$

as $n \to \infty$, where $\hat{\theta}_n$ is the MLE and $I(\theta_0)$ is Fisher information number at θ_0.

Solution. In the proof of asymptotic normality of posterior distribution, it was shown that with P_{θ_0}-probability one,

$$\int_{\mathbb{R}} \left| \pi\left(\widehat{\theta}_n + \frac{t}{\sqrt{n}}\right) \exp\left\{ L_n\left(\widehat{\theta}_n + \frac{t}{\sqrt{n}}\right) - L_n(\widehat{\theta}_n)\right\} - \pi(\theta_0)e^{-I(\theta_0)t^2/2} \right| dt \to 0$$

Clearly, it then follows that with P_{θ_0}-probability one,

$$C_n = \int_{\mathbb{R}} \pi\left(\widehat{\theta}_n + \frac{t}{\sqrt{n}}\right) \exp\left\{ L_n\left(\widehat{\theta}_n + \frac{t}{\sqrt{n}}\right) - L_n(\widehat{\theta}_n)\right\} dt \to \pi(\theta_0)\frac{\sqrt{2\pi}}{\sqrt{I(\theta_0)}} \tag{2}$$

Now note that,

$$C_n = \int_{\mathbb{R}} \pi\left(\widehat{\theta}_n + \frac{t}{\sqrt{n}}\right) \exp\left\{ L_n\left(\widehat{\theta}_n + \frac{t}{\sqrt{n}}\right) - L_n(\widehat{\theta}_n)\right\} dt$$

Substituting $\theta = \widehat{\theta}_n + (t/\sqrt{n})$, we obtain,

$$\begin{aligned} &= \int_{\mathbb{R}} \sqrt{n}\pi(\theta) \exp\left\{ L_n(\theta) - L_n(\widehat{\theta}_n)\right\} d\theta \\ &= \int_{\mathbb{R}} \sqrt{n}\pi(\theta) \exp\left\{ \log\left(\prod_{i=1}^{n} f(X_i \mid \theta)\right) - \log\left(\prod_{i=1}^{n} f(X_i \mid \widehat{\theta}_n)\right)\right\} d\theta \\ &= \frac{\sqrt{n}}{\prod_{i=1}^{n} f(X_i \mid \widehat{\theta}_n)} \int_{\mathbb{R}} \prod_{i=1}^{n} f(X_i \mid \theta)\pi(\theta)d\theta \end{aligned}$$

Combining the simplified form of C_n and Eq. (2), we obtain that with P_{θ_0}-probability one,

$$\frac{C_n}{\pi(\theta_0)\frac{\sqrt{2\pi}}{\sqrt{I(\theta_0)}}} = \frac{\sqrt{n}\sqrt{I(\theta_0)}}{\pi(\theta_0)\sqrt{2\pi}\prod_{i=1}^{n} f(X_i \mid \widehat{\theta}_n)} \int_{\mathbb{R}} \prod_{i=1}^{n} f(X_i \mid \theta)\pi(\theta)d\theta \to 1$$

i.e. with P_{θ_0}-probability one,

$$\begin{aligned} \log\left(\frac{C_n}{\pi(\theta_0)\frac{\sqrt{2\pi}}{\sqrt{I(\theta_0)}}}\right) &= \log \int_{\mathbb{R}} \prod_{i=1}^{n} f(X_i \mid \theta)\pi(\theta)d\theta - \sum_{i=1}^{n} \log f(X_i \mid \widehat{\theta}_n) \\ &\quad + \frac{1}{2}\log(n) - \frac{1}{2}\log(2\pi) + \frac{1}{2}\log(I(\theta_0)) - \log(\pi(\theta_0)) \to 0 \end{aligned} \tag{3}$$

However, the left hand side of Eq. (3) is a random variable depending on $X_1, X_2, \ldots X_n$, hence Eq. (3) simply means that the left hand side converges almost surely to 0. Since, almost sure convergence is stronger than convergence in probability, we have

$$\log \int_{\mathbb{R}} \prod_{i=1}^{n} f(X_i \mid \theta)\pi(\theta)d\theta - \sum_{i=1}^{n} \log f(X_i \mid \widehat{\theta}_n) + \frac{\log(n)}{2} - \frac{\log(2\pi)}{2} + \frac{\log(I(\theta_0))}{2} - \log(\pi(\theta_0)) \xrightarrow{P} 0$$

i.e.,

$$\log \int_{\mathbb{R}} \prod_{i=1}^{n} f(X_i \mid \theta)\pi(\theta)d\theta = \sum_{i=1}^{n} \log f(X_i \mid \widehat{\theta}_n) - \frac{\log n}{2} + \frac{\log 2\pi}{2} - \frac{\log I(\theta_0)}{2} + \log \pi(\theta_0) + o_p(1)$$

Problem 4. Show that the result on asymptotic normality of the posterior distribution of $\sqrt{n}(\theta - \hat{\theta}_n)$, proved in the class, implies consistency of the posterior distribution of θ at θ_0.

Solution. By the remark mentioned before Theorem 4.2 of the book[1], it follows that there exists a sequence of solutions $\widehat{\theta}_n$ of the likelihood equation, which strongly consistent for the true parameter θ_0. In other words, this means, with P_{θ_0} probability one, $\widehat{\theta}_n \to \theta_0$ as $n \to \infty$.
On the other hand, the result on asymptotic normality of the posterior distribution (i.e. Theorem 4.2), given the assumptions (A1)-(A4) holds, with P_{θ_0}-probability one,

$$\int_{\mathbb{R}} \left| \pi_n^*(t \mid X_1, \dots X_n) - \frac{\sqrt{I(\theta_0)}}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2 I(\theta_0)} \right| dt \to 0$$

as $n \to \infty$, where $t = \sqrt{n}(\theta - \widehat{\theta}_n)$, and $\pi_n^*(t \mid X_1, \dots X_n)$ denotes the posterior distribution of t conditional on the sample $X_1, \dots X_n$ of size n.
Let us assume $(\Omega, \mathcal{A}, \mathbb{P})$ be the probability measure space on which each of the random variables $X_1, X_2, \dots$ are defined and are measurable. Let, A and B be two subsets of Ω given as follows

$$\begin{aligned} A &= \left\{ \omega \in \Omega : \lim_{n \to \infty} \widehat{\theta}_n(\omega) = \theta_0 \right\} \\ B &= \left\{ \omega \in \Omega : \lim_{n \to \infty} \int_{\mathbb{R}} \left| \pi_n^*(t \mid X_1(\omega), \dots X_n(\omega)) - \frac{\sqrt{I(\theta_0)}}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2 I(\theta_0)} \right| dt = 0 \right\} \end{aligned}$$

where $\widehat{\theta}_n(\omega)$ is a strongly consistent solution to the likelihood equation

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta} \left(\log f(X_i(\omega) \mid \theta) \right) = 0$$

By the arguments on the existence of stongly consistent solution and the result on asymptotic normality of posterior distribution, it follows that $\mathbb{P}_{\theta_0}(A) = \mathbb{P}_{\theta_0}(B) = 1$. Clearly, $\mathbb{P}_{\theta_0}(A \cap B) = 1$. Now fix any $\epsilon > 0$, and choose some $\omega \in (A \cap B)$. It then follows that,

$$\begin{aligned} & \mathbb{P}(|\theta - \theta_0| < \epsilon \mid X_1(\omega), \dots X_n(\omega)) \\ \geq \; & \mathbb{P}\left(\left\{ |\theta - \widehat{\theta}_n(\omega)| < \epsilon/2 \right\} \cap \left\{ |\widehat{\theta}_n(\omega) - \theta_0| < \epsilon/2 \right\} \mid X_1(\omega), \dots X_n(\omega) \right) \\ & \qquad \text{since, by triangle inequality, } |\theta - \theta_0| \leq |\theta - \widehat{\theta}_n(\omega)| + |\widehat{\theta}_n(\omega) - \theta_0| \end{aligned}$$

Since, $\omega \in A$, there exists N_1 such that for any $n \geq N_1$, $|\widehat{\theta}_n(\omega) - \theta_0| < \epsilon/2$, for the chosen ω and ϵ. Thus, for all $n \geq N_1$, we have,

[1]An Introduction to Bayesian Analysis: Theory and Methods by J. K. Ghosh, M. Delampady and T. Samanta

$$
\begin{aligned}
& \mathbb{P}\left(\left\{|\theta-\widehat{\theta}_n(\omega)|<\epsilon/2\right\}\cap\left\{|\widehat{\theta}_n(\omega)-\theta_0|<\epsilon/2\right\}\mid X_1(\omega),\ldots X_n(\omega)\right)\\
= & \mathbb{P}\left(|\theta-\widehat{\theta}_n(\omega)|<\epsilon/2\mid X_1(\omega),\ldots X_n(\omega)\right)\\
= & \mathbb{P}\left(|\sqrt{n}(\theta-\widehat{\theta}_n(\omega))|<\sqrt{n}\epsilon/2\mid X_1(\omega),\ldots X_n(\omega)\right)\\
= & \int_{-\sqrt{n}\epsilon/2}^{\sqrt{n}\epsilon/2}\pi_n^*(t\mid X_1(\omega),\ldots X_n(\omega))dt
\end{aligned}
$$

Now, since $\omega\in B$, and since

$$
\begin{aligned}
& \int_{\mathbb{R}}\left|\pi_n^*(t\mid X_1(\omega),\ldots X_n(\omega))-\frac{\sqrt{I(\theta_0)}}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2I(\theta_0)}\right|dt\\
\geq & \int_{-\sqrt{n}\epsilon/2}^{\sqrt{n}\epsilon/2}\left|\pi_n^*(t\mid X_1(\omega),\ldots X_n(\omega))-\frac{\sqrt{I(\theta_0)}}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2I(\theta_0)}\right|dt\\
\geq & \left|\int_{-\sqrt{n}\epsilon/2}^{\sqrt{n}\epsilon/2}\pi_n^*(t\mid X_1(\omega),\ldots X_n(\omega))dt-\int_{-\sqrt{n}\epsilon/2}^{\sqrt{n}\epsilon/2}\frac{\sqrt{I(\theta_0)}}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2I(\theta_0)}dt\right|\\
= & \left|\int_{-\sqrt{n}\epsilon/2}^{\sqrt{n}\epsilon/2}\pi_n^*(t\mid X_1(\omega),\ldots X_n(\omega))dt-\left(\Phi\left(\frac{\sqrt{n}\epsilon}{2}I(\theta_0)\right)-\Phi\left(-\frac{\sqrt{n}\epsilon}{2}I(\theta_0)\right)\right)\right|
\end{aligned}
$$

where $\Phi(\cdot)$ is the cumulative distribution function of standard normal distribution. Since the dominating integral goes to zero as $n\to\infty$, therefore, for any chosen $\delta>0$, there would exist N_δ such that for any $n\geq N_\delta$,

$$
\int_{-\sqrt{n}\epsilon/2}^{\sqrt{n}\epsilon/2}\pi_n^*(t\mid X_1(\omega),\ldots X_n(\omega))dt\geq-\delta+\Phi\left(\frac{\sqrt{n}\epsilon}{2}I(\theta_0)\right)-\Phi\left(-\frac{\sqrt{n}\epsilon}{2}I(\theta_0)\right)
$$

Therefore, we have for any arbitrary $\delta>0$, and for $n\geq\max\{N_1,N_\delta\}$ and for any $\omega\in(A\cap B)$,

$$
\begin{aligned}
\mathbb{P}(|\theta-\theta_0|<\epsilon\mid X_1(\omega),\ldots X_n(\omega)) &\geq -\delta+\Phi\left(\frac{\sqrt{n}\epsilon}{2}I(\theta_0)\right)-\Phi\left(-\frac{\sqrt{n}\epsilon}{2}I(\theta_0)\right)\\
&\to(1-\delta),\ \text{as } n\to\infty
\end{aligned}
$$

Since $\delta>0$ is arbitrary, we have

$$
\mathbb{P}(|\theta-\theta_0|<\epsilon\mid X_1(\omega),\ldots X_n(\omega))\to 1
$$

as $n\to\infty$, for all $\omega\in(A\cap B)$, which has P_{θ_0}-probability one. Now since $\epsilon>0$ is arbitrary, it follows that the posterior distribution is consistent at θ_0.

Problem 5. Show that Condition (A4), used to prove asymptotic normality of posterior distribution, holds when $X_1, \cdots, X_n$ are i.i.d. $\mathcal{N}(\theta, 1)$.

Solution. We need to show that for any $\delta > 0$, with P_{θ_0}-probability one, the condition

$$\sup_{|\theta-\theta_0|>\delta} \frac{1}{n}(L_n(\theta) - L_n(\theta_0)) < (-\epsilon) \tag{4}$$

for some $\epsilon > 0$, and for sufficiently large n. Here, $L_n(\theta)$ denotes the log-likelihood of θ based on n datapoints $X_1, \dots X_n$, and the true parameter is indicated by θ_0.
For the given question, we note that the log-likelihood can be expressed as

$$\begin{aligned} L_n(\theta) &= \log\left(\prod_{i=1}^{n} f(X_i \mid \theta)\right) \\ &= \log\left(\frac{1}{(2\pi)^{(n/2)}} e^{-\sum_i (X_i-\theta)^2/2}\right) \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2 \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{1}{n}(L_n(\theta) - L_n(\theta_0)) &= \frac{1}{2n}\left[\sum_{i=1}^{n}(X_i-\theta_0)^2 - \sum_{i=1}^{n}(X_i-\theta)^2\right] \\ &= \frac{1}{2n}\left[2n\overline{X}_n(\theta-\theta_0) + n(\theta_0^2-\theta^2)\right], \qquad \text{where } \overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \\ &= (\theta-\theta_0)\left[\overline{X}_n - \frac{(\theta+\theta_0)}{2}\right] \end{aligned}$$

Now, under true parameter θ_0, and using Strong Law of Large Numbers, it follows that $\overline{X}_n \to \theta_0$ almost surely (i.e. with P_{θ_0}-probability one.) In other words,

$$\frac{1}{n}(L_n(\theta) - L_n(\theta_0)) \xrightarrow{a.s.} -\frac{(\theta-\theta_0)^2}{2}$$

For $|\theta-\theta_0| > \delta$, we have $-\frac{(\theta-\theta_0)^2}{2} < -\frac{\delta^2}{2}$. Therefore, with P_{θ_0}-probability one,

$$\frac{1}{n}(L_n(\theta) - L_n(\theta_0)) < -\frac{\delta^2}{4} \ \forall\theta \text{ such that } |\theta-\theta_0| > \delta$$

i.e.

$$\sup_{|\theta-\theta_0|>\delta} \frac{1}{n}(L_n(\theta) - L_n(\theta_0)) < -\frac{\delta^2}{4}$$

for sufficiently large n. Choosing $\epsilon = \delta^2/4$ now yields Eq. (4).

Problem 6. Let $X_1, \cdots, X_n$ be i.i.d. with a Cauchy density

$$f(x|\theta) = \frac{1}{\pi[1+(x-\theta)^2]}, \quad -\infty < x < \infty,\ \theta > 0$$

We want to find the $100(1-\alpha)\%$ credible set for θ.
Draw n (your choice) observations from this distribution with a chosen θ. Based on these observations, find 95% and 99% HPD credible sets for θ. Do this for three (or more) different values of n (small, moderately large, large/very large). Also describe your algorithm for finding the HPD credible sets.

Solution. In this case, we assume an objective uniform prior for θ, i.e. $\pi(\theta) = 1$ for all $\theta > 0$. Therefore, the posterior distribution,

$$\pi(\theta \mid X_1, \dots X_n) \propto \prod_{i=1}^{n} \frac{1}{[1+(X_i-\theta)^2]}, \quad \theta > 0$$

Since the integral $\int_0^\infty \prod_{i=1}^n \frac{1}{[1+(X_i-\theta)^2]} d\theta$ is not analytically tractable, hence the posterior density cannot be obtained in closed form. Hence, one must resort to usage of MCMC samples obtained from that posterior distribution. To compute the HPD credible interval in such case, we shall use **Chen-Shao HPD Estimation Algorithm**[2]. The algorithm works on the principle that the HPD credible interval is the interval with smallest width subject to the constraint that the posterior probability of θ lying inside the credible set is $100\alpha\%$. The steps are as follows:

1. Obtain an MCMC sample from the posterior, say $\{\theta_i : i = 1, 2, \dots B\}$.
2. Sort these samples to obtain the order statistics $\theta_{(1)} \le \theta_{(2)} \cdots \le \theta_{(B)}$.
3. Compute the $100\alpha\%$ credible intervals

$$R_j(B) = (\theta_{(j)}, \theta_{(j+[\alpha B])})$$

for $j = 1, 2, \dots B - [\alpha B]$.

4. The $100\alpha\%$ HPD credible interval is the one $R_{j^*}(B)$ with the smallest interval width.

The following function implements this Chen Shao Algorithm as `HPD.CS` function in `R`, given the input of MCMC samples from the posterior distribution.

```r
HPD.CS <- function(mcmc_samples, alpha = 0.95) {
    order_theta <- sort(mcmc_samples)    # perform step 2
    B <- length(order_theta)
    b <- floor(alpha * B)
    lengths <- numeric(B - b)   # array to store the lengths of the intervals

```

[2]See Section 7.3.1 in Monte Carlo Methods in Bayesian Computation authored by Ming-Hui Chen, Qi-Man Shao, Joseph G. Ibrahim.

```
    for (j in 1:(B-b)) {
        # compute the lengths (step 3)
        lengths[j] <- order_theta[j+b] - order_theta[j]
    }

    jstar <- which.min(lengths)  # find j* (step 4)

    # return the HPD credible interval
    return(c(order_theta[jstar], order_theta[jstar + b]))
}

```

In order to obtain the posterior samples using MCMC, we use Metropolis Hastings algorithm. In this case, we use the normal distribution with unit variance as a candidate distribution. The steps of this algorithm is as follows:

1. Pick an initial state θ_0. Set $t = 0$.
2. Iterate for $t = 0, 1, 2, \ldots$:
   (a) Generate θ' from the distribution $\mathcal{N}(\theta_t, 1)$.
   (b) Calculate the acceptance ratio $\alpha = \min\{f(\theta')/f(\theta_t), 1\}$, where $f(\theta)$ is the scaled version of posterior density i.e. $\prod_{i=1}^{n}[1 + (X_i - \theta)^2]^{-1}$.
   (c) Make
   $$\theta_{(t+1)} = \begin{cases} \theta' & \text{with probability } \alpha \\ \theta_t & \text{with probability } (1-\alpha) \end{cases}$$

Since the samples are usually correlated one in every few (say 20) samples is taken as a sample from the posterior distribution, while the rest are discarded. Also, to ensure convergence, the first few hundreds of iterations are discarded as a burn-in period (say first 1000 samples). Note that, in the given problem,

$$\alpha = \min\left\{1, \prod_{i=1}^{n} \frac{[1 + (X_i - \theta_t)^2]}{[1 + (X_i - \theta')^2]}\right\} = \exp\left[\min\left\{0, \sum_{i=1}^{n} \left(\log[1 + (X_i - \theta_t)^2] - \log[1 + (X_i - \theta')^2]\right)\right\}\right]$$

This representation of the probability is useful to avoid numerical underflow, which might occur due to multiplication of many small numbers. The following function in `R` implements this idea.

```
MH.samples <- function(X_samples, burnin = 1000, B = 1000, thin = 20) {
    mcmc_samples <- numeric(B)  # array to hold mcmc_samples
    nchain <- burnin + (B * thin) - 1

    # initialize theta_0
    theta <- mean(X_samples)
    pb <- txtProgressBar(min = 0, max = nchain, style = 3)  # set a progress bar

    for (i in 1:nchain) {
        accept <- FALSE
        while (!accept) {
```

```
            theta_prime <- rnorm(1, mean = theta, sd = 1)  # get theta'
            post_ratio <- sum(log(1 + (X_samples - theta)^2 )
                              - log(1 + (X_samples - theta_prime)^2 ))
            A <- min(1, exp(post_ratio))
            check <- rbinom(1, size = 1, prob = A)

            if (check == 1) {
                # if accepted, update theta_t
                theta <- theta_prime
                accept <- TRUE
            }
        }

        # if we have thin-th sample after burnin, make it iid posterior sample
        if ((i >= burnin) & ((i-burnin) %% thin == 0)) {
            index <- (i - burnin) %/% thin + 1
            mcmc_samples[index] <- theta
        }

        setTxtProgressBar(pb, value = i)  # update progress bar
    }

    close(pb)
    return(mcmc_samples)
}

```

I choose three different samples sizes $n = 5$ (small), $n = 25$ (moderately large) and $n = 100$ (large), and choose $\theta = 10$ for the simulation study.
The results are summarized in the following table. As seen from the table, all of the credible sets actually contains the true value of the parameter $\theta = 10$. Also, as the sample size n increases, the length of the HPD credible intervals decreases.

Sample size (n)	**Confidence Coefficient**	**Lower Bound**	**Upper Bound**
5	0.95	9.215795	12.885111
5	0.99	8.197846	13.262514
25	0.95	9.768901	10.895991
25	0.99	9.641768	11.046742
100	0.95	9.949556	10.593906
100	0.99	9.858396	10.684176

Table 2: 95% and 99% credible intervals for $\theta = 10$, for small ($n = 5$), moderately large ($n = 25$) and large ($n = 100$) samples from Cauchy distribution with location parameter θ, under objective uniform prior for θ.

Problem 7. Let $X_1, \cdots, X_n$ be i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$, μ, σ^2 both unknown. Consider the set up of Jeffreys test (see class note).

(a) Show that if $\overline{X} \to \infty$ and s^2 is bounded, BF_{01} goes to zero for the Cauchy prior but does not go to zero for normal prior.

(b) Consider Cauchy prior for this problem. Using the representation of the Cauchy density $g_1(\mu|\sigma)$ as a scale mixture of normals, express the integrated likelihood under H_1 as a one-dimensional integral (over the mixing variable τ).

Solution. The setup of Jeffreys test is as follows: Let $X_1, \ldots X_n$ be i.i.d. $\sim \mathcal{N}(\mu, \sigma^2)$, where μ, σ^2 are both unknown. The hypotheses are $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$, where μ_0 is a pre-specified known constant. Without loss of generality, one can assume $\mu_0 = 0$.
Under H_0, one may take the usual objective prior $g_0(\sigma) = (1/\sigma)$ for $\sigma > 0$. Under H_1, one may take the objective prior as $g_1(\mu, \sigma) = \dfrac{1}{\sigma} g_1(\mu \mid \sigma) = \dfrac{1}{\sigma^2} g_2(\mu/\sigma)$, where $g_2(\cdot)$ is a suitably chosen prior for the conditional mean.

(a) If one chooses $g_2(\cdot)$ to be the standard normal density $\mathcal{N}(0, 1)$, then we have the prior under H_1 as

$$g_1(\mu, \sigma^2) = \frac{1}{\sigma^2} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{\mu^2}{2\sigma^2}\right\}$$

To calculate the Bayes factor, we start by calculating its numerator first.

$$\begin{aligned}
& \int_0^\infty f(X_1, \ldots X_n \mid 0, \sigma^2) g_0(\sigma) d\sigma \\
= \quad & \int_0^\infty \frac{1}{(2\pi)^{n/2} \sigma^{(n+1)}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n X_i^2\right\} d\sigma \\
= \quad & \frac{1}{(2\pi)^{n/2}} \int_0^\infty (\sigma^2)^{-\frac{(n+2)}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n X_i^2\right\} \frac{1}{2} d(\sigma^2) \\
= \quad & \frac{1}{2(2\pi)^{n/2}} \Gamma\left(\frac{n}{2}\right) \left[\frac{\sum_{i=1}^n X_i^2}{2}\right]^{-(n/2)}
\end{aligned}$$

Next, we consider the normal prior under H_1, and compute the denominator of the Bayes factor,

$$
\begin{aligned}
& \int_0^\infty \int_{-\infty}^\infty f(X_1, \ldots X_n \mid \mu, \sigma^2) g_1(\mu, \sigma) d\mu d\sigma \\
= \quad & \int_0^\infty \int_{-\infty}^\infty \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu)^2\right\} \times \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left\{-\frac{\mu^2}{2\sigma^2}\right\} d\mu d\sigma \\
= \quad & \int_0^\infty \int_{-\infty}^\infty \frac{\sigma^{-(n+2)}}{(2\pi)^{(n+1)/2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n (X_i - \mu)^2 + \mu^2\right]\right\} d\mu d\sigma \\
= \quad & \int_0^\infty \frac{\sigma^{-(n+2)}}{(2\pi)^{(n+1)/2}} \exp\left\{-\frac{\sum_{i=1}^n X_i^2}{2\sigma^2}\right\} \int_{-\infty}^\infty \exp\left[-\frac{1}{2}\left(\mu^2 \frac{(n+1)}{\sigma^2} - 2\frac{n\overline{X}}{\sigma^2}\mu\right)\right] d\mu d\sigma \\
= \quad & \int_0^\infty \frac{\sigma^{-(n+2)}}{(2\pi)^{(n+1)/2}} \exp\left\{-\frac{\sum_{i=1}^n X_i^2}{2\sigma^2}\right\} \frac{\sqrt{2\pi}\sigma}{\sqrt{(n+1)}} \exp\left\{\frac{1}{2}\left(\frac{n\overline{X}}{\sigma^2}\right)^2 \frac{\sigma^2}{(n+1)}\right\} d\sigma \\
= \quad & \int_0^\infty \frac{\sigma^{-(n+1)}}{(2\pi)^{n/2}\sqrt{(n+1)}} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n X_i^2 - \frac{n^2}{(n+1)}\overline{X}^2\right]\right\} d\sigma \\
= \quad & \frac{1}{2(2\pi)^{n/2}\sqrt{(n+1)}} \Gamma\left(\frac{n}{2}\right)\left[\frac{\sum_{i=1}^n X_i^2 - \overline{X}^2 n^2/(n+1)}{2}\right]^{n/2}
\end{aligned}
$$

Therefore, the Bayes factor is,

$$
\begin{aligned}
BF_{01} &= \sqrt{(n+1)}\left[\frac{\sum_{i=1}^n X_i^2 - \overline{X}^2 n^2/(n+1)}{\sum_{i=1}^n X_i^2}\right]^{(n/2)} \\
&= \sqrt{(n+1)}\left[\frac{(n-1)s^2 + \overline{X}^2 \frac{n}{(n+1)}}{(n-1)s^2 + n\overline{X}^2}\right]^{(n/2)} \\
&\to \sqrt{(n+1)}\frac{1}{(n+1)^{(n/2)}}, \text{ as } \overline{X} \to \infty \text{ and } s^2 \text{ remains bounded} \\
&= (n+1)^{-(n-1)/2} > 0
\end{aligned}
$$

Therefore, the Bayes factor does not go to zero for normal prior.

To show that the Bayes factor go to zero for Cauchy prior, we shall first express the integrated likelihood under H_1 as a one-dimensional integral (over the mixing variable τ), as to be shown in part (b).

(b) In this case, the Cauchy prior under H_1 is given by

$$
g_1(\mu, \sigma) = \frac{1}{\sigma}\frac{1}{\sigma\pi(1 + \mu^2/\sigma^2)}
$$

An alternative representation of above prior can be obtained through a mixing variable λ such that $\mu \mid (\sigma, \lambda) \sim \mathcal{N}(0, \sigma^2/\lambda)$ and $\lambda \sim \text{Gamma}(1/2, 1/2)$. Therefore, the Cauchy prior under H_1 can also be expressed as

$$g_1(\mu, \sigma) = \frac{1}{\sigma} \int_0^\infty \frac{\sqrt{\lambda}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\lambda}{2\sigma^2}\mu^2\right) \frac{1}{\sqrt{2\pi}} \lambda^{(1/2)-1} e^{-\lambda/2} d\lambda$$

Therefore, the integrated likelihood under H_1 is

$$\begin{aligned}
& \int_0^\infty \int_{-\infty}^\infty f(X_1, \ldots X_n \mid \mu, \sigma^2) g_1(\mu, \sigma) d\mu d\sigma \\
= & \int_0^\infty \int_{-\infty}^\infty \frac{1}{(2\pi)^{n/2}\sigma^n} e^{-\dfrac{\sum_{i=1}^n (X_i - \mu)^2}{2\sigma^2}} \times \int_0^\infty \frac{1}{(2\pi)\sigma^2} \exp\left(-\frac{\lambda}{2\sigma^2}\mu^2\right) e^{-\lambda/2} d\lambda d\mu d\sigma \\
= & \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \frac{\sigma^{-(n+2)}}{(2\pi)^{(n+2)/2}} e^{-\lambda/2} e^{-\frac{\sum_{i=1}^n X_i^2}{2\sigma^2}} \exp\left[-\frac{1}{2}\left(\mu^2 \frac{(n+\lambda)}{\sigma^2} - 2\frac{n\overline{X}}{\sigma^2}\mu\right)\right] d\mu d\sigma d\lambda \\
& \qquad \text{where, the exchange of integrals is justified by Fubini's theorem} \\
= & \int_0^\infty \int_0^\infty \frac{\sigma^{-(n+2)}}{(2\pi)^{(n+2)/2}} e^{-\lambda/2} e^{-\frac{\sum_{i=1}^n X_i^2}{2\sigma^2}} \frac{\sqrt{2\pi}\sigma}{\sqrt{(n+\lambda)}} \exp\left\{\frac{1}{2}\left(\frac{n\overline{X}}{\sigma^2}\right)^2 \frac{\sigma^2}{(n+\lambda)}\right\} d\sigma d\lambda \\
= & \int_0^\infty \frac{e^{-\lambda/2}}{(2\pi)^{(n+1)/2}\sqrt{(n+\lambda)}} \int_0^\infty \sigma^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^n X_i^2 - \frac{n^2}{(n+\lambda)}\overline{X}\right]\right\} d\sigma d\lambda \\
= & \int_0^\infty \frac{e^{-\lambda/2}}{2(2\pi)^{(n+1)/2}\sqrt{(n+\lambda)}} \Gamma\left(\frac{n}{2}\right) \left[\frac{\sum_{i=1}^n X_i^2 - n^2\overline{X}^2/(n+\lambda)}{2}\right]^{-n/2} d\lambda
\end{aligned}$$

Therefore, Bayes factor for Cauchy prior is

$$\begin{aligned}
BF_{01} &= \left[\int_0^\infty \frac{1}{\sqrt{2\pi}\sqrt{(n+\lambda)}} e^{-\lambda/2} \left[\frac{\sum_{i=1}^n X_i^2}{\sum_{i=1}^n X_i^2 - n^2\overline{X}^2/(n+\lambda)}\right]^{n/2} d\lambda\right]^{-1} \\
&= \left[\int_0^\infty \frac{1}{\sqrt{2\pi}\sqrt{(n+\lambda)}} e^{-\lambda/2} \left[\frac{(n-1)s^2 + n\overline{X}^2}{(n-1)s^2 + \lambda n\overline{X}^2/(n+\lambda)}\right]^{n/2} d\lambda\right]^{-1}
\end{aligned}$$

Now, an application of Fatou's lemma yields,

$$
\begin{aligned}
\liminf_{\overline{X}\to\infty} BF_{01}^{-1} &= \liminf_{\overline{X}\to\infty} \int_0^\infty \frac{1}{\sqrt{2\pi}\sqrt{(n+\lambda)}} e^{-\lambda/2} \left[\frac{(n-1)s^2 + n\overline{X}^2}{(n-1)s^2 + \lambda n \overline{X}^2/(n+\lambda)}\right]^{n/2} d\lambda \\
&\geq \int_0^\infty \liminf_{\overline{X}\to\infty} \left[\frac{1}{\sqrt{2\pi}\sqrt{(n+\lambda)}} e^{-\lambda/2} \left[\frac{(n-1)s^2 + n\overline{X}^2}{(n-1)s^2 + \lambda n \overline{X}^2/(n+\lambda)}\right]^{n/2}\right] d\lambda \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}\sqrt{(n+\lambda)}} e^{-\lambda/2} \left[\frac{(n+\lambda)}{\lambda}\right]^{n/2} d\lambda \\
&= \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{e^{-\lambda/2}}{\sqrt{\lambda}} \left(1 + \frac{n}{\lambda}\right)^{(n-1)/2} d\lambda \\
&\geq \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{e^{-\lambda/2}}{\sqrt{\lambda}} \left(1 + \frac{n(n-1)}{2\lambda}\right) d\lambda, \quad \text{by Bernoulli's inequality for } n \geq 3 \\
&\geq \frac{n(n-1)}{2\sqrt{2\pi}} \int_0^\infty \lambda^{-3/2} e^{-\lambda/2} d\lambda, \quad \text{which diverges}
\end{aligned}
$$

If $n = 2$, then also

$$
\frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{e^{-\lambda/2}}{\sqrt{\lambda}} \left(1 + \frac{2}{\lambda}\right)^{1/2} d\lambda \geq \frac{1}{\sqrt{\pi}} \int_0^\infty \lambda^{-1} e^{-\lambda/2} d\lambda
$$

which again diverges to ∞. Therefore, by comparison test, it follows that $\liminf_{\overline{X}\to\infty} BF_{01}^{-1} = \infty$, in other words, $\limsup_{\overline{X}\to\infty} BF_{01} = 0$. But since $BF_{01} \geq 0$, therefore, it follows that $\lim_{\overline{X}\to\infty} BF_{01} = 0$.

Problem 8. Welch's paradox.

(a) Let X_1, X_2 be i.i.d. $\sim U\left(\theta - \dfrac{1}{2}, \theta + \dfrac{1}{2}\right)$, $\theta \in \mathbb{R}$. A frequentist 95% confidence interval is
$(\overline{X} - 0.3882, \overline{X} + 0.3882)$ where $\overline{X} = \frac{X_1+X_2}{2}$. Show that if X_1 and X_2 are sufficiently apart, say $X_1 - X_2 > d$ (find d) then θ must be in this confidence interval (but a frequentist reports the confidence level as only 95%).

Calculate $\mathbb{P}\left(\text{The interval } \overline{X} \mp 0.3882 \text{ covers } \theta | X_1 - X_2\right)$. Also find the posterior distribution of θ with the objective prior $\pi(\theta) \equiv 1$ and find an appropriate 95% credible interval for θ.

(b) Let X_1, X_2 be i.i.d. with a common density belonging to a location parameter family of densities with a location parameter θ. Assume without loss of generality that $\mathbb{E}_\theta(X_1) = \theta$. One can find a frequentist 95% confidence interval of the form $(\overline{X} - c, \overline{X} + c)$. Suppose now that $X_1 - X_2$ is known and one calculates $\mathbb{P}\left(\text{The interval } \overline{X} \mp c \text{ covers } \theta | X_1 - X_2\right)$. When can Welch's paradox occur in such a scenario?

Can Welch's paradox occur if X_1, X_2 are i.i.d. $\mathcal{N}(\theta, 1)$? (Explain.)

Solution. (a) Let, $X_1 - X_2 = d$. Without loss of generality, assume that $d \geq 0$, i.e. $X_1 \geq X_2$. Then, $\overline{X} = X_1 - (d/2) = X_2 + (d/2)$. Now, since both X_1, X_2 are i.i.d $U(\theta - 1/2, \theta + 1/2)$, and $X_1 - X_2 = d$, it follows that $(\theta - 1/2 + d) \leq X_1 \leq (\theta + 1/2)$ and $(\theta - 1/2) \leq X_2 \leq (\theta + 1/2 - d)$. Thus, $(\theta - 1/2 + d/2) \leq \overline{X} \leq (\theta + 1/2 - d/2)$, i.e.

$$|\overline{X} - \theta| \leq \frac{(1-d)}{2}$$

where d is the distance between two samples X_1 and X_2. However, if d is such that, $(1-d)/2 \leq 0.3882$, then clearly, θ must be in the confidence interval $(\overline{X} - 0.3882, \overline{X} + 0.3882)$. Therefore, $d \geq 1 - (2 \times 0.3882) = 0.2236$. Therefore, if X_1 and X_2 are atleast 0.2236 units apart, then the given frequentist 95% confidence interval must contain the true θ.

Turning to the calculation of the coverage probability, conditional on $X_1 - X_2 = d$ (assuming $d \geq 0$), we have

$$\begin{aligned}
& \mathbb{P}\left(\overline{X} - 0.3882 \leq \theta \leq \overline{X} + 0.3882 \mid X_1 - X_2 = d\right) \\
= \; & \mathbb{P}\left(X_2 + d/2 - 0.3882 \leq \theta \leq X_2 + d/2 + 0.3882 \mid X_1 - X_2 = d\right) \\
= \; & \mathbb{P}\left(\theta - 0.3882 - d/2 \leq X_2 \leq \theta + 0.3882 - d/2 \mid X_1 - X_2 = d\right)
\end{aligned}$$

Now, as indicated above, both X_1, X_2 are i.i.d $U(\theta - 1/2, \theta + 1/2)$, and $X_1 - X_2 = d$, it follows that $X_2 \sim U(\theta - 1/2, \theta + 1/2 - d)$. Therefore,

$$
\begin{aligned}
& \mathbb{P}\left(\overline{X} - 0.3882 \le \theta \le \overline{X} + 0.3882 \mid X_1 - X_2 = d\right) \\
= \; & \frac{\min\left\{\theta - \frac{d}{2} + 0.3882, \theta + \frac{1}{2} - d\right\} - \max\left\{\theta - \frac{d}{2} - 0.3882, \theta - \frac{1}{2}\right\}}{(\theta + \frac{1}{2} - d) - (\theta - \frac{1}{2})} \\
= \; & \frac{\min\left\{\theta - \frac{d}{2} + 0.3882, \theta + \frac{1}{2} - d\right\} - \max\left\{\theta - \frac{d}{2} - 0.3882, \theta - \frac{1}{2}\right\}}{(1 - d)} \\
= \; & \begin{cases} 1 & \text{if } d \ge 0.2236 \\ \dfrac{0.7764}{(1-d)} & \text{if } d < 0.2236 \end{cases}
\end{aligned}
$$

However, since we assumed that $d > 0$, by symmetry, it now follows that

$$
\mathbb{P}\left(\overline{X} - 0.3882 \le \theta \le \overline{X} + 0.3882 \mid X_1 - X_2 = d\right) = \begin{cases} 1 & \text{if } |d| \ge 0.2236 \\ \dfrac{0.7764}{(1-|d|)} & \text{if } 0 \le |d| < 0.2236 \end{cases}
$$

Now, to find the posterior distribution of θ with respect to the improper prior $\pi(\theta) = 1$, we note that

$$
\begin{aligned}
\pi(\theta \mid X_1, X_2) &\propto \pi(\theta) \mathbf{1}_{\{(\theta - 1/2) \le X_1 \le (\theta + 1/2)\}} \mathbf{1}_{\{(\theta - 1/2) \le X_2 \le (\theta + 1/2)\}} \\
&= \mathbf{1}_{\{\theta \ge X_{(2)} - 1/2\}} \mathbf{1}_{\{\theta \le X_(1) + 1/2\}}
\end{aligned}
$$

where $X_{(i)}$ is the i-th order statistic among X_1, X_2, for $i = 1, 2$. Hence, the posterior distribution of θ is uniform distribution $\left(X_{(2)} - 1/2, X_{(1)} + 1/2\right)$. Starting with the midpoint of this interval as the center, we consider the credible intervals of form $[\overline{X} - c, \overline{X} + c]$ for some choice of c. We wish to have,

$$
\begin{aligned}
& \mathbb{P}\left[\overline{X} - c \le \theta \le \overline{X} + c\right] = 0.95 \\
\Rightarrow \quad & \frac{2c}{(X_{(1)} - X_{(2)} + 1)} = 0.95 \\
\Rightarrow \quad & c = 0.475\left(1 + X_{(1)} - X_{(2)}\right)
\end{aligned}
$$

Therefore, the 95% credible interval for θ is given by

$$
\left[\overline{X} - 0.475\left(1 + X_{(1)} - X_{(2)}\right), \overline{X} + 0.475\left(1 + X_{(1)} - X_{(2)}\right)\right]
$$

(b) Let, X_1, X_2 be i.i.d. with a common density function $f(x, \theta)$ belonging to a location parameter family. Thus, $f(x, \theta) = g(x - \theta)$ for some suitable function g. In other words, the distribution of $Y_i = (X_i - \theta)$ is free of θ.

Now note that, by definition of the 95% confidence interval given by $(\overline{X} - c, \overline{X} + c)$,

$$\begin{aligned} 0.95 &= \mathbb{P}(\text{The interval } \overline{X} \mp c \text{ covers } \theta) \\ &= \mathbb{E}\left[\mathbb{P}(-c < \overline{X} - \theta < c \mid X_{(2)} - X_{(1)} = d)\right] \end{aligned}$$

where, $X_{(i)}$ denotes the i-th order statistic of the sample X_1, X_2. Welch's paradox occurs if $\mathbb{P}(\text{The interval } \overline{X} \mp c \text{ covers } \theta \mid X_{(2)} - X_{(1)} = d)$ is dependent on d. Because then, the coverage probability of the interval $(\overline{X} - c, \overline{X} + c)$ of containing the true parameter θ becomes a function of d, which at expectation over the choices of d equals the frequentist coverage probability 95%. Therefore, it is immediate that there will remain some choice of d for which the coverage probability will actually be higher than 95%.

Assume without loss of generality that $X_1 > X_2$, then,

$$\begin{aligned} \mathbb{P}(-c < \overline{X} - \theta < c \mid X_{(2)} - X_{(1)} = d) &= \mathbb{P}(-c < \overline{Y} < c \mid Y_1 - Y_2 = d) \\ &= \mathbb{P}(-c < Y_1 - (d/2) < c \mid Y_1 - Y_2 = d) \\ &= \mathbb{P}(-c + (d/2) < Y_1 < c + (d/2) \mid Y_1 - Y_2 = d) \end{aligned}$$

Since $(Y_1 - Y_2)$ is ancillary, the knowledge of $(Y_1 - Y_2)$ does not generally contain any knowledge of the sample Y_1, hence Welch's paradox will not occur generally. However, if the specification of $(Y_1 - Y_2)$ restricts the support of Y_1, then the above probability becomes dependent on the choice of d, and hence Welch's paradox will occur.

For the case of normal distribution, where X_1, X_2 are i.i.d. $\mathcal{N}(\theta, 1)$, we know that $\overline{X}$ is a complete sufficient statistic, while $(X_1 - X_2)$ is ancillary for θ. Therefore, by Basu's theorem, they are independent. Hence,

$$\mathbb{P}\left(\overline{X} - c \leq \theta \leq \overline{X} + c \mid X_1 - X_2 = d\right) = \mathbb{P}\left(\overline{X} - c \leq \theta \leq \overline{X} + c\right)$$

which is simply the frequentist notion of coverage probability of confidence interval $(\overline{X} - c, \overline{X} + c)$.

Problem 9. Let $X_1, \cdots, X_m$ and let $Y_1, \cdots, Y_n$ be two independent random samples from $\mathcal{N}(\mu_1, \sigma^2)$ and $\mathcal{N}(\mu_2, \sigma^2)$ respectively. Assume that the prior distribution of $(\mu_1, \mu_2, \log\sigma^2)$ is improper uniform where (μ_1, μ_2, σ^2) are independent. Find the posterior distribution of $\mu_1 - \mu_2$.

Solution. We first note that, an improper uniform prior distribution of $(\mu_1, \mu_2, \log\sigma^2)$ results in the prior distribution

$$\pi(\mu_1, \mu_2, \sigma^2) \propto \sigma^{-2},\ \sigma^2 > 0$$

Since, $X_1, \ldots X_m \sim \mathcal{N}(\mu_1, \sigma^2)$ and independently $Y_1, \ldots Y_n \sim \mathcal{N}(\mu_2, \sigma^2)$, it follows that the posterior distribution (let $\boldsymbol{X} = X_1, \ldots X_m$ and $\boldsymbol{Y} = Y_1, \ldots Y_n$) is

$$\begin{aligned}\pi(\mu_1, \mu_2, \sigma^2 \mid \boldsymbol{X}, \boldsymbol{Y}) &\propto \frac{1}{(2\pi)^{(m+n)/2}\sigma^{(m+n+2)}} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{m}(X_i - \mu_1)^2 + \sum_{j=1}^{n}(Y_j - \mu_2)^2\right\}\right] \\ &\propto \sigma^{-(m+n+2)} \exp\left[-\frac{1}{2\sigma^2}\left\{S_x^2 + S_y^2 + m(\mu_1 - \overline{X})^2 + n(\mu_2 - \overline{Y})^2\right\}\right]\end{aligned}$$

where $S_x^2 = \sum_{i=1}^m (X_i - \overline{X})^2$ and $S_y^2 = \sum_{j=1}^n (Y_j - \overline{Y})^2$.
We first start by considering the posterior distribution of σ^2. Clearly,

$$\begin{aligned}\pi(\sigma^2 \mid \boldsymbol{X}, \boldsymbol{Y}) &= \int_{\mathbb{R}}\int_{\mathbb{R}} \pi(\mu_1, \mu_2, \sigma^2 \mid \boldsymbol{X}, \boldsymbol{Y}) d\mu_1 d\mu_2 \\ &\propto \sigma^{-(m+n+2)} e^{-\frac{S_x^2 + S_y^2}{2\sigma^2}} \int_{\mathbb{R}} e^{-\frac{m}{2\sigma^2}(\mu_1 - \overline{X})^2} d\mu_1 \int_{\mathbb{R}} e^{-\frac{n}{2\sigma^2}(\mu_2 - \overline{Y})^2} d\mu_2 \\ &\propto \sigma^{-(m+n)} e^{-\frac{S_x^2 + S_y^2}{2\sigma^2}} \int_{\mathbb{R}} \frac{\sqrt{m}}{\sqrt{2\pi}\sigma} e^{-\frac{m}{2\sigma^2}(\mu_1 - \overline{X})^2} d\mu_1 \int_{\mathbb{R}} \frac{\sqrt{n}}{\sqrt{2\pi}\sigma} e^{-\frac{n}{2\sigma^2}(\mu_2 - \overline{Y})^2} d\mu_2 \\ &\propto (\sigma^2)^{-(m+n)/2} e^{-\frac{S_x^2 + S_y^2}{2\sigma^2}}\end{aligned}$$

i.e. the posterior distribution of σ^2 is an inverse-gamma distribution with shape parameter $\left(\dfrac{m+n}{2} - 1\right)$ and scale parameter $\dfrac{S_x^2 + S_y^2}{2}$. Next, we shall try to find the conditional distribution of μ_1, μ_2, conditional on the data $\boldsymbol{X}, \boldsymbol{Y}$ and also the variance parameter σ^2.

$$\begin{aligned}\pi(\mu_1, \mu_2 \mid \sigma^2, \boldsymbol{X}, \boldsymbol{Y}) &\propto \pi(\mu_1, \mu_2, \sigma^2 \mid \boldsymbol{X}, \boldsymbol{Y}) \\ &\propto \exp\left[-\frac{1}{2\sigma^2}\left\{m(\mu_1 - \overline{X})^2 + n(\mu_2 - \overline{Y})^2\right\}\right] \\ &\propto \exp\left[-\frac{m}{2\sigma^2}(\mu_1 - \overline{X})^2\right] \times \exp\left[-\frac{n}{2\sigma^2}(\mu_2 - \overline{Y})^2\right]\end{aligned}$$

In other words, μ_1 and μ_2 are independently distributed given σ^2 and $(\mu_1 \mid \sigma^2, \boldsymbol{X}, \boldsymbol{Y}) \sim \mathcal{N}(\overline{X}, \sigma^2/m)$ and $(\mu_2 \mid \sigma^2, \boldsymbol{X}, \boldsymbol{Y}) \sim \mathcal{N}(\overline{Y}, \sigma^2/n)$. Therefore, it easily follows that $u = (\mu_1 - \mu_2)$, has the conditional distribution

$$u \mid \sigma^2, \boldsymbol{X}, \boldsymbol{Y} \sim \mathcal{N}\left(\overline{X} - \overline{Y}, \sigma^2\left(\frac{1}{m} + \frac{1}{n}\right)\right)$$

Finally, we have the posterior distribution of $u = (\mu_1 - \mu_2)$ given by

$$\begin{aligned}
\pi(u \mid \boldsymbol{X}, \boldsymbol{Y}) &\propto \int_0^\infty \pi(u \mid \sigma^2, \boldsymbol{X}, \boldsymbol{Y}) \pi(\sigma^2 \mid \boldsymbol{X}, \boldsymbol{Y}) d\sigma^2 \\
&\propto \int_0^\infty \frac{\sqrt{mn}}{\sigma\sqrt{(m+n)}} \exp\left[-\frac{mn}{2\sigma^2(m+n)}(u - (\overline{X} - \overline{Y}))^2\right] (\sigma^2)^{-(m+n)/2} e^{-\frac{S_x^2+S_y^2}{2\sigma^2}} d\sigma^2 \\
&\propto \int_0^\infty (\sigma^2)^{-(m+n+1)/2} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{mn}{(m+n)}(u - (\overline{X} - \overline{Y}))^2 + S_x^2 + S_y^2\right)\right] d\sigma^2 \\
&\propto \left[\frac{mn}{(m+n)}(u - (\overline{X} - \overline{Y}))^2 + S_x^2 + S_y^2\right]^{-(m+n-1)/2}
\end{aligned}$$

where the last line follows from identifying the integral as an integral of inverse-gamma density for σ^2. Therefore,

$$\pi(u \mid \boldsymbol{X}, \boldsymbol{Y}) \propto \left[1 + \frac{\dfrac{mn(m+n-2)}{(m+n)} \dfrac{(u - (\overline{X} - \overline{Y}))^2}{(S_x^2 + S_y^2)}}{(m+n-2)}\right]^{-\dfrac{(m+n-2)+1}{2}}$$

i.e. denoting $s^2 = \dfrac{1}{(m+n-2)}\left[\sum_{i=1}^m (X_i - \overline{X})^2 + \sum_{j=1}^n (Y_j - \overline{Y})^2\right]$, we have

$$T = \frac{(\mu_1 - \mu_2) - (\overline{X} - \overline{Y})}{s\sqrt{\frac{1}{m} + \frac{1}{n}}} \mid \boldsymbol{X}, \boldsymbol{Y} \sim t_{m+n-2}$$

Therefore, $(\mu_1 - \mu_2)$, when properly centered and scaled as shown above, has a posterior distribution as the student's t distribution with $(m+n-2)$ degrees of freedom.

Problem 10. Let $X_1, X_2, \dots X_n$ be i.i.d. $\sim \mathcal{N}(\theta, \sigma^2)$, σ^2 known. Assume $\sigma^2 = 1$.

(a) Consider the problem of $H_0 : \theta \le \theta_0$ vs $H_1 : \theta > \theta_0$. We reject H_0 if $T = \sqrt{n}(\overline{X} - \theta_0)$ is large. A classical (frequentist) measure of evidence against H_0 is the P-value defined by

$$P = \sup_{\theta \le \theta_0} P_\theta[\sqrt{n}(\overline{X} - \theta_0) > t]$$

where t is the observed value of T (We reject H_0 at level α if $P \le \alpha$). Find the P-value (in terms of t). Consider now the uniform prior $\pi(\theta) \equiv 1$. Find the posterior probability of H_0 (note that it is the same as the P-value).

(b) Suppose we want to test $H_0 : \theta = \theta_0$ vs $H_1 : \theta \ne \theta_0$. We reject H_0 if $T = |\sqrt{n}(\overline{X} - \theta_0)|$ is large. Here P-value $= P_{\theta_0}[|\sqrt{n}(\overline{X} - \theta_0)| > t]$, where t is the observed value of T. Find the P-value in terms of t. Consider now a $\mathcal{N}(\theta_0, 1)$ prior for θ under H_1 and find the Bayes factor BF_{01}. Assuming prior probabilities $P(H_0) = P(H_1) = (1/2)$, find the posterior probability $P(H_0 \mid X_1, X_2, \dots X_n)$. When $n = 50, t = 1.96$, show that P-value is 0.05, $BF_{01} = 1.08$ and $\mathbb{P}(H_0 \mid X_1, \dots X_n) = 0.52$. (This shows a conflict between frequentist and Bayesian answers.)

Solution. (a) Under θ, we know that $\overline{X} \sim \mathcal{N}(\theta, 1/n)$ (since $\sigma^2 = 1$ is known), i.e. $\sqrt{n}(\overline{X} - \theta) \sim \mathcal{N}(0, 1)$. Therefore, the P-value can be expressed as

$$\begin{aligned}
P &= \sup_{\theta \le \theta_0} P_\theta[\sqrt{n}(\overline{X} - \theta_0) > t] \\
&= \sup_{\theta \le \theta_0} P_\theta[\sqrt{n}(\overline{X} - \theta) > t + \sqrt{n}(\theta_0 - \theta)] \\
&= \sup_{\theta \le \theta_0} \left(1 - \Phi(t + \sqrt{n}(\theta_0 - \theta))\right) \quad \text{where, } \Phi(\cdot) \text{ is the standard normal c.d.f.} \\
&= 1 - \inf_{\theta \le \theta_0} \Phi\left(t + \sqrt{n}(\theta_0 - \theta)\right) \\
&= 1 - \Phi(t), \qquad \text{by non-decreasing nature of } \Phi
\end{aligned}$$

Now, if we consider the uniform improper prior $\pi(\theta) \equiv 1$, then note that the posterior distribution can be expressed as,

$$\begin{aligned}
\pi(\theta \mid X_1, \dots X_n) &\propto \pi(\theta) \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(X_i - \theta)^2} \\
&\propto \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(X_i - \theta)^2\right] \\
&\propto \exp\left[-\frac{1}{2}(n\theta^2 - 2n\overline{X})\right]
\end{aligned}$$

which is proportional to the normal density with mean $\overline{X}$ and variance $(1/n)$. In other words, the posterior distribution of $\sqrt{n}(\theta - \overline{X})$ is a standard normal distribution. Thus, the posterior probability of H_0 can be obtained as

$$
\begin{aligned}
\mathbb{P}(H_0 \mid X_1, \dots X_n) &= \int_{\theta \leq \theta_0} \pi(\theta \mid X_1, \dots X_n) d\theta \\
&= \mathbb{P}_{\theta \mid X_1, \dots X_n} (\theta \leq \theta_0) \\
&= \mathbb{P}_{\theta \mid X_1, \dots X_n} \left(\sqrt{n}(\theta - \overline{X}) \leq \sqrt{n}(\theta_0 - \overline{X}) \right) \\
&= \Phi \left(\sqrt{n}(\theta_0 - \overline{X}) \right)
\end{aligned}
$$

Since, the observed value of the statistic T is $t = \sqrt{n}(\overline{X} - \theta_0)$, we have $\mathbb{P}(H_0 \mid X_1, \dots X_n) = \Phi(-t) = 1 - \Phi(t)$, the same quantity as P-value.

(b) Note that, under θ_0, $\overline{X} \sim \mathcal{N}(0, 1/n)$. In this case of testing two sided hypothesis, the P-value can be expressed as

$$
\begin{aligned}
P &= P_{\theta_0}[\sqrt{n}|\overline{X} - \theta_0| > t] \\
&= 1 - P_{\theta_0}[\sqrt{n}|\overline{X} - \theta_0| \leq t] \\
&= 1 - P_{\theta_0}[(-t) \leq \sqrt{n}(\overline{X} - \theta_0) \leq t] \\
&= 1 - 2P_{\theta_0}[0 \leq \sqrt{n}(\overline{X} - \theta_0) \leq t] \quad \text{by symmetry of normal distribution about } \theta_0 \\
&= 1 - 2(\Phi(t) - 1/2), \qquad \text{since, } \Phi(0) = (1/2) \\
&= 2(1 - \Phi(t))
\end{aligned}
$$

Now, the Bayes Factor can be obtained as,

$$
BF_{01} = \frac{\prod_{i=1}^n f(X_i, \theta_0)}{\int_{\mathbb{R}} \prod_{i=1}^n f(X_i, \theta) \pi(\theta) d\theta} \tag{5}
$$

where $f(X_i, \theta)$ is the $\mathcal{N}(\theta, 1)$ density and $\pi(\theta)$ is $\mathcal{N}(\theta_0, 1)$ density. The denominator in Eq. (5) can be explicity computed as follows

$$
\begin{aligned}
& \int_{\mathbb{R}} \prod_{i=1}^n f(X_i, \theta) \pi(\theta) d\theta \\
= \quad & \int_{\mathbb{R}} \left(\frac{1}{\sqrt{2\pi}} \right)^{(n+1)} e^{-\frac{1}{2}\left[\sum_{i=1}^n (X_i - \theta)^2 + (\theta - \theta_0)^2\right]} d\theta \\
= \quad & \left(\frac{1}{\sqrt{2\pi}} \right)^{(n+1)} \exp\left[-\frac{(\sum_{i=1}^n X_i^2 + \theta_0^2)}{2} \right] \int_{\mathbb{R}} \exp\left[-\frac{1}{2} \left(\theta^2 (n+1) - 2\theta(n\overline{X} + \theta_0) \right) \right] d\theta
\end{aligned}
$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{(n+1)} \exp\left[-\frac{(\sum_{i=1}^n X_i^2 + \theta_0^2)}{2}\right] \frac{\sqrt{2\pi}}{\sqrt{(n+1)}} \exp\left[\frac{(n\overline{X} + \theta_0)^2}{2(n+1)}\right]$$
$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{n} \frac{1}{\sqrt{(n+1)}} \exp\left[-\frac{(\sum_{i=1}^n X_i^2 + \theta_0^2)}{2} + \frac{(n\overline{X} + \theta_0)^2}{2(n+1)}\right]$$

and the numerator is simply, $\left(\frac{1}{\sqrt{2\pi}}\right)^{n} \exp\left[-\frac{1}{2}\sum_{i=1}^n (X_i - \theta_0)^2\right]$. Therefore,

$$BF_{01} = \sqrt{(n+1)} \exp\left[-\frac{1}{2}\left(\sum_{i=1}^n X_i^2 - 2n\overline{X}\theta_0 + n\theta_0^2 - \sum_{i=1}^n X_i^2 - \theta_0^2 + \frac{n^2\overline{X}^2}{(n+1)} + \frac{2n\overline{X}\theta_0}{(n+1)} + \frac{\theta_0^2}{(n+1)}\right)\right]$$
$$= \sqrt{(n+1)} \exp\left[-\frac{1}{2}\left(\theta_0^2\frac{n^2}{(n+1)} + \frac{n^2\overline{X}^2}{(n+1)} - \frac{2n^2\overline{X}\theta_0}{(n+1)}\right)\right]$$
$$= \sqrt{(n+1)} \exp\left[-\frac{n^2}{2(n+1)}(\overline{X} - \theta_0)^2\right]$$
$$= \sqrt{(n+1)} \exp\left[-\frac{n}{2(n+1)}t^2\right]$$

where $t = \sqrt{n}(\overline{X} - \theta)$. Hence, the posterior probability is

$$\mathbb{P}(H_0 \mid X_1, \ldots X_n) = \frac{\mathbb{P}(X_1, \ldots X_n \mid H_0)P(H_0)}{\mathbb{P}(X_1, \ldots X_n \mid H_0)P(H_0) + \mathbb{P}(X_1, \ldots X_n \mid H_1)P(H_1)}$$
$$= \frac{\mathbb{P}(X_1, \ldots X_n \mid H_0)}{\mathbb{P}(X_1, \ldots X_n \mid H_0) + \mathbb{P}(X_1, \ldots X_n \mid H_1)}, \quad \text{since, } \mathbb{P}(H_0) = \mathbb{P}(H_1) = (1/2)$$
$$= \frac{1}{1 + BF_{01}^{(-1)}}$$
$$= \frac{1}{1 + \frac{1}{\sqrt{n+1}}\exp\{\frac{nt^2}{2(n+1)}\}}$$

Clearly, when $n = 50$ and $t = 1.96$, P-value is $1 - \Phi(1.96) = (1 - 0.95) = 0.05$. In this case, the Bayes factor is $BF_{01} = \sqrt{51}\exp\left[-\frac{50}{102} \times 1.96^2\right] = 1.07559526995 \approx 1.08$. Also, the posterior probability of H_0 then turns out to be $1/(1 + (1/1.08)) = 0.5192307 \approx 0.52$.

Problem 11. Let the sample space be $\{1, 2, \dots k\}$, and $P = (p_1, \dots, p_k)$ be a random probability distribution on this sample sapce. Let $X_1, \dots X_n$ be i.i.d. $\sim P$, and $P \sim \text{Dirichlet}(\alpha_1, \dots \alpha_k), \alpha_i > 0 \ \forall i$. Show that for any subset A of the sample space, the posterior mean of $P(A)$ is a weighted average of its prior mean and $P_n(A)$, where P_n denotes the empirical distribution of $X_1, \dots X_n$.

Solution. Let, $n_j,\ j = 1, 2, \dots k$, denotes the number of samples X_i that are exactly equal to $j \in \Omega = \{1, 2, \dots k\}$. Therefore, $n_j = \sum_{i=1}^n \mathbf{1}_{\{X_i = j\}}$. Now, the posterior distribution can be expressed as

$$\begin{aligned}
\pi(\theta \mid X_1, \dots X_n) &\propto \prod_{j=1}^k p_j^{\sum_{i=1}^n \mathbf{1}_{\{X_i = j\}}} \times \prod_{j=1}^k p_j^{\alpha_j - 1} \\
&= \prod_{j=1}^k p_j^{n_j} \times \prod_{j=1}^k p_j^{\alpha_j - 1} \\
&= \prod_{j=1}^k p_j^{n_j + \alpha_j - 1}
\end{aligned}$$

Therefore, the posterior is a Dirichlet distribution with parameters $(n_1 + \alpha_1, n_2 + \alpha_2, \dots n_k + \alpha_k)$. Now, let us consider the expectation of a single probability p_i, for $i = 1, 2, \dots k$. Therefore, the posterior mean of p_i is

$$\mathbb{E}(p_i \mid X_1, \dots X_n) = \frac{n_i + \alpha_i}{\sum_{j=1}^k (n_j + \alpha_j)},\ i = 1, 2, \dots k$$

Since $A \subseteq \{1, 2, \dots k\}$, then $P(A) = \sum_{i=1}^k p_i \mathbf{1}_{\{i \in A\}}$, where $\mathbf{1}_{\{i \in A\}}$ is the indicator whether i is contained in the chosen set A (clearly $\mathbf{1}_{\{i \in A\}}$ is a real constant). Therefore,

$$\begin{aligned}
\mathbb{E}(P(A) \mid X_1, \dots X_n) &= \mathbb{E}\left(\sum_{i=1}^k p_i \mathbf{1}_{\{i \in A\}} \mid X_1, \dots X_n \right) \\
&= \sum_{i=1}^k \mathbf{1}_{\{i \in A\}} \mathbb{E}(p_i \mid X_1, \dots X_n) \\
&= \sum_{i=1}^k \mathbf{1}_{\{i \in A\}} \frac{(n_i + \alpha_i)}{\sum_{j=1}^k (n_j + \alpha_j)} \\
&= \sum_{i=1}^k \mathbf{1}_{\{i \in A\}} \left[\frac{n_i}{\sum_{j=1}^k n_j} \frac{\sum_{j=1}^k n_j}{\sum_{j=1}^k (n_j + \alpha_j)} + \frac{\alpha_i}{\sum_{j=1}^k \alpha_j} \frac{\sum_{j=1}^k \alpha_j}{\sum_{j=1}^k (n_j + \alpha_j)} \right] \\
&= \frac{\sum_{j=1}^k n_j}{\sum_{j=1}^k (n_j + \alpha_j)} \left(\sum_{i=1}^k \mathbf{1}_{\{i \in A\}} \frac{n_i}{\sum_{j=1}^k n_j} \right) + \frac{\sum_{j=1}^k \alpha_j}{\sum_{j=1}^k (n_j + \alpha_j)} \left(\sum_{i=1}^k \mathbf{1}_{\{i \in A\}} \frac{\alpha_i}{\sum_{j=1}^k \alpha_j} \right) \\
&= \frac{\sum_{j=1}^k n_j}{\sum_{j=1}^k (n_j + \alpha_j)} P_n(A) + \frac{\sum_{j=1}^k \alpha_j}{\sum_{j=1}^k (n_j + \alpha_j)} \mathbb{E}(P(A))
\end{aligned}$$

where $\mathbb{E}(P(A))$ is the prior mean of $P(A)$. Also, the $P_n(A)$ is obtained by the similar formula of $P(A)$ where each probability is obtained by its empirical estimate, $\widehat{p}_i = n_i/(\sum_{j=1}^{k} n_j)$.

Thank You