

Homework 3

Sara Shao

9/6/2021

```
rm(list=ls())
set.seed(123)
```

Task 3

```
loss_function = function(theta, c){
  if (c < theta){
    return(10*abs(theta - c))
  }
  else{
    return(1 = abs(theta - c))
  }
}

posterior_risk = function(c, a_prior, b_prior, sum_x, n, s = 30000){
  # random draws from beta distribution
  a_post = a_prior + sum_x
  b_post = b_prior + n - sum_x
  theta = rbeta(s, a_post, b_post)
  loss <- apply(as.matrix(theta), 1, loss_function, c)
  # average values from the loss function
  risk = mean(loss)
}

n=30
a=0.05
b=1

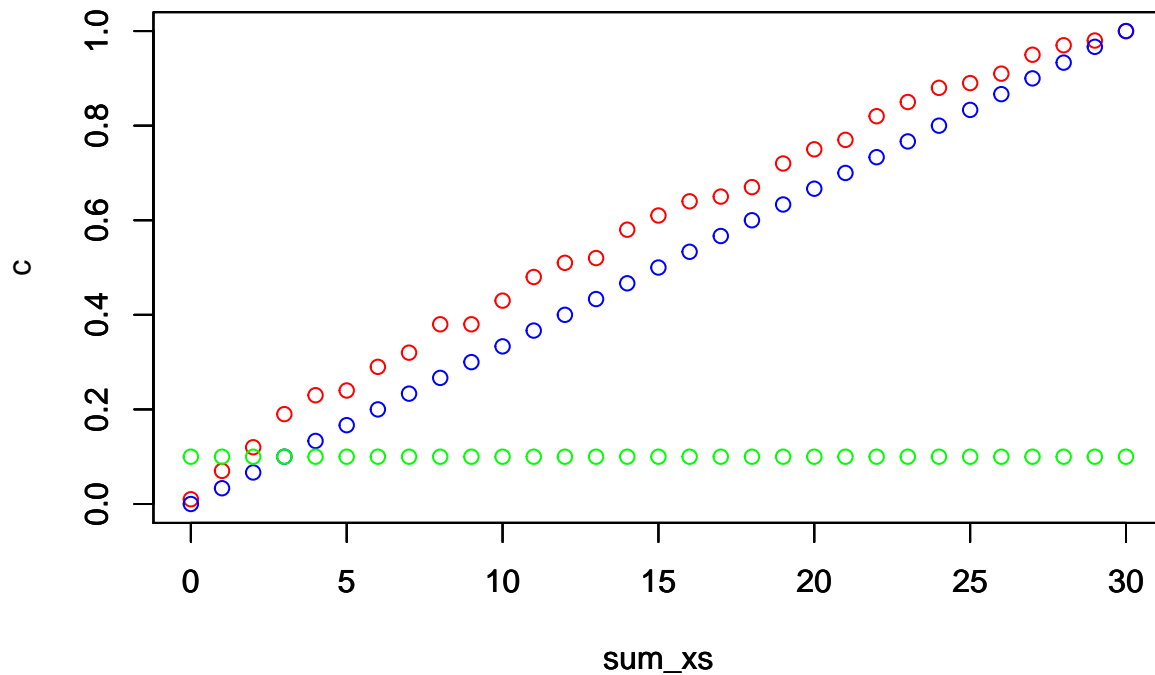
sum_xs = seq(0,n)

bayes_c <- function(sum_x,a_prior, b_prior, n, s=500) {
  cs = seq(0,1,by=0.01)
  post_risk = apply(as.matrix(cs), 1,posterior_risk,a_prior, b_prior, sum_x, n, s)
  cs[which.min(post_risk)]
}

c1 = apply(as.matrix(sum_xs),1,bayes_c,a,b,n)
c2 = sum_xs/n
c3 = rep(0.1, 31)

plot(sum_xs, c1, col = "red", ylim = c(0,1), ylab = "c")
par(new=TRUE)
plot(sum_xs, c2, col = "blue", ylim = c(0,1), ylab = "")
```

```
par(new=TRUE)
plot(sum_xs, c3, col = "green", ylim = c(0,1), ylab = "")
```



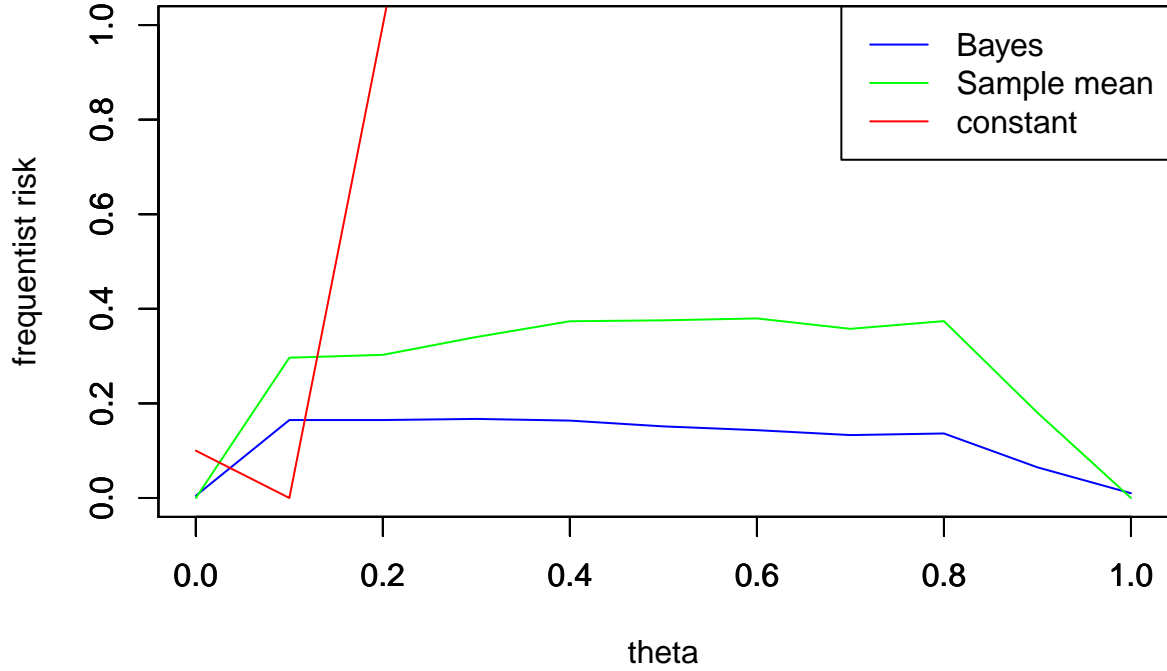
Task 4

```
thetas = seq(0, 1, by=0.1)
# frequentist risk for the 3 estimators given a theta
frequentist_risk = function(theta){
  sum_xs = rbinom(100, 30, theta)
  Bayes_optimal = apply(as.matrix(sum_xs), 1, bayes_c, a, b, n, s = 100)
  mean_c = sum_xs / 30
  loss1 = apply(as.matrix(Bayes_optimal), 1, loss_function, theta = theta)
  loss2 = apply(as.matrix(mean_c), 1, loss_function, theta = theta )
  risk1 = mean(loss1)
  risk2 = mean(loss2)
  risk3 = loss_function(theta, 0.1)
  return(c(risk1, risk2, risk3))
}
frequentist_risk(0.1)
```

```
## [1] 0.1461000 0.2406667 0.0000000
```

```
# given a sequence a theta, compute frequentist risk for each theta
R = apply(as.matrix(thetas), 1, frequentist_risk)
# plot
plot(thetas, R[1,], col='blue', type = "l",
ylab = "frequentist risk", xlab = '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"))
```



Task 5

None of the estimators are inadmissible. Bayes is admissible because it's lower than the constant and sample mean in most places. The constant is admissible for thetas close to the constant, and the sample mean is admissible at $\theta = 0$ or 1 .

2. $p(X = x|\theta) = \frac{1}{\theta}$ for $0 < x < \theta$, otherwise $p(X = x|\theta) = 0$

$p(\theta|X = x) \propto \frac{1}{\theta} \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} I_{(\beta,\infty)}(\theta)$ for $\theta > x$, otherwise $p(\theta|X = x) \propto 0 \times \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} I_{(\beta,\infty)}(\theta) = 0$

Therefore,

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

$$Pareto(\alpha + 1, \max(x, \beta))$$

3a. $p(\delta(x), x_{1:n}) = E(L(\theta, \delta(x))|x_{1:n}) = E(c(\theta - \delta(x))^2|x_{1:n}) = cE(\theta^2 - 2\theta\delta(x) + \delta^2(x)|x_{1:n})$

$$= cE(\theta^2|x_{1:n}) - 2c\delta(x)E(\theta|x_{1:n}) + c\delta^2(x), \hat{\theta} = \delta(x),$$

$$\frac{\partial(cE(\theta^2|x_{1:n}) - 2c\hat{\theta}E(\theta|x_{1:n}) + c\hat{\theta}^2)}{\partial\hat{\theta}} = -2cE(\theta|x_{1:n}) + 2c\hat{\theta} = 0$$

Solving $-2cE(\theta|x_{1:n}) + 2c\hat{\theta} = 0$ for $\hat{\theta}$ yields $\hat{\theta} = E(\theta|x_{1:n})$, which is the posterior mean.

3b. $p(\delta(x), x_{1:n}) = E(L(\theta, \delta(x))|x_{1:n}) = E(w(\theta)(g(\theta) - \delta(x))^2|x_{1:n})$

$$= E(w(\theta)g^2(\theta) - 2w(\theta)g(\theta)\delta(x) + w(\theta)\delta^2(x)|x_{1:n})$$

$$= E(w(\theta)g^2(\theta)|x_{1:n}) - 2\delta(x)E(w(\theta)g(\theta)|x_{1:n}) + \delta^2(x)E(w(\theta)|x_{1:n}), \hat{\theta} = \delta(x),$$

$$\frac{\partial(E(w(\theta)g^2(\theta)|x_{1:n})-2\hat{\theta}E(w(\theta)g(\theta)|x_{1:n})+\hat{\theta}^2E(w(\theta)|x_{1:n}))}{\partial\hat{\theta}} = -2E(w(\theta)g(\theta)|x_{1:n}) + 2\hat{\theta}E(w(\theta)|x_{1:n}) = 0$$

Solving the above equation for $\hat{\theta}$ yields:

$$\hat{\theta} = \frac{E(w(\theta)g(\theta)|x_{1:n})}{E(w(\theta)|x_{1:n})}$$

The Bayes rule(s) are unique because as long as $w(\theta)$ is positive, the $\hat{\theta}^2$ term ($\delta^2(x)$ term) has a positive coefficient, so the posterior risk is a convex function of $\hat{\theta}$.