# STA602 Homework 3

Zach Calhoun

Due 1/21/2022

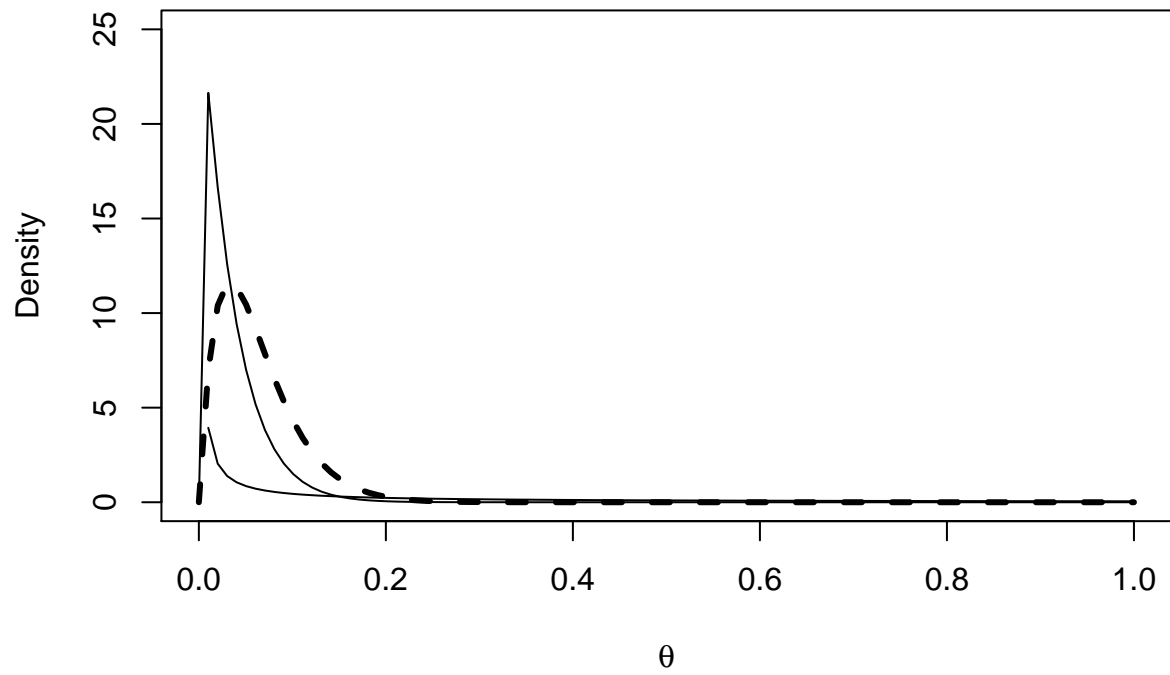## 1 - Reproduce figure 1 from module 2

```r
set.seed(123)

th <- seq(0,1, length.out =100)
c <- seq(0,1, length.out = 100)

# Calculate the prior
a = 0.05
b = 1
sum_x = 1
n = 30
prior <- dbeta(th, a, b)
like <- dbeta(th, sum_x+1, n - sum_x + 1)
post <- dbeta(th, sum_x+a, n-sum_x+b)

# Graph this:

plot(th, like, type="l", main="Likelihood", ylab="Density", lty=2, lwd=3,
     xlab=expression(theta), ylim=c(0,25))
lines(th, prior)
lines(th, post)
```
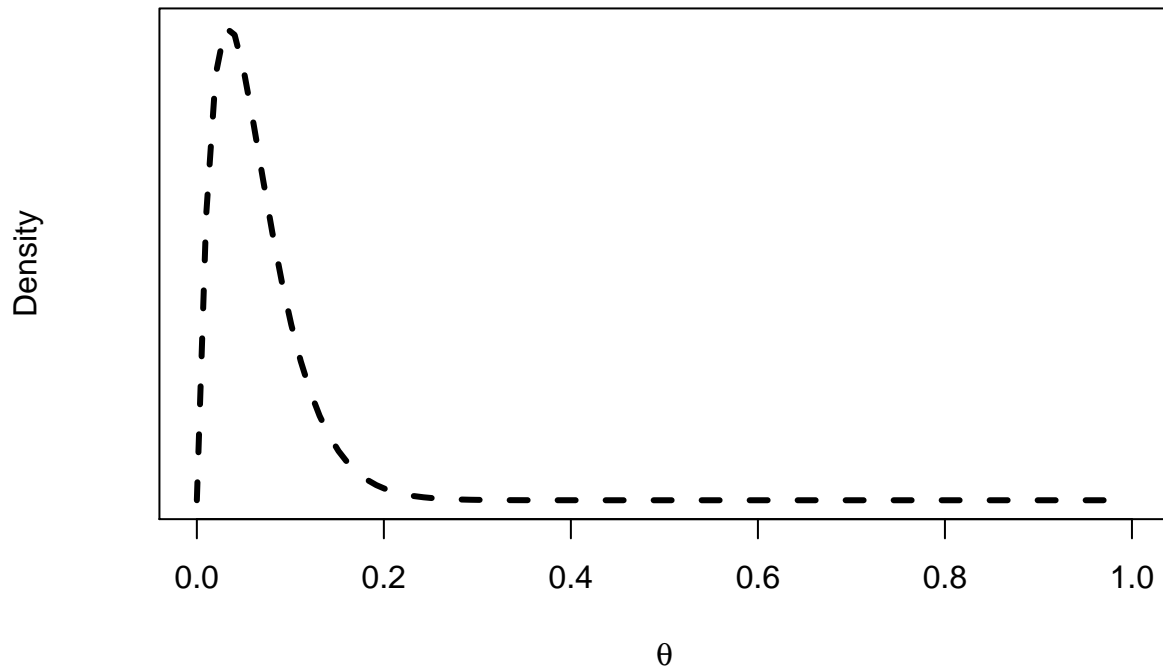
**Likelihood**



```
plot(th,like, type="l", ylab="Density", lty=2, lwd=3,
     xlab=expression(theta), yaxt="n",
     main="With a non-informative prior")
```

## With a non-informative prior



```r
# Calculate the loss
loss <- function(theta, c) {
  if(c < theta) {
    return(10*abs(theta-c))
  } else {
    return(abs(theta-c))
  }
}
an = a + sum_x
bn = b + n - sum_x
posterior_risk <- function(c, s = 30000) {

  theta <- rbeta(s, an, bn)

  loss <- apply(as.matrix(theta), 1, loss, c)

  risk <- mean(loss)
}

c <- seq(0, 0.5, by=0.01)
post_risk <- apply(as.matrix(c), 1, posterior_risk)

plot(c, post_risk, type="l",col="blue", lwd=3, ylab='p(c,x)')
```
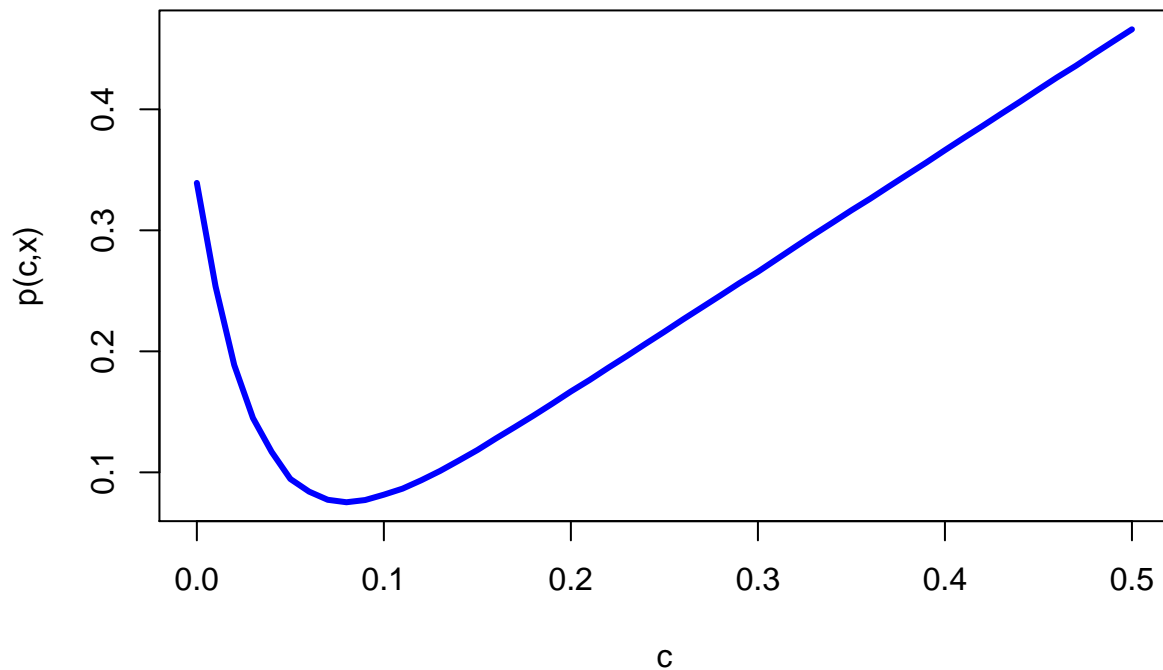
```
# Calculate the minimum
print(c[which.min(post_risk)])
```

```
## [1] 0.08
```

## 2 - Calculate the posterior risk

```r
posterior_risk <- function(c, a_prior, b_prior, sum_x, n, s=30000) {

  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, c)

  risk <- mean(loss)
}
as <- c(0.5, 1, 0.05); bs <- c(1,1,10)
post_risk <- matrix(NA, 3, length(c))

for(i in 1:3) {
  a_prior = as[i]
  b_prior = bs[i]
```
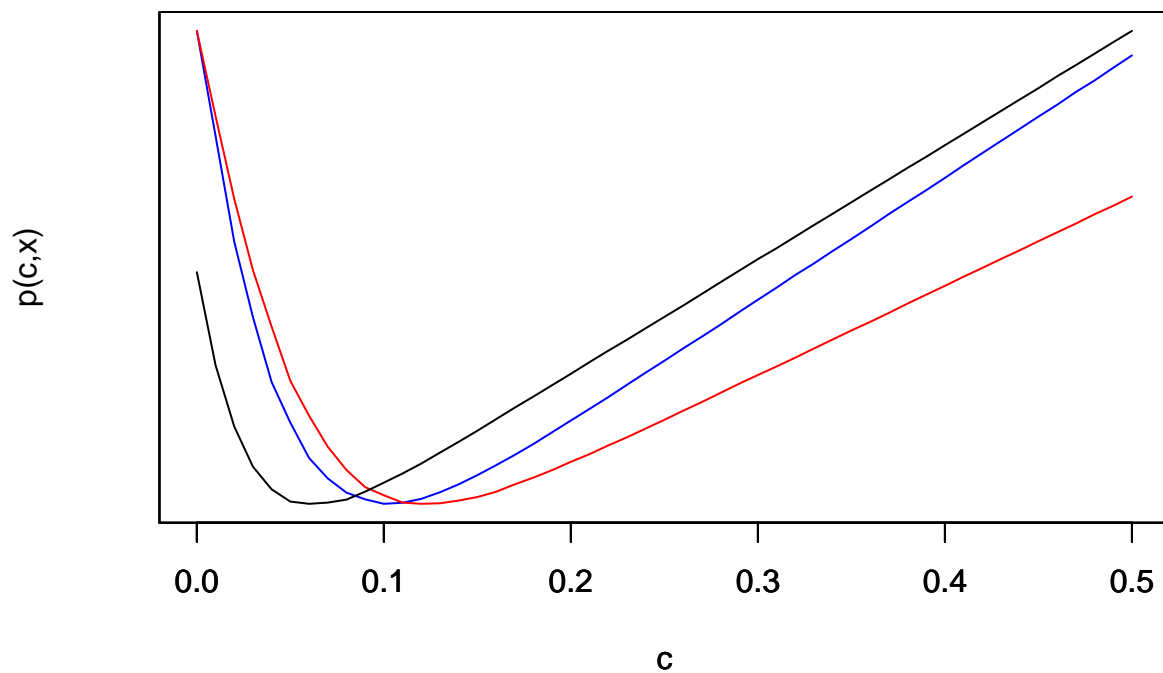
```
  post_risk[i,] <- apply(as.matrix(c), 1, posterior_risk, a_prior, b_prior, sum_x, n)
}


plot(c, post_risk[1,], type="l", col="blue", lty=1, yaxt="n", ylab="p(c,x)")
par(new=T)
plot(c, post_risk[2,], type="l", col="red", lty=1, yaxt="n",ylab="")
par(new=T)
plot(c, post_risk[3,], type="l", col="black", lty=1, yaxt="n", ylab="")
```



## Task 3. Consider the Bayes procedure, $c = \bar{x}$, and c = 0.1.

```
sum_xs = seq(0,30)
n <- 30
a <- 0.05; b <-1

min_c = matrix(NA, 3, length(sum_xs))

optimal_c <- function(sum_x, a_prior, b_prior, n, s=500) {

  # Set up values of C to consider.
  c <- seq(0,1, by=0.01)
  # Calculate posterior risk for each c
  post_risk <- apply(as.matrix(c), 1, posterior_risk, a_prior,b_prior, sum_x, n, s)
```

```
  # Return the optimal C
  c[which.min(post_risk)]
}

# Return the bayesian risk
min_c[1,] = apply(as.matrix(sum_xs), 1, optimal_c, a, b, n)

min_c[2,] = sum_xs/n
min_c[3,] = 0.1

plot(sum_xs, min_c[1,], col='blue',type='o', pch=16,
     ylab="resources allocated", xlab="observed number of disease cases",
     ylim=c(0,1))
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"))
```
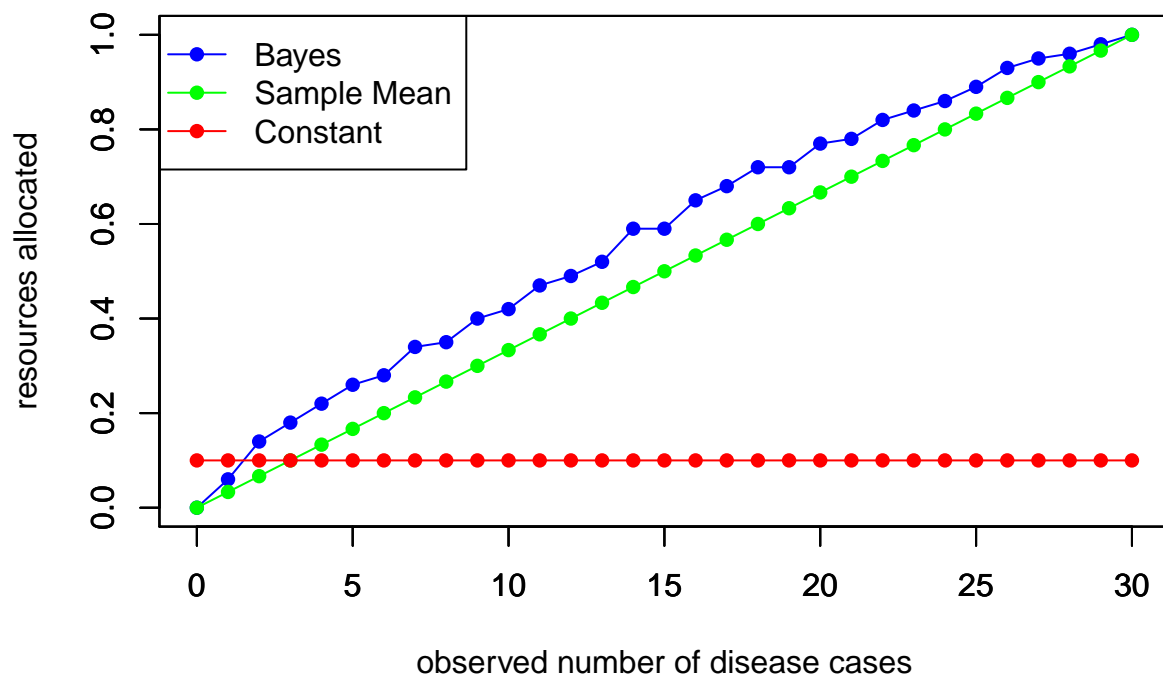
# Task 4 - Plot the frequentist risk as a function of $\theta$.

Frequentist risk is defined as:

$$R(\theta, \delta) = \mathbb{E}(\ell(\theta, \delta(X))|\theta = \theta)$$

Therefore, we must calculate the loss as a function of $\theta$.

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


frequentist_risk <- function(theta) {
  sum_xs = rbinom(100, 30, theta)

  bayes_optimal <- apply(as.matrix(sum_xs), 1, optimal_c, a, b, n, s=100)

  mean_c <- sum_xs/30

  risk = matrix(NA, 3, length(sum_xs))

  # Calculate bayes risk
  loss1 <- apply(as.matrix(bayes_optimal), 1, loss, theta=theta)
  loss2 <- apply(as.matrix(mean_c), 1, loss, theta=theta)

  r1 <- mean(loss1)
  r2 <- mean(loss2)
  r3 <- loss(theta,0.1)
  return(c(r1, r2, r3))
}

risk <- apply(as.matrix(th), 1, frequentist_risk)

plot(th, risk[1,], col='blue', type='l', ylab='frequentist risk',
     xlab=expression(theta), ylim=c(0,1))

par(new=T)

plot(th, risk[2,], col='green', type='l', ylab='',xlab='', ylim=c(0,1))
par(new=T)
plot(th, risk[3,], col='red', type='l', ylab='',xlab='', ylim=c(0,1))
legend("topright", lty=c(1,1,1), col=c("blue","green", "red"),
       legend=c("Bayes", "Sample Mean", "constant"))
```
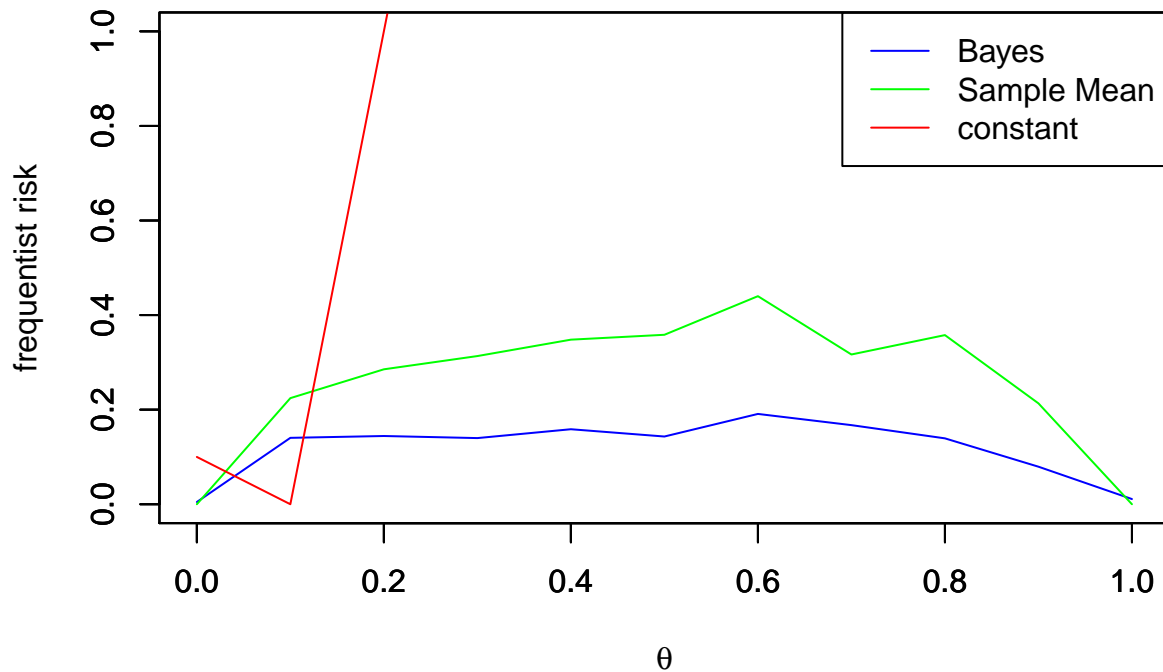
In general, the Bayes estimator tends to yield the lowest frequentist risk.

## Task 5

Because there are thetas for which each of the estimators has the lowest risk, all of the estimators are considered admissible. The sample mean estimator appears to be the best estimator when theta is either really close to 0 or really close to 1, and the constant is optimal when theta is close to 0.1. However, in general, the Bayes estimator has the lowest frequentist risk.

2. (15 points total) *The Uniform-Pareto*

   **The goal of this problem is to continue getting more practice calculating the posterior distribution.**

   Suppose $a < x < b$. Consider the notation $I_{(a,b)}(x)$, where $I$ denotes the indicator function. We define $I_{(a,b)}(x)$ to be the following:

   $$I_{(a,b)}(x) = \begin{cases} 1 & \text{if } a < x < b, \\ 0 & \text{otherwise.} \end{cases}$$

Let X be a random variable and let x be an observed value. Let

$$X = x \mid \theta \sim \text{Uniform}(0, \theta)$$
$$\theta \sim \text{Pareto}(\alpha, \beta),$$

where $p(\theta) = \dfrac{\alpha \beta^\alpha}{\theta^{\alpha+1}} I_{(\beta,\infty)}(\theta)$. Write out the likelihood $p(X = x \mid \theta)$. Then calculate the posterior distribution of $\theta|X = x$.

**Answer to problem 2**

The likelihood is given by:

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

Given this likelihood, we can calculate the posterior as follows:

$$\begin{aligned}
p(\theta|X = x) &= p(X = x|\theta)p(\theta) \\
&= \frac{1}{\theta} I_{(0,\theta)}(x) \frac{\alpha\beta^\alpha}{\theta^{\alpha+1}} I_{(\beta,\infty)}(\theta) \\
&\propto \frac{1}{\theta} I_{(0,\theta)}(x) \frac{1}{\theta^{\alpha+1}} I_{(\beta,\infty)}(\theta) \\
&\propto \frac{1}{\theta^{\alpha+2}} I_{(\max(\beta,x),\infty)}(\theta) \\
&= \text{Pareto}(\alpha + 1, \max(\beta, x))
\end{aligned}$$

Thus, we see that we have a conjugate relationship between the prior and the posterior with respect to $\theta$. We note that as we observe $x$, we either update the lower bound on $\theta$ or keep it the same, and that we continue to increment $\alpha$.

3. (15 points total) *The Bayes estimator or Bayes rule*
   **The goal of this problem is to practice a similar problem that we considered in Module 2, where we derived the Bayes rule under squared error loss and found the result was the posterior mean.**

(a) (5 pts) Find the Bayes estimator (or Bayes rule) when the loss function is $L(\theta, \delta(x)) = c\ (\theta - \delta(x))^2$, where $c > 0$ is a constant.

**Answer**

To find the Bayes rule, we want to find the estimator $\hat{\theta}$ for which the posterior risk is minimized. In this case:

$$\begin{aligned}
\rho(\delta(x), x) &= \mathbb{E}(\ell(\theta, \delta(x))|x) = \mathbb{E}(c(\theta - \delta(x))^2|x) \\
&= c\mathbb{E}(\theta^2|x) - c\mathbb{E}(2\theta\delta(x)|x) + c\mathbb{E}(\delta(x)^2|x)
\end{aligned}$$

Given the above, we can solve for the first derivative and solve for zero to get the minimum value of $\delta(x)$, which will give Bayes rule.

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

Solving for zero, we get:

$$\delta(x) = \mathbb{E}(\theta|x)$$

Intuitively, this makes sense. We should expect that our best estimate is still the posterior mean, because we are multiplying our loss function with a positive constant.

(b) (10 pts) Derive the Bayes estimator (or Bayes rule) when $L(\theta, \delta(x)) = w(\theta)(g(\theta) - \delta(x))^2$. Do so without writing any integrals. Note that you can write $\rho(\pi, \delta(x)) = E[L(\theta, \delta(x))|X]$. You may assume that $w(\theta) > 0$. Don't forget to prove or state why the Bayes rule(s) are unique.

**Answer**

Again, we start by writing out the posterior risk:

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

We then solve for the partial derivative of $\rho$:

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

Setting equal to zero, we can then solve for $\delta(x)$.

$$\mathbb{E}[w(\theta)g(\theta)|X] = \mathbb{E}[w(\theta)\delta(x)|X]$$

And thus,

$$\delta(x) = \frac{\mathbb{E}[w(\theta)g(\theta)|X]}{\mathbb{E}[w(\theta)|X]}$$

We can confirm that this is the maximum by solving for the second derivative of $\rho$.

$$\frac{\partial^2 \rho(\pi, \delta(x))}{\partial \delta(x)^2} = 2\mathbb{E}[w(\theta)|X]$$

Since this value will always be positive (since $w(\theta)$ is always positive), then we know that our solution is unique.