From the plot of the log-likelihood functions based on Data Set #1 and #2, we can see that both curves are almost in the same shape and reach their maximum at around \( \theta = 2 \), which does provide some reliable indication that both data sets are generated from the given model with \( \theta = 2 \).
From the plot above, we can see that the curve of the log-likelihood function of the model with left censoring is in the same shape with the one of the original model; and they almost have the same vertical position.

(c)

\[
F(x|\theta) = 1 - \exp(-x/\theta) \quad \text{for} \quad X \sim \text{Exp}(\theta),
\]

\[
L_i(\theta) = \prod_{i=1}^{10} (F(x_i + r|\theta) - F(x_i - r|\theta))
\]

\[
= \prod_{i=1}^{10} \left( \exp\left(\frac{-x_i - r}{\theta}\right) + \exp\left(\frac{-x_i + r}{\theta}\right) \right),
\]

where \( r = \begin{cases} 
0.0005, & \text{for 3 decimal place} \\
0.005, & \text{for 2 decimal place} \\
0.05, & \text{for 1 decimal place}
\end{cases} \)

The curves of the log-likelihood function of the model II rounding to 3, 2, and 1 decimal place are all in the same pattern with the one of original model I.

# 3.

(1) The 1st experiment: \( \text{Bin}(100, p_{++}) \)

(2) The 2nd experiment: \( \text{Bin}(100, p_{++}) \)

(3) The 3rd experiment: \( \text{Multinomial}(50, p_{++}, p_{+-}, p_{-+}) \quad p_{-} = 1 - (p_{++} + p_{+-} + p_{-+}) \)

where \( p_{++} = p_{++} + p_{--} \) and \( p_{++} = p_{++} + p_{-+}. \)

Since these three experiments are mutually independent,
\[
\log L(p|x) = \log\left(\frac{100!}{10!90!} (p_+ \cdot)^{10} (1-p_+)^{90}\right) + \log\left(\frac{100!}{20!80!} (p_- \cdot)^{10} (1-p_-)^{80}\right)
+ \log\left(\frac{50!}{4!20!36!} (p_{+ +})^4 (p_{+ -})^2 (p_{- +})^2 (1-p_{+ +} - p_{+ -} - p_{- +})^{36}\right).
\]

# 4.

(a)

\[
\pi(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\int \limits_{\theta} p(x|\theta)\pi(\theta)d\theta} = \frac{\frac{2^x x!}{\theta^2} I(0 < x < \theta)I(0 \leq \theta \leq 1)}{\int \limits_{x}^{1} \frac{2^x x!}{\theta^2} d\theta} = \frac{\frac{2^x x!}{\theta^2} I(0 < x < \theta \leq 1)}{\frac{[-\frac{2x}{\theta}]^1}{1}}
= \frac{x}{\theta^2(1-x)} I(0 < x < \theta \leq 1)
\]

(b)

\[
\pi(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{\int \limits_{\theta} p(x|\theta)\pi(\theta)d\theta} = \frac{\frac{2^x x!}{\theta^2} 3^2 I(0 < x < \theta)I(0 \leq \theta \leq 1)}{\int \limits_{x}^{1} \frac{2^x x!}{\theta^2} 3^2 d\theta} = \frac{\frac{6x x!}{\theta^2} I(0 < x < \theta \leq 1)}{[2x(3\theta)]_x^1}
= \frac{6x}{6x(1-x)} I(0 < x < \theta \leq 1) = \frac{1}{(1-x)} I(0 < x < \theta \leq 1)
\]

(c)

For the prior in (a),

\[
E(\theta|X) = \int \limits_{x}^{1} \frac{2^x x!}{\theta^2(2-2x)} d\theta = \int \limits_{x}^{1} \frac{1}{2^x (2-2x)} \frac{1}{\theta} d\theta = \frac{2x}{(2-2x)} \left[ \log \theta \right]_x^1
= \frac{x}{(1-x)} \left( -\log x \right), \quad 0 < x < 1.
\]

For the prior in (b),

\[
E(\theta|X) = \int \limits_{x}^{1} \frac{1}{(1-x)} \frac{1}{\theta} d\theta = \frac{1}{(1-x)} \left[ \frac{1}{2} \theta^2 \right]_x^1 = \frac{1-x^2}{2(1-x)}
= \frac{1+x}{2}, \quad 0 < x < 1.
\]
(d) 
\[
\pi(\theta|\mathbf{x}) = \frac{\int_{\theta} p(x|\theta)\pi(\theta)p(\theta)d\theta}{\int_{\theta} \prod_{i=1}^{n} \left(\frac{2x_i}{\theta^2}\right)I(0 < x_i < \theta)d\theta} = \frac{\prod_{i=1}^{n} \left(2x_i\right)I(0 < x_{(n)} < \theta < 1)}{\int_{x_{(n)}}^{1} \prod_{i=1}^{n} \left(\frac{2x_i}{\theta^2}\right)d\theta}
\]

\[
= \frac{\prod_{i=1}^{n} x_i I(x_{(n)} < \theta < 1)}{\prod_{i=1}^{n} x_i \frac{1}{\theta^{2n}} (1 - x_{(n)}^{1 - 2n})}
\]

where \(x_{(n)}\) is the maximum of \(x_1, \ldots, x_n\).

\[\square\] B&D 1.2.3.

(a) Given the prior distribution of \(\theta\), \(\pi(\theta) = \frac{1}{3}\) for \(\theta = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}\).

The posterior density of \(\theta\) given \(X = 2\) is as follows.

\[
\pi(\theta|x = 2) = \frac{p(x = 2|\theta)\pi(\theta)}{\sum_{\theta} p(x = 2|\theta)\pi(\theta)} = \frac{(1 - \theta)^2 \theta(1/3)}{(1 - 1/4)^2(1/4) + (1 - 1/2)^2(1/2) + (1 - 3/4)^2(3/4)(1/3)} = \frac{(1 - \theta)^2 \theta}{5/16},
\]

where \(\theta = \frac{1}{4}, \frac{1}{2}, \) or \(\frac{3}{4}\).

(b) 
\[
\pi(\theta|x = 2) = \begin{cases} 
9/20, & \theta = 1/4 \\
2/5, & \theta = 1/2 \\
3/20, & \theta = 3/4 
\end{cases}
\]

The most probable value of \(\theta\) given \(x = 2\) is \(\theta = 1/4\).

Given \(x = k\),
\[
\pi(\theta|x = 2) = \frac{p(x = k\theta)\pi(\theta)}{\sum_{\theta}p(x = k\theta)\pi(\theta)} = \frac{(1-\theta)^{k\theta}(1/3)}{(1-1/4)^k(1/4) + (1-1/2)^k(1/2) + (1-3/4)^k(3/4)}(1/3)
\]
\[
= \frac{4^{k+1}(1-\theta)^{k\theta}}{3^k + 2^{k+1} + 3}
\]
\[
= \begin{cases} 
(3^k/(3^k + 2^{k+1} + 3)), & \theta = 1/4 \\
(2^{k+1}/(3^k + 2^{k+1} + 3)), & \theta = 1/2, \\
(3/(3^k + 2^{2k+1} + 3)), & \theta = 3/4
\end{cases}
\]
The most probable values of \(\theta\) given \(x = k\) are \(\theta = 3/4\) if \(k = 0\), \(\theta = 1/2\) if \(k = 1\) and \(\theta = 1/4\) if \(k \geq 2\).

(c)

\[
\pi(\theta|X = k) = \frac{p(X = k\theta)\pi(\theta)}{\int_{\theta} p(X = k\theta)\pi(\theta) d\theta} = \frac{(1-\theta)^{k\theta} \frac{1}{\beta(r,s)} \theta^{-1}(1-\theta)^{r-1}I(0 < \theta < 1)}{\int_{0}^{1} (1-\theta)^{k\theta} \frac{1}{\beta(r,s)} \theta^{-1}(1-\theta)^{r-1} d\theta}
\]
\[
= \frac{(1-\theta)^{s+k-1} \theta^r I(0 < \theta < 1)}{\int_{0}^{1} (1-\theta)^{s+k-1} \theta^r d\theta} = \frac{1}{\beta(r+1, s+k)} (1-\theta)^{s+k-1} \theta^r I(0 < \theta < 1)
\]
Therefore, \(\theta|X = k \sim Beta(r+1, s+k)\).

© B&D 1.2.14.

(a)

\[
f(x_{n+1}|\theta) = \frac{1}{\sqrt{2\pi} \sigma_0} e^{\frac{-(x_{n+1} - \theta)^2}{2\sigma_0^2}} \text{ where } \pi(\theta) = \frac{1}{\sqrt{2\pi} \tau_0} e^{\frac{-(\theta - \theta_0)^2}{2\tau_0^2}}
\]
\[
f(x_{n+1}) = \int_{-\infty}^{\infty} f(x_{n+1}|\theta)\pi(\theta) d\theta
\]
\[
= \frac{1}{2\pi \sigma_0 \tau_0} \int_{-\infty}^{\infty} \exp\left\{-(\frac{(x_{n+1} - \theta)^2}{2\sigma_0^2} - \frac{(\theta - \theta_0)^2}{2\tau_0^2})\right\} d\theta
\]
\[
= \frac{1}{2\pi \sigma_0 \tau_0} \int_{-\infty}^{\infty} \exp\left\{-(\theta - \frac{x_{n+1}}{\sigma_0^2} + \frac{\sigma_0^2 \theta_0}{\sigma_0^2 + \tau_0^2}) \right\}
\]
\[
\text{ where } \frac{(x_{n+1} - \theta_0)^2}{\sigma_0^2 + \tau_0^2} = \frac{2\sigma_0^2 \tau_0^2}{\sigma_0^2 + \tau_0^2}
\]
Thus,  \( X_{n+1} \sim \mathcal{N}(\theta_0, \tau_0^2 + \sigma_0^2) \).

Since

\[
    f(x_1, \cdots, x_{n+1} | \theta) = \prod_{i=1}^{n+1} \frac{1}{\sqrt{2\pi} \sigma_0} \exp \left\{ -\frac{(x_i - \theta)^2}{2\sigma_0^2} \right\},
\]

and

\[
    \pi(\theta) = \frac{1}{\sqrt{2\pi} \tau_0} \exp \left\{ -\frac{(\theta - \theta_0)^2}{2\tau_0^2} \right\},
\]

\[
    f(x_1, \cdots, x_{n+1}) = \int_{-\infty}^{\infty} f(x_1, \cdots, x_{n+1} | \theta) \pi(\theta) \, d\theta
\]

\[
= \frac{1}{(2\pi\sigma_0^2)^{(n+1)/2} \tau_0} \int_{-\infty}^{\infty} \exp \left\{ -\frac{\sum_{i=1}^{n+1} (x_i - \theta)^2}{2\sigma_0^2} \right\} \exp \left\{ -\frac{(\theta - \theta_0)^2}{2\tau_0^2} \right\} d\theta
\]

\[
= \frac{1}{(2\pi\sigma_0^2)^{(n+1)/2} \tau_0} \left\{ \left( \frac{\tau_0^2 \sum_{i=1}^{n+1} x_i + \sigma_0^2 \theta_0}{(n+1)\tau_0^2 + \sigma_0^2} \right)^2 + \frac{(n+1)\tau_0^2 \sum_{i=1}^{n+1} (x_i - \bar{x})^2 + \sigma_0^2 \tau_0 \sum_{i=1}^{n+1} (x_i - \theta_0)^2}{(n+1)\tau_0^2 + \sigma_0^2} \right\}
\]

\[
= \frac{2\tau_0^2 \sigma_0^2}{(n+1)\tau_0^2 + \sigma_0^2}
\]

\[
= \frac{1}{(2\pi)^{n/2} \sigma_0^2 \sqrt{(n+1)\tau_0^2 + \sigma_0^2}} \exp \left\{ \frac{(n+1)\tau_0^2 \sum_{i=1}^{n+1} (x_i - \bar{x})^2 + \sigma_0^2 \tau_0 \sum_{i=1}^{n+1} (x_i - \theta_0)^2}{2((n+1)\tau_0^2 + \sigma_0^2)^2} \right\}.
\]

Similarly,

\[
    f(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{(n-1)/2} \sigma_0^{n-1} \sqrt{n\tau_0^2 + \sigma_0^2}} \exp \left\{ -\frac{n\tau_0^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 + \sigma_0^2 \tau_0 \sum_{i=1}^{n} (x_i - \theta_0)^2}{2((n)\tau_0^2 + \sigma_0^2)^2} \right\}.
\]

Thus, \( \frac{f(x_{n+1} | x_1, \cdots, x_n)}{f(x_1, \cdots, x_n)} = \frac{\sqrt{n\sigma_0^2 + \tau_0^2}}{\sqrt{2\pi} \sigma_0 \sqrt{(n+1)\sigma_0^2 + \tau_0^2}} \).
\[
\frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\sigma_0^2(1 + \frac{\gamma_0^2}{n\tau_0^2 + \sigma_0^2})}} \cdot \exp\left(-\frac{\left(x_{n+1} - \frac{n\tau_0^2 x_n + \sigma_0^2 \theta_0}{n\tau_0^2 + \sigma_0^2}\right)^2}{2\sigma_0^2\left(1 + \frac{\gamma_0^2}{n\tau_0^2 + \sigma_0^2}\right)}\right).
\]

Then, the posterior predictive distribution is as follows.
\[
X_{n+1} | X_1, \ldots, X_n \sim N\left(\frac{n\tau_0^2 x_n + \sigma_0^2 \theta_0}{n\tau_0^2 + \sigma_0^2}, \frac{\gamma_0^2}{n\tau_0^2 + \sigma_0^2}\right).
\]

**Solution 2**

Consider \(X_i = \theta + \epsilon_i\), where \(\epsilon_i \sim iid \ N(0, \sigma_0^2)\), and \(\theta \sim N(\theta_0, \tau_0^2)\) and \(\epsilon_i\) and \(\theta\) are independent. Then,
\[
(X_1, X_2, \ldots, X_n, \theta^T - MVN(\theta_0, \Sigma),
\]
where \(\Sigma = \sigma_0^2 I_{(n+2) \times (n+2)} + \frac{\gamma_0^2}{\tau_0} I_{(n+2) \times (n+2)}\) and \(J\) is a matrix of 1's.

So,
\[
(X_1, X_2, \ldots, X_n)^T \sim MVN(\theta_0, \Sigma_{XX}),
\]
where \(\Sigma_{XX} = \sigma_0^2 I_{n \times n} + \frac{\gamma_0^2}{\tau_0} I_{n \times n}\). Also, \(X_{n+1} \sim N(\theta_0, \sigma_0^2 + \tau_0^2)\).

Let \(W = (X_1, \ldots, X_n)^T\) and \(Z = X_{n+1}\). Then \(W \sim MVN(\theta_0, \Sigma_{XX}), Z \sim N(\theta_0, \sigma_0^2 + \tau_0^2)\)
and \(\Sigma_{WZ} = \frac{\gamma_0^2}{\tau_0} I_{(n-1) \times 1}\).

By using the formula of
\[
E(Z | W) = \mu_Z + \Sigma_{WZ}^{-1} \Sigma_{WZ} (W - \mu_W),
\]
and
\[
\text{Cov}(Z | W) = \Sigma_{ZZ} - \Sigma_{ZW} \Sigma_{WZ}^{-1} \Sigma_{WZ},
\]
we can get
\[
E(X_{n+1} | X_1, \ldots, X_n) = \frac{n\tau_0^2 x_n + \sigma_0^2 \theta_0}{n\tau_0^2 + \sigma_0^2}
\]
and
\[
\text{Var}(X_{n+1} | X_1, \ldots, X_n) = \frac{\gamma_0^2}{n\tau_0^2 + \sigma_0^2}.
\]

Hence, \(X_{n+1} | X_1, \ldots, X_n \sim N\left(\frac{n\tau_0^2 x_n + \sigma_0^2 \theta_0}{n\tau_0^2 + \sigma_0^2}, \sigma_0^2\left(1 + \frac{\gamma_0^2}{n\tau_0^2 + \sigma_0^2}\right)\right)\).

* Notice that the inverse of a matrix in the form of
\[
aI + bJ
\]
has the form of
\[ \alpha I + \beta J \]

for \( \alpha = \frac{1}{a} \) and \( \beta = -\frac{b}{a(a+kb)} \), where \( k \) is the dimension of the square matrix \( I \) (or \( J \)).

(b) 
As \( n \to \infty \), the posterior predictive distribution becomes close to \( N(\theta_0, \sigma_0^2) \), while the predictive distribution of \( X_{n+1} \) is still \( N(\theta_0, r_0^2 + \sigma_0^2) \).

© B&D 1.3.3.
(a) 
\[ R(\theta, r_{r,s}) = E_{\theta}(L(\theta, r_{r,s}(X))) \]

\[
\begin{align*}
&= \begin{cases} 
0P(X < r) + cP(r < X < s) + (b + c)P(X > s), & \theta < 0 \\
(b + c)P(X < r) + 0P(r < X < s) + bP(X > s), & \theta = 0 \\
bP(X < r) + cP(r < X < s) + 0(b + c)P(X > s), & \theta > 0 
\end{cases} \\
&= \begin{cases} 
c(\Phi(\sqrt{n}(s-\theta)) - \Phi(\sqrt{n}(r-\theta))) + (b + c)(1 - \Phi(\sqrt{n}(s-\theta))), & \theta < 0 \\
\Phi(\sqrt{n}(r-\theta)) + b(1 - \Phi(\sqrt{n}(s-\theta))), & \theta = 0 \\
bP(X < r) + cP(r < X < s) - \Phi(\sqrt{n}(r-\theta)), & \theta > 0 
\end{cases}
\]

\[
= \begin{cases} 
c\Phi(\sqrt{n}(r-\theta)) + b\Phi(\sqrt{n}(s-\theta)), & \theta < 0 \\
\Phi(\sqrt{n}(r-\theta)) + b\Phi(\sqrt{n}(s-\theta)), & \theta = 0 \\
b\Phi(\sqrt{n}(r-\theta)) + c\Phi(\sqrt{n}(s-\theta)), & \theta > 0 
\end{cases}
\]

where \( \Phi = 1 - \Phi \) and \( \Phi \) is the normal cdf.
From the plot, we can see that when $\theta > 0$, the procedure with $r=-s=-1$ has smaller risk than the procedure with $r=-s/2=-1$.

© B&D 1.3.9.

$X \sim \text{Bin}(n, \theta_0)$,

$E(\hat{p}) = E\left(\frac{X}{n}\right) = \theta_0, \quad \text{Var}(\hat{p}) = \text{Var}\left(\frac{X}{n}\right) = \frac{\theta_0(1-\theta_0)}{n},$

$E(\hat{\theta}) = (0.2)(0.10) + (0.8)E(\hat{p}) = 0.02 + 0.8\theta_0$ and $\text{Var}(\hat{\theta}) = (0.8)^2\text{Var}(\hat{p}) = 0.64 \frac{\theta_0(1-\theta_0)}{n}$

$\text{MSE}(\hat{p}) = \frac{\theta_0(1-\theta_0)}{n}$, $\text{MSE}(\hat{\theta}) = (0.2\theta_0 - 0.02)^2 + 0.64 \frac{\theta_0(1-\theta_0)}{b = n}$

$\frac{\text{MSE}(\hat{\theta})}{\text{MSE}(\hat{p})} < 1 \Rightarrow (0.2\theta_0 - 0.02)^2 + 0.64 \frac{\theta_0(1-\theta_0)}{n} < \frac{\theta_0(1-\theta_0)}{n}$

$\Rightarrow (n+9)\theta_0^2 - (0.2n+9)\theta_0 + 0.01n < 0$

$0.0187 < \theta_0 < 0.3931$ for $n = 25,$

$0.0407 < \theta_0 < 0.2253$ for $n = 100.$

© B&D 1.3.10.

From part (a) in problem 1.3.3,

$E_0(L(\theta, \delta_{r,s})(x)) = \begin{cases} 
\tilde{\Phi}(\sqrt{n}(r-\theta)) + b\tilde{\Phi}(\sqrt{n}(s-\theta)) & , \theta < 0 \\
b\Phi(\sqrt{n}s) + b\Phi(\sqrt{n}r) & , \theta = 0 \\
b\Phi(\sqrt{n}(r-\theta)) + c\Phi(\sqrt{n}(s-\theta)) & , \theta > 0
\end{cases}$

$= \begin{cases} 
\tilde{\Phi}(r-\theta) + \tilde{\Phi}(s-\theta) & , \theta < 0 \\
\Phi(s) + \Phi(r) & , \theta = 0 \\
\Phi((r-\theta)) + \Phi((s-\theta)) & , \theta > 0
\end{cases}$

where $\tilde{\Phi} = 1 - \Phi$ and $\Phi$ is the standard normal distribution function and the prior distribution is as follows.

$\pi(0) = \pi(-1/2) = \pi(1/2) = 1/3.$

(a)

$E_0(L(\theta, \delta_{r=-1,s=1})(x)) = \begin{cases} 
\tilde{\Phi}((-1-\theta)) + \tilde{\Phi}((1-\theta)) & , \theta < 0 \\
\tilde{\Phi}(1) + \tilde{\Phi}(-1) & , \theta = 0 \\
\Phi((-1-\theta)) + \Phi((1-\theta)) & , \theta > 0
\end{cases}$

The the Bayes risk of $\delta_{-1,1}(x)$ is
\[ R(G, \delta_{-1,1}(x)) = \sum_\theta E_\theta(L(\theta, \delta_{-1,1}(x)) \pi(\theta) \]
\[ = \frac{\{-\phi(-\frac{1}{2}) + \phi(\frac{3}{2}) + (\phi(1) + \phi(-1)) + \phi(-\frac{3}{2}) + \phi(\frac{1}{2})\}}{3} \]
\[ = 0.61128 \]

(b)
\[ E_\theta(L(\theta, \delta_{r=-1,s=2}(x)) = \begin{cases} 
\Phi((-1-\theta)) + \Phi((2-\theta)) & \theta < 0 \\
\Phi(2) + \Phi(-1) & \theta = 0 \\
\Phi((-1-\theta)) + \Phi((2-\theta)) & \theta > 0 
\end{cases} \]

The Bayes risk of \( \delta_{-1,2}(x) \) is
\[ R(G, \delta_{-1,2}(x)) = \sum_\theta E_\theta(L(\theta, \delta_{-1,2}(x)) \pi(\theta) \]
\[ = \frac{\{-\phi(-\frac{1}{2}) + \phi(\frac{5}{2}) + (\phi(2) + \phi(-1)) + \phi(-\frac{3}{2}) + \phi(\frac{3}{2})\}}{3} \]
\[ = 0.626359. \]

From the Bayes point of view, \( \delta_{-1,1}(x) \) is a better decision rule.

© B&D 1.3.18.

(a) Risk function is \( R(\theta, \delta_k) = L(\theta, 1)P(\delta = 1) + L(\theta, 0)P(\delta = 0) \)
\[ = \begin{cases} 
sP_{\theta}(X \geq k) + rN\theta P_{\theta}(X < k) & \theta < \theta_0 \\
rN\theta P_{\theta}(X < k) & \theta > \theta_0 \end{cases} \]
where the loss function is \( L(\theta, 1) = s, \theta < \theta_0 \)
\( L(\theta, 1) = 0, \theta \geq \theta_0 \)
\( l(\theta, 0) = rN\theta \)

(b) For \( N = 10, s = r = 1, \theta_0 = .1 \) and \( k = 3 \)
\[ R(\theta, \delta_k) = \begin{cases} 
P_{\theta}(X \geq 3) + 10\theta P_{\theta}(X < 3), \theta < 0.1 \\
10\theta P_{\theta}(X < 3), \theta \geq 0.1 \end{cases} \]

(c)
For $N=10$, $s=r=1$, $\theta_0 = .1$ and $k=2$

$$R(\theta, \delta_2) = \begin{cases} 
  P_\theta(X \geq 2) + 10\theta P_\theta(X < 2), & \theta < 0.1 \\
  10\theta P_\theta(X < 2), & \theta \geq 0.1 
\end{cases}$$

, where $P(X < k) = \sum_{i=0}^{k-1} \binom{N\theta}{i} \binom{N-N\theta}{n-i} \binom{N}{n}$ and $\max\{n-N(1-\theta),0\} \leq i \leq \min\{N\theta,n\}$

From the plot, $R(\theta, \delta_2)$ is smaller than $R(\theta, \delta_3)$ over all possible values of $\theta$, which means that $\delta_2$ is a better decision rule.