Problem 0: Readings
Read first part of note 6 on the course web page.

Problem 1: Normal observations
Suppose that you are given 12 observations from a normal distribution:
15.644  16.437  17.287  14.448
15.308  15.169  18.123  17.635
17.259  16.311  15.390  17.252
and we are told that the variance $\sigma^2 = 1$. Find the 90% high density region for the posterior distribution of the mean using a suitable prior of your choice. The 90% high density region is defined as the region that leaves 5% probability on each side.

Solution: We compute $\bar{x} = 16.355$. With a conjugate prior $N(\beta, \tau^2)$ for $\mu$ we get (see note 6):
$$
\mu | \bar{x} \sim N\left(\frac{\sigma^2 \beta / n + \tau^2 \bar{x}}{\sigma^2 / n + \tau^2}, \frac{\sigma^2 \tau^2 / n}{\sigma^2 / n + \tau^2}\right)
$$
Now we have two choices: either set values for $\beta$ and $\tau^2$, or assume that we have no such knowledge. The latter case is easier, and we can do it by making $\tau^2 \to \infty$ (the prior becomes improper uniform). In this case, we get:
$$
\mu | \bar{x} \sim (\bar{x}, \sigma^2 / n) = N(16.355, 1/12)
$$
The standard normal $N(0, 1)$ gives 90% high density region in $[-1.65, 1.65]$; therefore,
$$
\frac{\mu_1 - 16.355}{\sqrt{1/12}} = -1.65 \Rightarrow \mu_1 = 15.879
$$
$$
\frac{\mu_2 - 16.355}{\sqrt{1/12}} = 1.65 \Rightarrow \mu_2 = 16.831
$$
$$
\mu \in [15.879, 16.831]
$$

Problem 2: Two groups
Consider the following two groups:
\[
\begin{array}{c|c}
\bar{x} = 216 & \bar{y} = 210 \\
\sigma_x^2 = 2210 & \sigma_y^2 = 2618 \\
n = 130 & m = 119 \\
\end{array}
\]

where given \(\mu_x, x_1, \ldots, x_n\) are independent and identically distributed normal random variables with mean \(\mu_x\) and variance \(\sigma_x^2\) (which is known). Same for the \(y\)'s.

(a) Use a P-value approach to decide whether \(\mu_x = \mu_y\).

**Solution:** we compute
\[
z = (\bar{x} - \bar{y}) / \sqrt{\sigma_x^2/n + \sigma_y^2/m} = 0.96.
\]
Looking at the normal table, we find that P-value = \(P(z > 0.96) \approx 0.1685\), which is not small enough to reject the hypothesis that \(\mu_x = \mu_y\).

(b) Use a Bayesian approach with an improper uniform prior to find \(P(\mu_x - \mu_y > b | \bar{x}, \bar{y})\) for \(b = 0, 1, 2, \ldots, 10\). Plot the curve \(P(\mu_x - \mu_y > b | \bar{x}, \bar{y})\) vs. \(b\) to have a visual representation.

**Solution:** With an improper uniform prior we have:
\[
\mu_x - \mu_y | \bar{x}, \bar{y} \sim N(\bar{x} - \bar{y}, \sigma_x^2/n + \sigma_y^2/m) = N(6, 39)
\]

Therefore,
\[
P(\mu_x - \mu_y > b | \bar{x}, \bar{y}) = 1 - \Phi((b - 6)/\sqrt{39}) = \Phi((6 - b)/\sqrt{39}).
\]

<table>
<thead>
<tr>
<th>(b)</th>
<th>(0.8315)</th>
<th>(0.7811)</th>
<th>(0.7389)</th>
<th>(0.6844)</th>
<th>(0.6255)</th>
<th>(0.5636)</th>
<th>(0.5)</th>
<th>(0.4364)</th>
<th>(0.3745)</th>
<th>(0.3156)</th>
<th>(0.2611)</th>
</tr>
</thead>
</table>

(c) Compare both approaches.

**Solution:** In the first approach, we simply declare that \(\mu_x = \mu_y\). In the second approach, we have better information. For instance, both \(\bar{x}\) and \(\bar{y}\) are in the 200 range. If we assume that 5% makes a considerable difference, which is about 10, then we can clearly see that \(P(\mu_x - \mu_y > 10 | \bar{x}, \bar{y})\) is 0.26. So, there is about 25% chance that \(\mu_x\) and \(\mu_y\) are “different”.

**Problem 3: Cauchy distribution**

The Cauchy distribution given by:
\[
f(x) = \frac{1}{\pi} \frac{b}{b^2 + x^2}
\]
where \(b > 0\). It is known that if \(x_1, \ldots, x_n\) are Cauchy independent observations, then \(\bar{x} = (x_1 + \ldots + x_n)/n\) also satisfies:
\[
f(\bar{x}) = \frac{1}{\pi} \frac{b}{b^2 + \bar{x}^2}
\]
(a) Why is that not a contradiction of the central limit theorem?

*(Hint: Find $E[X]$)*

**Solution:** The central limit theorem requires finite mean and variance. For the Cauchy distribution, the mean (and variance) are not defined.

Let $u = b^2 + x^2$, then $du = 2xdx$ and $u \in [b^2, \infty)$ when $x \in (-\infty, 0]$ and when $x \in [0, \infty)$:

$$E[X] = \int_{-\infty}^{\infty} \frac{bx}{\pi(b^2 + x^2)} dx = \int_{x\leq 0} \frac{bx}{\pi(b^2 + x^2)} dx + \int_{x\geq 0} \frac{bx}{\pi(b^2 + x^2)} dx =$$

$$2 \int_{b^2}^{\infty} \frac{b}{2\pi} \frac{du}{u} = \frac{b}{\pi} [\log \infty - \log b^2] = \infty$$

(b) Assume that $b$ is unknown and consider a prior for $b$ of the form

$$f(b) \propto \frac{1}{b}$$

What kind of prior is that?

**Solution:** This is an improper prior.

(c) Find the posterior density of $b$, i.e. $f(b|\bar{x})$.

**Solution:**

$$f(b|\bar{x}) \propto f(\bar{x}|b)f(b) \propto \frac{1}{x^2 + b^2}$$

Therefore, since $\int_{0}^{\infty} f(b|\bar{x})db = 1$, we have

$$f(b|\bar{x}) = \frac{1}{\pi} \frac{2\bar{x}}{x^2 + b^2}$$