Homework 3

Chavez Cheong

1/18/2022

1 Lab Component

```r
### Set seed
set.seed(123)

### Data
x_sum <- 1
n <- 30

### Prior Parameters
a <- 0.5
b <- 1
```

```r
### Loss Function
# Inputs: theta, c
# Output: loss

loss_function <- function(theta, c){
  if (c > theta) {
    loss <- abs(theta - c)
  }
  else {
    loss <- 10 * abs(theta - c)
  }
  return(loss)
}
```

```r
### Posterior Risk Function
# Inputs: c, a_prior, b_prior, x_sum, n, s(number of rolls)

posterior_risk <- function(c, prior_a, prior_b, x_sum, n, s = 30000) {
  # random draws from posterior distribution
  post_a <- prior_a + x_sum
  post_b <- prior_b + n - x_sum
  theta <- rbeta(s, post_a, post_b)
  loss <- apply(as.matrix(theta),1,loss_function,c)
  risk <- mean(loss)
  return(risk)
}
```

Task 3

```r
sum_xs <- seq(0,30)
min_c <- matrix(NA,3, length(sum_xs))

find_optimal_C <- function(x_sum, prior_a, prior_b, n, s = 1000) {
  c = seq(0, 1, by=0.01)
  risk <- apply(as.matrix(c), 1, posterior_risk, prior_a, prior_b, x_sum, n, s)
  c[which.min(risk)]
}

min_c[1,] = apply(as.matrix(sum_xs), 1, find_optimal_C, a, b, n)
min_c[2,] = (sum_xs/n)
min_c[3,] = 0.1
```

```r
### Plot the curves
plot(sum_xs, min_c[1,], col = "blue", type = "o", pch = 16, ylab = "resources allocated", xlab = "observ
par(new = T)
plot(sum_xs, min_c[2,], type = 'o', col='green',
pch = 16, ylab = "", xlab = '', ylim = c(0,1))
par(new = T)
plot(sum_xs, min_c[3,], type = 'o',col = 'red',
pch = 16, ylab = "", xlab = '', ylim = c(0,1))

legend("topleft", lty = c(1,1,1), pch = c(16,16,16),
col = c("blue", "green", "red"),
legend = c("Bayes", "Sample mean", "constant"))
```
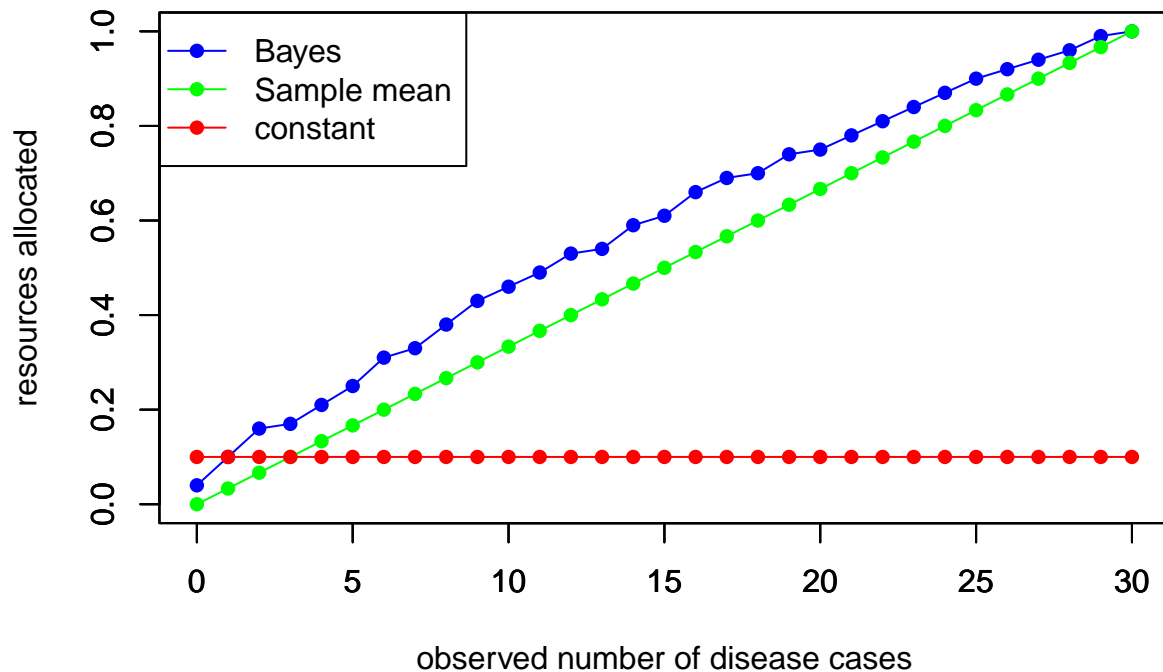
The Bayesian procedure always picks an estimator that is slightly higher than the sample mean $\bar{x}$. This is because the Bayes estimator is a weighted average between the prior mean and the sample mean, and the prior mean in this case is always slightly higher than the sample mean due to our choice of prior.

Task 4

```r
thetas <- seq(0, 1, by = 0.1)

frequentist_risk <- function(theta) {
  sum_xs <- rbinom(100,30, theta)
  bayes_optimal = apply(as.matrix(sum_xs), 1, find_optimal_C, a, b, n, s = 100)
  mean_c = sum_xs / 30
  bayes_loss <- apply(as.matrix(bayes_optimal), 1, loss_function, theta = theta)
  sample_loss <- apply(as.matrix(mean_c), 1, loss_function, theta)
  bayes_risk <- mean(bayes_loss)
  sample_risk <- mean(sample_loss)
  constant_risk <- loss_function(theta, 0.1)
  return(c(bayes_risk, sample_risk, constant_risk))
}

R <- apply(as.matrix(thetas), 1, frequentist_risk)

plot(thetas, R[1,], col='blue', type = "l",
ylab = "frequentist risk", xlab = expression(theta),ylim = c(0,1))
par(new = T)
plot(thetas, R[2,], type = 'l', col='green',
```
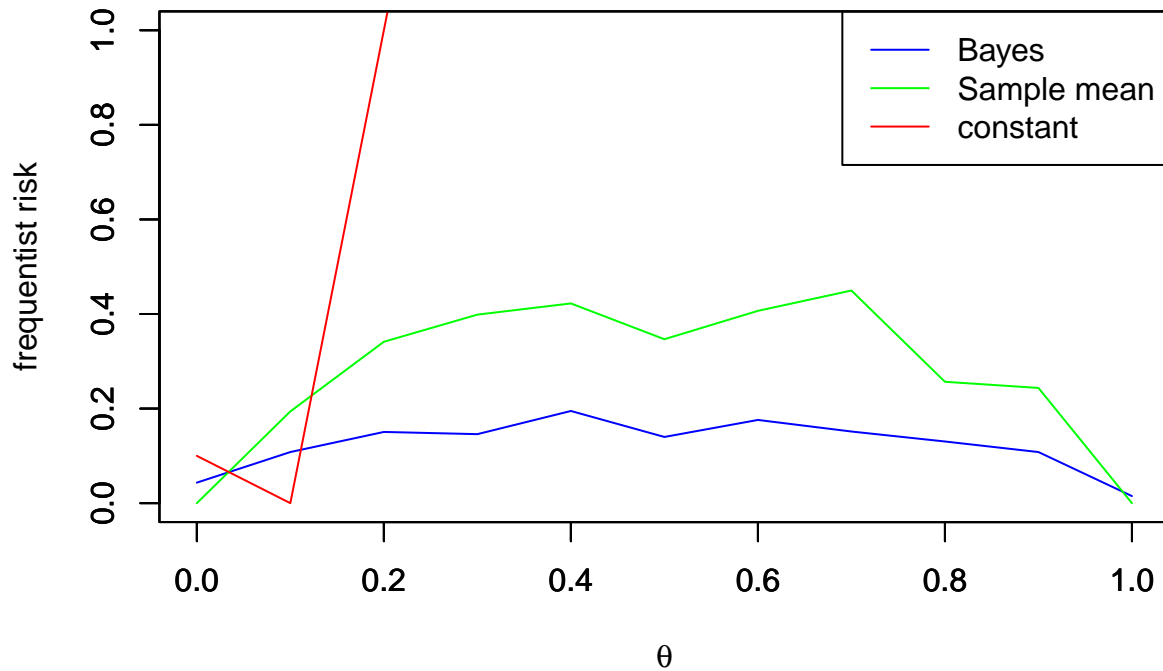
```
ylab = "", xlab = '', ylim = c(0,1))
par(new = T)
plot(thetas, R[3,], type = 'l',col = 'red',
ylab = "", xlab = '', ylim = c(0,1))
legend("topright", lty = c(1,1,1), col = c("blue", "green", "red"),
legend = c("Bayes", "Sample mean", "constant"))
```


The first observation is that none of the estimators are non-admissible, as each estimator has ranges of θ where the frequentist risk is lower than other estimators.

The second observation is that the constant estimator has a very low risk for θ close to its value, but the risk increases rapidly before and after.

The third observation is that the Bayesian estimator generally has a lower risk than the sample mean except for values of θ close to 0 and close to 1.

Task 5

All estimators are admissible, as each estimator has ranges of θ where the frequentist risk is lower than other estimators.

For θ close to 0.1, the constant estimator will have the lowest frequentist risk out all estimators, but it increases quickly before and after 0.1.

For θ close to 0 or 1, the sample mean has a lower frequentist risk than the Bayesian estimator. For all other values of θ, the Bayesian estimator has a lower frequentist risk than the sample mean.

2 The Uniform-Pareto

Likelihood

$$p(X = x|\theta) = \frac{1}{\theta} I_{(0,\theta)}(x)$$

Posterior Distribution

$$p(\theta|X = x) \overset{\theta}{\propto} p(\theta)p(X = x|\theta)$$
$$= \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} I_{(\beta,\infty)}(\theta)\frac{1}{\theta} I_{(0,\theta)}(x)$$
$$\overset{\theta}{\propto} \frac{1}{\theta^{\alpha+2}} I_{(\max(\beta,x),\infty)}(\theta)$$

Hence, $\theta|X = x \sim \text{Pareto}(\alpha + 1, \max(x, \beta))$.

3 The Bayes Estimator or Bayes Rule

a

$$\rho(\delta(x), x) = \mathbb{E}(L(\theta, \delta(x))|X)$$
$$= \mathbb{E}(c(\theta - \delta(x))^2|X)$$
$$= \mathbb{E}(c(\theta^2 - 2\delta(x)\theta + (\delta(x))^2)|X)$$
$$= c \cdot \mathbb{E}(\theta^2|X) - 2c\delta(x) \cdot \mathbb{E}(\theta|X) + c \cdot (\delta(x))^2$$

Taking the partial with respective to $\delta(x)$

$$\frac{\partial \rho(\delta(x), x)}{\partial \delta(x)} = -2c \cdot \mathbb{E}(\theta|X) + 2c \cdot \delta(x)$$

We set the partial to 0 to obtain a stationary point.

$$\frac{\partial \rho(\delta(x), x)}{\partial \delta(x)} = -2c \cdot \mathbb{E}(\theta|X) + 2c \cdot \delta(x) = 0$$
$$\iff 2c \cdot \delta(x) = 2c \cdot \mathbb{E}(\theta|X)$$
$$\iff \delta(x) = \mathbb{E}(\theta|X)$$

Thus, the Bayesian estimator is $\delta(x) = \mathbb{E}(\theta|X)$.

To show that this is unique, we take the second partial with respective to $\delta(x)$

$$\frac{\partial^2 \rho(\delta(x), x)}{\partial \delta(x)^2} = 2c > 0 \because c > 0$$

Hence, since $\frac{\partial^2 \rho(\delta(x),x)}{\partial \delta(x)^2} > 0$, we know that $\rho(\delta(x), x)$ is a convex function with respect to $\delta(x)$, hence the stationary point we found is a global minimum, and $\delta(x) = \mathbb{E}(\theta|X)$ is a unique Bayesian estimator.

b

$$\rho(\pi, \delta(x)) = \mathbb{E}(L(\theta, \delta(x))|X)$$
$$= \mathbb{E}(w(\theta)(g(\theta) - \delta(x))^2|X)$$
$$= \mathbb{E}(w(\theta)((g(\theta))^2 - 2\delta(x)g(\theta) + (\delta(x))^2)|X)$$
$$= \mathbb{E}(w(\theta)(g(\theta))^2|X) - 2\delta(x) \cdot \mathbb{E}(w(\theta)g(\theta)|X) + (\delta(x))^2 \cdot \mathbb{E}(w(\theta)|X)$$

Taking the partial with respective to $\delta(x)$

$$\frac{\partial \rho(\pi, \delta(x))}{\partial \delta(x)} = -2 \cdot \mathbb{E}(w(\theta)g(\theta)|X) + 2\delta(x) \cdot \mathbb{E}(w(\theta)|X)$$

We set the partial to 0 to obtain a stationary point.

$$\frac{\partial \rho(\delta(x), x)}{\partial \delta(x)} = -2 \cdot \mathbb{E}(w(\theta)g(\theta)|X) + 2\delta(x) \cdot \mathbb{E}(w(\theta)|X) = 0$$
$$\iff \delta(x) = \frac{\mathbb{E}(w(\theta)g(\theta)|X)}{\mathbb{E}(w(\theta)|X)}$$

Thus, the Bayesian estimator is $\delta(x) = \frac{\mathbb{E}(w(\theta)g(\theta)|X)}{\mathbb{E}(w(\theta)|X)}$.

To show that this is unique, we take the second partial with respective to $\delta(x)$,

$$\frac{\partial^2 \rho(\delta(x), x)}{\partial \delta(x)^2} = 2 \cdot \mathbb{E}(w(\theta)|X) > 0 \because w(\theta) > 0, \mathbb{E}(\theta|X) > 0$$

Hence, since $\frac{\partial^2 \rho(\delta(x), x)}{\partial \delta(x)^2} > 0$, we know that $\rho(\delta(x), x)$ is a convex function with respect to $\delta(x)$, hence the stationary point we found is a global minimum, and $\delta(x) = \frac{\mathbb{E}(w(\theta)g(\theta)|X)}{\mathbb{E}(w(\theta)|X)}$ is a unique Bayesian estimator.