CS7DS3 Assignment 1

08-03-2025

We define θ as the probability of an individual component failure X. So it can be said:

$$P(X \text{ fails}) = \theta$$

For a system with two components $X1$ and $X2$ in parallel, the system $S1$ failure probability θ_1 is given by:

$$\theta_1 = P(S1 \text{ fails})$$

Since the components are independent (meaning one failing doesn't affect the other), we can use the multiplication rule for probability:

$$\theta_1 = P(X1 \text{ fails}) * P(X2 \text{ fails})$$

As both $X1$ and $X2$ have common probability θ, we can say:

$$P(X \text{ fails}) = P(X1 \text{ fails}) = P(X2 \text{ fails}) = \theta$$

Hence, $\theta 1$ can be represented as the following:

$$\theta_1 = \theta^2$$

Answer 2

```
# we define the component failure probability and number of systems
n <- 1000
theta <- 0.08

# calculate system failure probability theta1
theta1 <- theta^2
theta1
```

Part 1

```
## [1] 0.0064
```

```
# calculate Expectation and Variance
E_A <- n * theta1
E_A
```

```
## [1] 6.4
```

```
Var_A <- n * theta1 * (1 - theta1)
Var_A
```

```
## [1] 6.35904
```

We can actually conclude from these values that parallel configurations significantly improve reliability: the system failure rate (0.64%) is much lower than individual component failure rates (8%), resulting in only about 6-7 expected failures among 1000 systems with minimal variance of 6.35904 as well.

Part 2 We need to find the probability $P(5 \leq A \leq 10)$, where A is the total number of observed failures out of 1,000 tested $S1$ systems.

We know from Part 1 that the probability of a single $S1$ system failing is $\theta_1 = 0.0064$. When we test 1,000 independent systems, the total number of failures follows a binomial distribution:

$$A = Binomial(n = 1000, p = 0.0064)$$

To calculate $P(5 \leq A \leq 10)$, we can use the cumulative binomial probability function:

$$P(5 \leq A \leq 10) = P(A \leq 10) - P(A \leq 4)$$

We shall calculate this in R:

```
n <- 1000     # number of systems tested
theta <- 0.08
theta1 <- theta^2     # probability of a single system failing

# calculate P(5 <= A <= 10)
prob_5_to_10 <- pbinom(10, size=n, prob=theta1) - pbinom(4, size=n, prob=theta1)
prob_5_to_10
```

```
## [1] 0.7050234
```

The probability that between 5 and 10 systems (inclusive) will fail out of the 1,000 systems is approximately 0.705 or about 70.5%. This makes sense because we calculated earlier that the expected number of failures is 6.4, and the range from 5 to 10 captures values both slightly below and above this expected value.

Part 3 For this part, we need to find the minimum number of failures $k*$ such that $P(A \leq k*) \geq 0.95$. In other words, we're looking for the smallest value $k*$ where there's at least a 95% chance that the number of failures will be $k*$ or fewer.

1. We already established that A (the number of system failures) follows a binomial distribution with $n = 1000$ and $p = \theta_1 = 0.0064$.

2. We need to find the smallest value $k*$ where: $P(A \leq k*) \geq 0.95$

3. The cumulative distribution function (CDF) of the binomial distribution can provide $P(A \leq k*)$ for any value of $k*$.

4. In R, we can use the quantile `qbinom()` function of the binomial distribution, which performs the inverse of the CDF.

```
k_star <- qbinom(0.95, size=1000, prob=0.0064)
k_star
```

[1] 11

The result is $k* = 11$.

Answer 3

Part 1 For system $S2$, which has components $X1$ and $X2$ in series, we need to find $\theta_2 = P(S2 \text{ fails})$.

In a series configuration, the system fails if ANY component fails. This is the opposite of the parallel configuration given in Question 1.

To find the probability of system failure, let's first consider that the system $S2$ works ONLY if BOTH $X1$ AND $X2$ work. Therefore:

$$P(S2 \text{ works}) = P(X1 \text{ works AND } X2 \text{ works})$$

Since the components are independent:

$$P(S2 \text{ works}) = P(X1 \text{ works}) \times P(X2 \text{ works})$$
$$P(S2 \text{ works}) = (1 - \theta) \times (1 - \theta) = (1 - \theta)^2$$

Now, to find the probability of system failure:

$$P(S2 \text{ fails}) = 1 - P(S2 \text{ works})$$

$$P(S2 \text{ fails}) = 1 - (1 - \theta)^2$$
$$P(S2 \text{ fails}) = 1 - (1 - 2\theta + \theta^2)$$
$$P(S2 \text{ fails}) = 2\theta - \theta^2$$

Therefore:
$$\theta_2 = P(S2 \text{ fails}) = 2\theta - \theta^2$$

Part 2 For system $S3$, which has a mixed configuration as shown in the diagram, we need to find $\theta3 = P(S3 \text{ fails})$ in terms of θ.

From looking at the structure of $S3$, let's say the system contains consists of four components $(X1, X2, X3, X4)$ arranged in a specific configuration i.e. $X1$ and $X2$ in series and $X3$ and $X4$ in series. Both these pairs are then connected in parallel.

We shall calculate the failure probability for each series pair in terms of θ.

For a series pair, the pair works only if both components work (this was already calculated in previous question):

$$P(Pair1 \text{ works}) = (1 - \theta)^2$$
$$P(Pair1 \text{ fails}) = 1 - (1 - \theta)^2 = 2\theta - \theta^2$$

Similarly,
$$P(Pair2 \text{ fails}) = 2\theta - \theta^2$$

Kaaviya Paranji Ramkumar | 24358906

Now the two series pairs are connected in parallel. So, for the parallel configuration, the system fails only if both pairs fail and we can express it in θ in the following way:

$$P(S3 \text{ fails}) = P(Pair1 \text{ fails}) \times P(Pair2 \text{ fails})$$

$$P(S3 \text{ fails}) = (2\theta - \theta^2) \times (2\theta - \theta^2)$$

Therefore:

$$\theta_3 = P(S3 \text{ fails}) = (2\theta - \theta^2)^2$$

Answer 4

Part 1 The proposed prior distribution is Beta(3, 30). We can derive few key properties in the following way:

1. The mean of this distribution is $\frac{\alpha}{\alpha+\beta} = \frac{3}{3+30} = \frac{3}{33} \approx 0.091$ or about 9.1%.

2. The mode is $\frac{\alpha-1}{\alpha+\beta-2} = \frac{2}{31} \approx 0.065$ or about 6.5%.

3. The variance is $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{3\times30}{33^2\times34} \approx 0.0025$.

We can also visualize and analyze this distribution in R:

```r
# perform Beta(3, 30) prior analysis
x <- seq(0, 0.3, by=0.001)
prior_density <- dbeta(x, 3, 30)

# plot the prior
plot(x, prior_density, type="l", lwd=2,
     xlab=expression(theta), ylab="Density",
     main="Beta(3, 30) Prior Distribution")
abline(v=3/33, col="red", lty=2)   # the calculated mean
abline(v=2/31, col="blue", lty=2)  # the calculated mode
```

Kaaviya Paranji Ramkumar | 24358906

Beta(3, 30) Prior Distribution
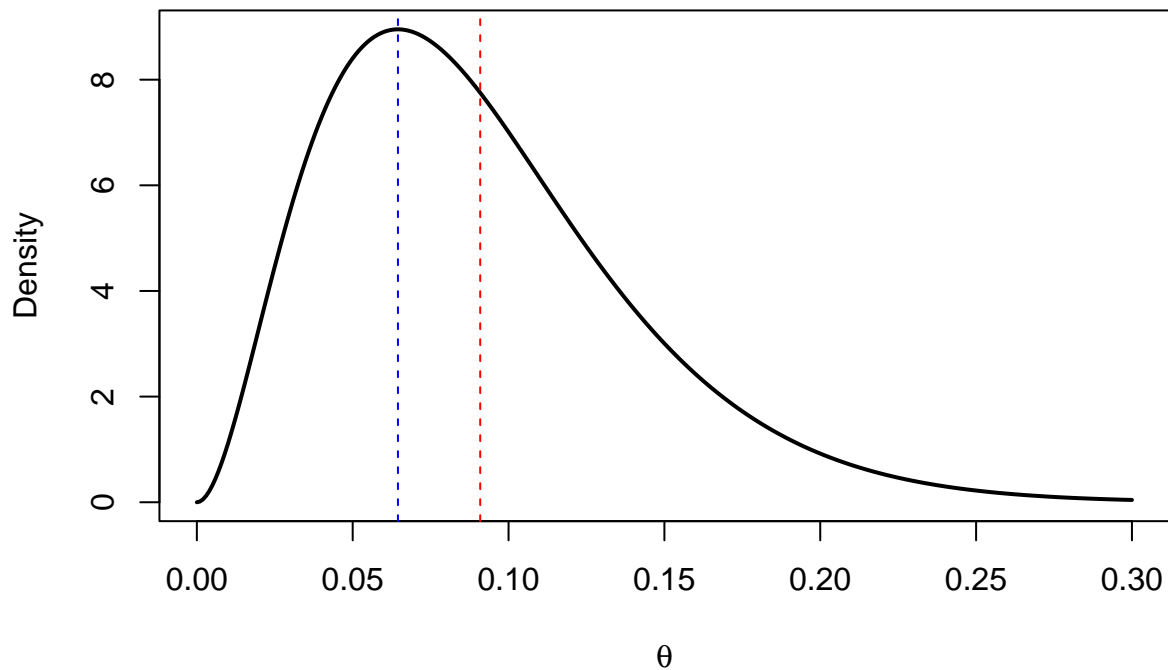

```r
# probability that theta is between 0.05 and 0.10
prob_5_10 <- pbeta(0.10, 3, 30) - pbeta(0.05, 3, 30)
prob_5_10
```

```
## [1] 0.419431
```

```r
# probability that theta exceeds 0.20
prob_exceed_20 <- 1 - pbeta(0.20, 3, 30)
prob_exceed_20
```

```
## [1] 0.03169127
```

The Beta(3, 30) prior matches the expert's description generally well. Its mean of 9.1% falls within the expert's expected range of 5% to 10%. The probability that θ exceeds 0.20 is small (about 3.2%), which aligns with the expert's assertion that values above 20% are "highly unusual". The distribution places a significant probability mass (42%) in the 5-10% range, which is consistent with the expert's belief that the component failure probability is likely in this interval.

Part 2 To construct the posterior distribution, we combine the Beta(3, 30) prior with our binomial data of 3 failures in 25 trials. For the Beta-Binomial model, the posterior follows a Beta distribution with parameters (below formula is similarly described in Lecture 6 pg. 15/31):

$$\alpha_{post} = \alpha_{prior} + \text{number of failures} = 3 + 3 = 6$$

$$\beta_{post} = \beta_{prior} + \text{number of successes} = 30 + (25 - 3) = 30 + 22 = 52$$

Therefore, the posterior distribution can be written as:

$$p(\theta|x) \sim \text{Beta}(6, 52)$$

We can derive the key features of this posterior distribution similarly to the way done in prior distribution:

1. Mean: $\frac{\alpha_{post}}{\alpha_{post} + \beta_{post}} = \frac{6}{6+52} = \frac{6}{58} \approx 0.103$ or about 10.3%

2. Mode: $\frac{\alpha_{post} - 1}{\alpha_{post} + \beta_{post} - 2} = \frac{5}{56} \approx 0.089$ or about 8.9%

3. Variance: $\frac{\alpha_{post}\beta_{post}}{(\alpha_{post} + \beta_{post})^2(\alpha_{post} + \beta_{post} + 1)} = \frac{6 \times 52}{58^2 \times 59} \approx 0.0016$

Now, we can visualize the posterior distribution:

```r
# perform posterior Beta(6, 52) analysis
x <- seq(0, 0.3, by=0.001)
posterior_density <- dbeta(x, 6, 52)

# plot the posterior
plot(x, posterior_density, type="l", lwd=2,
     xlab=expression(theta), ylab="Density",
     main="Beta(6, 52) Posterior Distribution")
abline(v=6/58, col="red", lty=2)   # Mean
abline(v=5/56, col="blue", lty=2)   # Mode
```

Beta(6, 52) Posterior Distribution

Kaaviya Paranji Ramkumar | 24358906
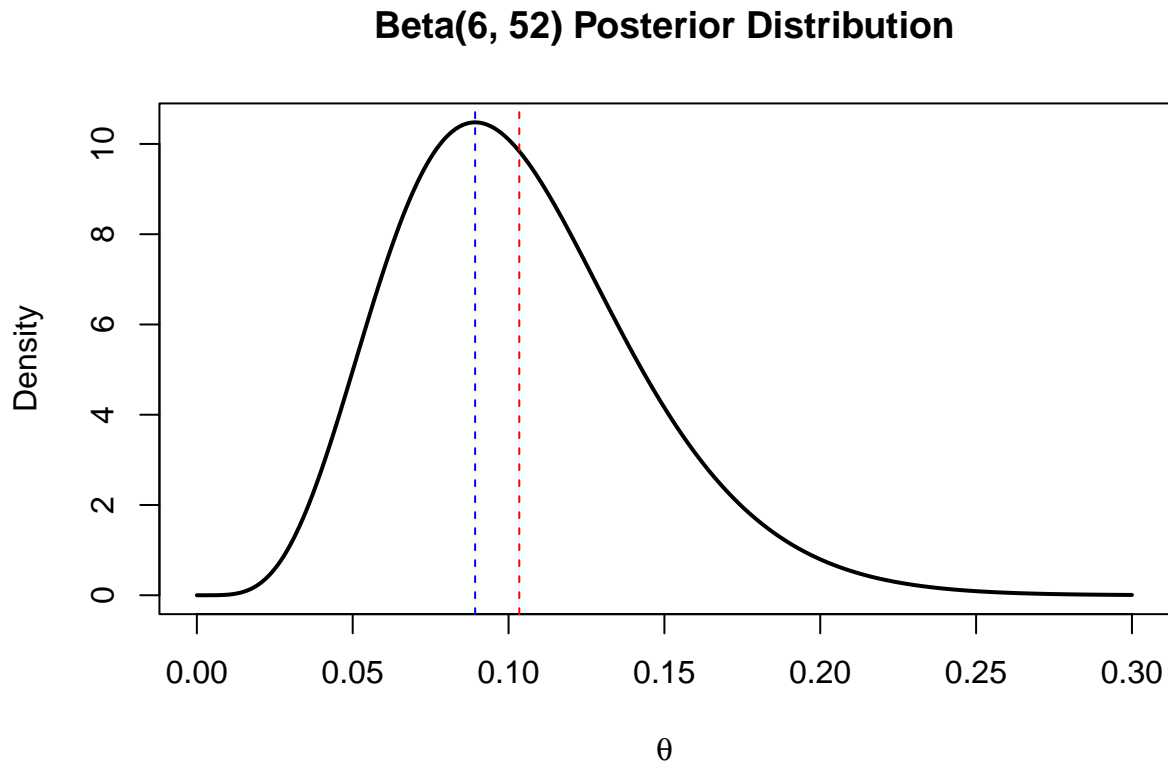
The posterior distribution has a mean of 10.3%, suggesting that based on our prior beliefs and observed data, the most likely component failure rate is around 10.3%.

Part 3   To compare the posterior distribution $p(\theta|x)$ with the prior distribution $p(\theta)$, we can plot them together in R:

```r
# Compare prior and posterior
x <- seq(0, 0.3, by=0.001)
prior_density <- dbeta(x, 3, 30)
posterior_density <- dbeta(x, 6, 52)

# Plotting both distributions
plot(x, posterior_density, type="l", lwd=2, col="red",
     xlab=expression(theta), ylab="Density",
     main="Prior vs Posterior Distribution")
lines(x, prior_density, lwd=2, col="blue")
legend("topright", legend=c("Prior Beta(3, 30)", "Posterior Beta(6, 52)"),
       col=c("blue", "red"), lty=1, lwd=2)
```

Prior vs Posterior Distribution
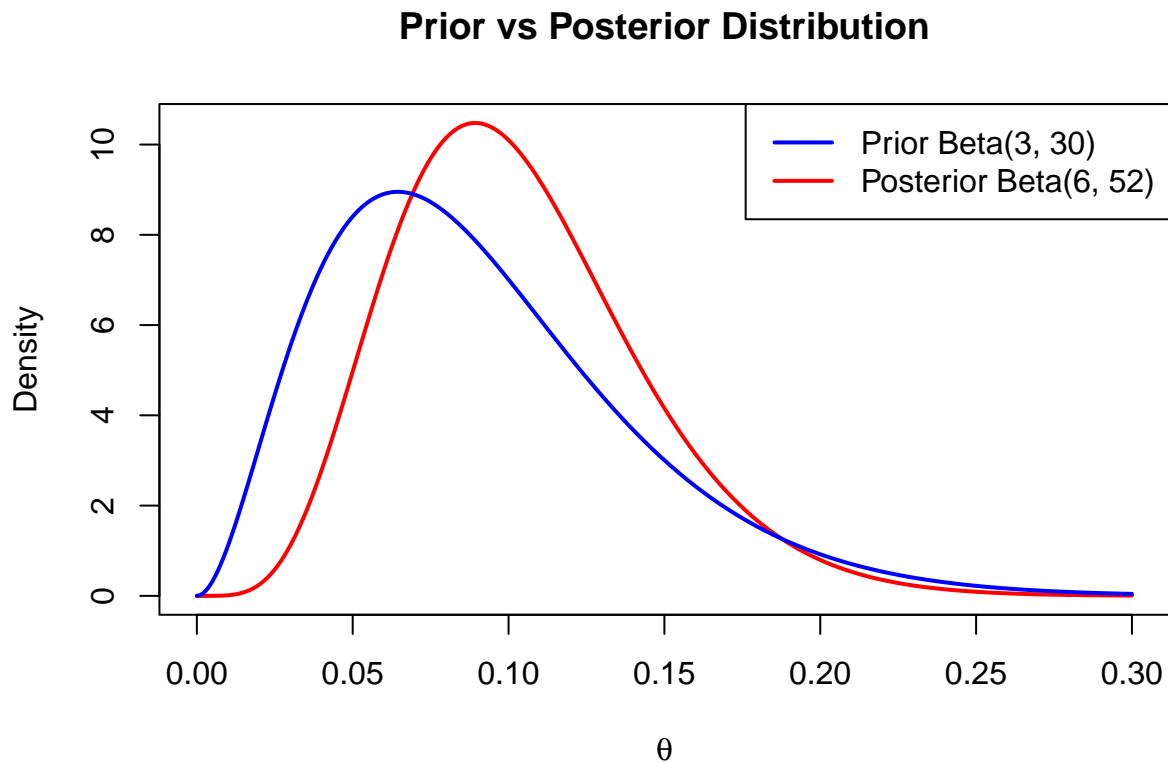
We know that the prior mean is 9.1% while the posterior mean is 10.3% which means that the data has slightly increased our estimate of the failure probability. The variance of the posterior (0.0016) is smaller than the prior variance (0.0025), which could mean certainty has increased after observing the data. From the figure, the posterior distribution looks more narrower and concentrated. We can measure this by comparing the "effective sample size" of the prior ($\alpha + \beta = 33$) to the actual sample size (n = 25). Since these are of similar magnitude, the prior has a substantial influence on the posterior. The prior contributes information

Kaaviya Paranji Ramkumar | 24358906

equivalent to having observed 3 failures in 33 trials, which is comparable to our actual data of 3 failures in 25 trials. Thus, the prior is quite informative to the data.

Part 4   Now, we can apply the posterior distribution for θ to derive posterior distributions for θ_1, θ_2, and θ_3 using Monte Carlo methods:

```r
# perform Monte Carlo sampling from posterior
set.seed(123)  # set seed for reproducibility
n_samples <- 10000
theta_samples <- rbeta(n_samples, 6, 52)

# calculate the 3 system failure probabilities
theta1_samples <- theta_samples^2                      # already derived this for parallel system
theta2_samples <- 2*theta_samples - theta_samples^2    # already derived this for series system
theta3_samples <- (2*theta_samples - theta_samples^2)^2  # already derived this for mixed system

# create density plots
par(mfrow=c(2,2))
hist(theta_samples, main="Component Failure Rate", xlab=expression(theta), probability=TRUE)
hist(theta1_samples, main="Parallel System Failure Rate", xlab=expression(theta[1]), probability=TRUE)
hist(theta2_samples, main="Series System Failure Rate", xlab=expression(theta[2]), probability=TRUE)
hist(theta3_samples, main="Mixed System Failure Rate", xlab=expression(theta[3]), probability=TRUE)
```


Kaaviya Paranji Ramkumar | 24358906
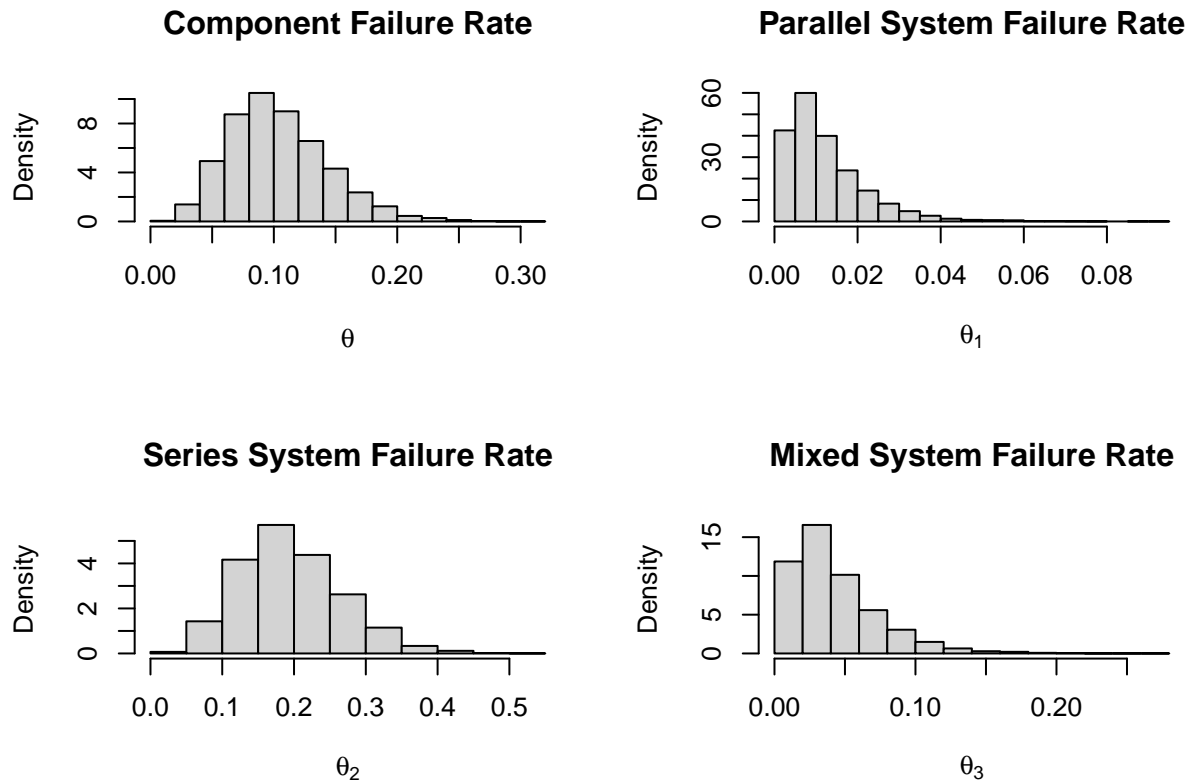
```
# get summary statistics
theta_mean <- mean(theta_samples)
theta1_mean <- mean(theta1_samples)
theta2_mean <- mean(theta2_samples)
theta3_mean <- mean(theta3_samples)

c(Component=theta_mean, Parallel=theta1_mean, Series=theta2_mean, Mixed=theta3_mean)
```

```
##   Component    Parallel      Series       Mixed
## 0.10355393  0.01231831  0.19478956  0.04292777
```

From the 2nd sub figure for θ_1 (i.e. parallel system), the distribution is highly skewed toward zero with a much lower mean (1.23%) than the component failure rate (10.3%). This reflects the high reliability of parallel systems, where both components must fail for the system to fail. For θ_2 (series system), the distribution is shifted to the right of the component failure rate distribution, with a higher mean (19.48%). So, we can conclude that series systems is less reliable, where any component failure can cause entire system failure. Finally for mixed system with θ_2, the distribution shows characteristics between the parallel and series systems where mean of the sample is 4.29%, but is more similar to the parallel system because of the redundancy provided by the parallel configuration of the series pairs.

Answer 5

Part 1 For this part, we need to calculate the expected price the company will pay for 1,000 type S3 systems and consider the following refund conditions:

- Condition 1: If failures < 50: Pay full price of €1,000
- Condition 2: If 50 ≤ failures < 100: Pay €(1,000 - number of failures)
- Condition 3: If failures ≥ 100: Pay €0 (full refund)

Next, we can calculate the expected price of each condition. The overall expected price of the systems will be nothing but the sum of the expected price of the 3 conditions.

$$E[\text{Price on Condition 1}] = €1,000 \times P(failures < 50)$$

$$E[\text{Price on Condition 2}] = €(1,000 - E[failures|50 \leq failures < 100]) \times P(50 \leq failures < 100)$$

$$E[\text{Price on Condition 3}] = €0 \times P(failures \geq 100)$$

Using our posterior distribution from Answer 4, we can compute these probabilities using Monte Carlo simulation in R. So far, we know: - θ $Beta(6, 52)$ for the component failure probability - $\theta_3 = (2\theta - \theta^2)^2$ for the S3 system failure probability

```
# perform Monte Carlo sampling from the posterior
set.seed(123)  # set seed for reproducibility
n_samples <- 100000
theta_samples <- rbeta(n_samples, 6, 52)

# calculate only S3 system failure probability for each sample
theta3_samples <- (2*theta_samples - theta_samples^2)^2

# For each theta3, let's simulate 1000 systems with binomial distribution
failures <- rbinom(n_samples, size=1000, prob=theta3_samples)
```

```
# calculate price according to conditions
price <- rep(1000, n_samples)
price[failures >= 50 & failures < 100] <- 1000 - failures[failures >= 50 & failures < 100]
price[failures >= 100] <- 0

# get expected price from mean
expected_price <- mean(price)
expected_price
```

[1] 922.0821

```
# calculate probabilities of different scenarios
prob_less_than_50 <- mean(failures < 50)
prob_50_to_99 <- mean(failures >= 50 & failures < 100)
prob_100_or_more <- mean(failures >= 100)

c(Expected_Price = expected_price,
  P_less_than_50 = prob_less_than_50,
  P_50_to_99 = prob_50_to_99,
  P_100_or_more = prob_100_or_more)
```

Expected_Price P_less_than_50   P_50_to_99 P_100_or_more
922.08212   0.67665   0.26338   0.05997

Based on our Monte Carlo simulation above using the posterior distribution for component reliability, the expected price for 1,000 S3 systems is €922.08, rather than the given €1,000. This can account for all possible refund scenarios. With the output from the simulation, we can infer:

- 67.7% probability of fewer than 50 failures (full payment of €1,000)
- 26.3% probability of 50-99 failures (partial refund)
- 6.0% probability of 100+ failures (full refund)

These results indicate that while the company will most likely pay the full price, there's a significant chance (about 32.3%) they'll receive some form of refund, reducing the expected cost by approximately €77.92 per 1,000 systems.

Part 2 To assess the sensitivity of the expected price estimate to the choice of prior, we should repeat the calculation using a non-informative Beta(1,1) prior instead of the expert prior.

With the original data (3 failures in 25 tests) and a Beta(1,1) prior, the posterior would be:

$$\theta = Beta(1 + 3, 1 + 22) = Beta(4, 23)$$

Now, we can compare the expected prices using both priors:

```
# perform Monte Carlo sampling from the non-informative prior posterior
set.seed(123)
theta_samples_noninf <- rbeta(n_samples, 4, 23)

# calculate only S3 system failure probability
```

```
theta3_samples_noninf <- (2*theta_samples_noninf - theta_samples_noninf^2)^2

# simulate failures
failures_noninf <- rbinom(n_samples, size=1000, prob=theta3_samples_noninf)

# calculate price according to conditions specified
price_noninf <- rep(1000, n_samples)
price_noninf[failures_noninf >= 50 & failures_noninf < 100] <- 1000 - failures_noninf[failures_noninf >=
price_noninf[failures_noninf >= 100] <- 0

# get expected price with non-informative prior
expected_price_noninf <- mean(price_noninf)

# compare both expected prices
c(Expert_Prior = expected_price,
  Non_Informative_Prior = expected_price_noninf,
  Difference = expected_price - expected_price_noninf)
```

```
##       Expert_Prior Non_Informative_Prior          Difference
##          922.0821             655.0043            267.0779
```

Using the non-informative Beta(1,1) prior leads to an expected price of approximately €655, which is about €267 different from our original estimate.

This significant difference (approximately 29% of the full price) shows that the expected price estimate is highly sensitive to the choice of prior distribution. The non-informative prior results in a much lower expected price because because the Beta(1,1) prior assigns equal probability to all possible values of θ between 0 and 1, whereas the expert prior concentrates probability mass around the 5-10% range. The Beta(1,1) prior results in a posterior distribution with a higher mean for θ ($4/27 = 0.148$ compared to $6/58 = 0.103$), which leads to a higher system failure probability.

This sensitivity analysis suggests that expert knowledge does meaningfully influence our results, but the data still plays a substantial role in shaping the posterior distribution and thus our expected price. Using the expert's beliefs about component reliability does produce a more optimistic (and likely more realistic) estimate, while the non-informative prior, because it assigns equal weight to all possible failure rates, leads to a more pessimistic pricing expectation.

Answer 6

Part 1 In this part, we need to express the likelihood for data y in terms of both θ_1 and θ.

We have a study where $S1$ systems (the parallel configuration) are tested directly. Out of n systems tested, k systems failed. Each system has a failure probability of θ_1, and we're observing $y = (y_1, ..., y_n)$, where $y_i = 1$ if the system failed and 0 otherwise.

a) In terms of θ_1: As similarly shown in Lecture 6 (Beta-Binomial Model), the likelihood for binomial data is:

$$L(y|\theta_1) = \binom{n}{k} \theta_1^k (1 - \theta_1)^{n-k}$$

This represents the probability of observing k failures in n trials, where each trial has probability θ_1 of failure.

b) In terms of θ: From Answer 1, we know that $\theta_1 = \theta^2$ for the parallel system. Substituting this into the likelihood:

$$L(y|\theta) = \binom{n}{k}(\theta^2)^k(1-\theta^2)^{n-k}$$

$$L(y|\theta) = \binom{n}{k}\theta^{2k}(1-\theta^2)^{n-k}$$

Part 2 To find the maximum likelihood estimate (MLE) of θ, following the approach in Lecture 4 (Frequentist Inference), we:

1. Take the log of the likelihood:

$$\ell(\theta) = \log\binom{n}{k} + 2k\log\theta + (n-k)\log(1-\theta^2)$$

2. Differentiate with respect to θ and set equal to zero:

$$\frac{d\ell}{d\theta} = \frac{2k}{\theta} - (n-k)\frac{2\theta}{1-\theta^2} = 0$$

3. Solve for θ:

$$\frac{2k}{\theta} = (n-k)\frac{2\theta}{1-\theta^2}$$

Simplifying:

$$k(1-\theta^2) = (n-k)\theta^2$$

$$k = n\theta^2$$

$$\theta^2 = \frac{k}{n}$$

Therefore:

$$\hat{\theta} = \sqrt{\frac{k}{n}}$$

This is the maximum likelihood estimate for θ.

Part 3 Based on Lecture 6 (Bayesian Inference - Beta-Binomial Model) and Lecture 7 (Bayesian Inference - Normal-Normal Model), a Bayesian approach would be more challenging in this case simply because we don't have a standard conjugate prior.

As discussed in Lecture 6, the Beta distribution is conjugate for the Binomial likelihood. However, in this case, our likelihood has the form:

$$L(y|\theta) \propto \theta^{2k}(1-\theta^2)^{n-k}$$

This doesn't match the standard Beta-Binomial form. Similar to what was mentioned in Lecture 7 about the case where both μ and τ are unknown in the normal model, we would need to use computational methods to handle this non-conjugate case.