Lab 3: Decision Theory

Olivier Binette

28/08/2020

Agenda

1. Review of decision theory
2. Walkthrough of lab 3
   - A little bit of functional programming
3. Questions

Note: I can talk more about homework 3 and conjugate families at my OH (this lab is pretty packed already).

Review of decision theory

Review of decision theory

Statistical ingredients:

- Statistical model for your data X defined by a likelihood function $p(X \mid \theta)$.
- Prior $p(\theta)$.

Decision ingredients:

- Action $a = a(X)$.
- Loss function $\ell(\theta_0, a)$ which gives you the loss of taking action a if the true value of the parameter is θ_0.

Review of decision theory

Statistical ingredients:

- Statistical model for your data X defined by a likelihood function $p(X \mid \theta)$.
- Prior $p(\theta)$.

Decision ingredients:

- Action $a = a(X)$.
- Loss function $\ell(\theta_0, a)$ which gives you the loss of taking action a if the true value of the parameter is θ_0.

Review of decision theory

Statistical ingredients:

- Statistical model for your data X defined by a likelihood function $p(X \mid \theta)$.
- Prior $p(\theta)$.

Decision ingredients:

- Action $a = a(X)$.
- Loss function $\ell(\theta_0, a)$ which gives you the loss of taking action a if the true value of the parameter is θ_0.

Review of decision theory

Example: You want to estimate a normal mean with known variance.

- Model: $X \mid \mu \sim N(0, 1)$.
- Prior $p(\mu) \propto e^{-\mu^2/2}$.

- Action (how you choose to estimate μ): $\hat{\mu} = X$.
- Loss: $\ell(\mu, \hat{\mu}) = (\mu - \hat{\mu})^2$.

Review of decision theory

Example: You want to estimate a normal mean with known variance.

- Model: $X \mid \mu \sim N(0, 1)$.
- Prior $p(\mu) \propto e^{-\mu^2/2}$.

- Action (how you choose to estimate μ): $\hat{\mu} = X$.
- Loss: $\ell(\mu, \hat{\mu}) = (\mu - \hat{\mu})^2$.

Review of decision theory

Example: You want to estimate a normal mean with known variance.

- Model: $X \mid \mu \sim N(0, 1)$.
- Prior $p(\mu) \propto e^{-\mu^2/2}$.

- Action (how you choose to estimate μ): $\hat{\mu} = X$.
- Loss: $\ell(\mu, \hat{\mu}) = (\mu - \hat{\mu})^2$.

Review of decision theory

How do we evaluate the quality of an action?

Frequentist risk:

- Think of it as the “what if?” risk.
- What if $\theta = 0$? What if $\theta = 0.1$? What if ...
- For all possible values of θ, we compute the expected loss with respect to $X \sim p(X \mid \theta)$:

$$R(\theta, a) = \mathbb{E}_{X \sim p(X \mid \theta)} \left[\ell(\theta, a(X)) \right] = \int \ell(\theta, a(X)) p(x \mid \theta) \, dx.$$

Review of decision theory

How do we evaluate the quality of an action?

Frequentist risk:

- Think of it as the "what if?" risk.
- What if $\theta = 0$? What if $\theta = 0.1$? What if ...
- For all possible values of θ, we compute the expected loss with respect to $X \sim p(X \mid \theta)$:

$$R(\theta, a) = \mathbb{E}_{X \sim p(X|\theta)} [\ell(\theta, a(X))] = \int \ell(\theta, a(X)) p(x \mid \theta) \, dx.$$

Review of decision theory

How do we evaluate the quality of an action?

Frequentist risk:

- Think of it as the "what if?" risk.
- What if $\theta = 0$? What if $\theta = 0.1$? What if ...
- For all possible values of θ, we compute the expected loss with respect to $X \sim p(X \mid \theta)$:

$$R(\theta, a) = \mathbb{E}_{X \sim p(X \mid \theta)} [\ell(\theta, a(X))] = \int \ell(\theta, a(X)) p(x \mid \theta)\, dx.$$

Review of decision theory

How do we evaluate the quality of an action?

Frequentist risk:

- Think of it as the "what if?" risk.
- What if $\theta = 0$? What if $\theta = 0.1$? What if ...
- For all possible values of θ, we compute the expected loss with respect to $X \sim p(X \mid \theta)$:

$$R(\theta, a) = \mathbb{E}_{X \sim p(X|\theta)} \left[\ell(\theta, a(X))\right] = \int \ell(\theta, a(X)) p(x \mid \theta) \, dx.$$

Review of decision theory

How do we evaluate the quality of an action?

Posterior risk:

- It comes after observing the data.
- Given the observed X and given my prior on θ, what do I expect to be my loss if I take the action $a = a(X)$?

$$\rho(a, X) = \mathbb{E}_{\theta \sim p(\theta|X)} [\ell(\theta, a)] = \int \ell(\theta, a) p(\theta \mid X) \, d\theta.$$

Review of decision theory

How do we evaluate the quality of an action?

Posterior risk:

- It comes after observing the data.
- Given the observed X and given my prior on θ, what do I expect to be my loss if I take the action $a = a(X)$?

$$\rho(a, X) = \mathbb{E}_{\theta \sim p(\theta|X)} \left[\ell(\theta, a)\right] = \int \ell(\theta, a) p(\theta \mid X) \, d\theta.$$

Review of decision theory

How do we evaluate the quality of an action?

Posterior risk:

- It comes after observing the data.
- Given the observed X and given my prior on θ, what do I expect to be my loss if I take the action $a = a(X)$?

$$\rho(a, X) = \mathbb{E}_{\theta \sim p(\theta|X)} [\ell(\theta, a)] = \int \ell(\theta, a) p(\theta \mid X) \, d\theta.$$

Review of decision theory

How do we evaluate the quality of an action?

Posterior risk:

- It comes after observing the data.
- Given the observed X and given my prior on θ, what do I expect to be my loss if I take the action $a = a(X)$?

$$\rho(a, X) = \mathbb{E}_{\theta \sim p(\theta \mid X)} [\ell(\theta, a)] = \int \ell(\theta, a) p(\theta \mid X) \, d\theta.$$

Review of decision theory

Summary:

- The **frequentist risk** considers fixed values of θ and random X.
- The **posterior risk** considers fixed observed X and random values of θ (following the posterior distribution).

Review of decision theory

Summary:

- The **frequentist risk** considers fixed values of θ and random X.
- The **posterior risk** considers fixed observed X and random values of θ (following the posterior distribution).

Review of decision theory

Summary:

- The **frequentist risk** considers fixed values of θ and random X.
- The **posterior risk** considers fixed observed X and random values of θ (following the posterior distribution).

Review of decision theory

Finally:

- The **Bayes rule** is to take the action which minimizes the posterior risk:

$$a = a(X) = \arg\min_{\delta} \rho(\delta, X) = \arg\min_{\delta} \mathbb{E}_{\theta \sim p(\theta|X)} \left[\ell(\theta, \delta) \right].$$

- A rule is **inadmissible** if its frequentist risk is always worst than for some other rule. It is **admissible** otherwise.

Review of decision theory

Finally:

- The **Bayes rule** is to take the action which minimizes the posterior risk:

$$a = a(X) = \arg\min_{\delta} \rho(\delta, X) = \arg\min_{\delta} \mathbb{E}_{\theta \sim p(\theta|X)} \left[\ell(\theta, \delta) \right].$$

- A rule is **inadmissible** if its frequentist risk is always worst than for some other rule. It is **admissible** otherwise.

Review of decision theory

Finally:

- The **Bayes rule** is to take the action which minimizes the posterior risk:

$$a = a(X) = \arg\min_{\delta} \rho(\delta, X) = \arg\min_{\delta} \mathbb{E}_{\theta \sim p(\theta|X)} \left[\ell(\theta, \delta) \right].$$

- A rule is **inadmissible** if its frequentist risk is always worst than for some other rule. It is **admissible** otherwise.

Lab 3

Lab 3

Problem (resource allocation):

- Public health officials need to allocate resources for infected individuals in a small city.
- They want to choose the fraction c of the population which will be covered by the resources.
- Unknown proportion θ of the population is infected, prior $\theta \sim \text{Beta}(a, b)$ with $a = 0.05$ and $b = 1$.
- Loss function loss function $\ell(\theta, c) = |\theta - c|$ if $c \geq \theta$, $\ell(\theta, c) = 10|\theta - c|$ if $c < \theta$.
- Sample $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernouilli}(\theta)$, observed $\sum_{i=1}^{n} X_i = 1$, $n = 30$.

Lab 3

Problem (resource allocation):

- Public health officials need to allocate resources for infected individuals in a small city.
- They want to choose the fraction c of the population which will be covered by the resources.
- Unknown proportion θ of the population is infected, prior $\theta \sim \text{Beta}(a, b)$ with $a = 0.05$ and $b = 1$.
- Loss function loss function $\ell(\theta, c) = |\theta - c|$ if $c \geq \theta$, $\ell(\theta, c) = 10|\theta - c|$ if $c < \theta$.
- Sample $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernouilli}(\theta)$, observed $\sum_{i=1}^{n} X_i = 1$, $n = 30$.

Lab 3

Problem (resource allocation):

- Public health officials need to allocate resources for infected individuals in a small city.
- They want to choose the fraction c of the population which will be covered by the resources.
- Unknown proportion θ of the population is infected, prior $\theta \sim \text{Beta}(a, b)$ with $a = 0.05$ and $b = 1$.
- Loss function loss function $\ell(\theta, c) = |\theta - c|$ if $c \geq \theta$, $\ell(\theta, c) = 10|\theta - c|$ if $c < \theta$.
- Sample $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernouilli}(\theta)$, observed $\sum_{i=1}^{n} X_i = 1$, $n = 30$.

Lab 3

Problem (resource allocation):

- Public health officials need to allocate resources for infected individuals in a small city.
- They want to choose the fraction c of the population which will be covered by the resources.
- Unknown proportion θ of the population is infected, prior $\theta \sim \text{Beta}(a, b)$ with $a = 0.05$ and $b = 1$.
- Loss function loss function $\ell(\theta, c) = |\theta - c|$ if $c \geq \theta$, $\ell(\theta, c) = 10|\theta - c|$ if $c < \theta$.
- Sample $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernouilli}(\theta)$, observed $\sum_{i=1}^{n} X_i = 1$, $n = 30$.

Lab 3

Problem (resource allocation):

- Public health officials need to allocate resources for infected individuals in a small city.
- They want to choose the fraction c of the population which will be covered by the resources.
- Unknown proportion θ of the population is infected, prior $\theta \sim \text{Beta}(a, b)$ with $a = 0.05$ and $b = 1$.
- Loss function loss function $\ell(\theta, c) = |\theta - c|$ if $c \geq \theta$, $\ell(\theta, c) = 10|\theta - c|$ if $c < \theta$.
- Sample $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernouilli}(\theta)$, observed $\sum_{i=1}^{n} X_i = 1$, $n = 30$.

Task 1

Plot the posterior risk $\rho(c, \{X_i\})$ as a function of c and find where the minimum occurs.

- First let's code up the loss function in R:

```
loss <- function(theta, c){
  (c >= theta)*(c - theta) + (c < theta)*10*(theta - c)
}
```

- We know that the posterior distribution of θ is $\text{Beta}(a+1, b+29)$.
- Therefore the posterior risk is

```
post_risk <- Vectorize(function(c, a=0.05, b=1, nsim=5000){
  theta.post = rbeta(nsim, a+1, b+29)
  mean(loss(theta.post, c))
})
```

Task 1

Plot the posterior risk $\rho(c, \{X_i\})$ as a function of c and find where the minimum occurs.

- First let's code up the loss function in R:

```
loss <- function(theta, c){
  (c >= theta)*(c - theta) + (c < theta)*10*(theta - c)
}
```

- We know that the posterior distribution of θ is $\text{Beta}(a+1, b+29)$.
- Therefore the posterior risk is

```
post_risk <- Vectorize(function(c, a=0.05, b=1, nsim=5000){
  theta.post = rbeta(nsim, a+1, b+29)
  mean(loss(theta.post, c))
})
```

Task 1

Plot the posterior risk $\rho(c, \{X_i\})$ as a function of c and find where the minimum occurs.

- First let's code up the loss function in R:

```
loss <- function(theta, c){
  (c >= theta)*(c - theta) + (c < theta)*10*(theta - c)
}
```

- We know that the posterior distribution of θ is $\text{Beta}(a+1, b+29)$.
- Therefore the posterior risk is

```
post_risk <- Vectorize(function(c, a=0.05, b=1, nsim=5000){
  theta.post = rbeta(nsim, a+1, b+29)
  mean(loss(theta.post, c))
})
```

Task 1

Plot the posterior risk $\rho(c, \{X_i\})$ as a function of c and find where the minimum occurs.

- First let's code up the loss function in R:

```
loss <- function(theta, c){
  (c >= theta)*(c - theta) + (c < theta)*10*(theta - c)
}
```

- We know that the posterior distribution of θ is $\text{Beta}(a+1, b+29)$.

- Therefore the posterior risk is

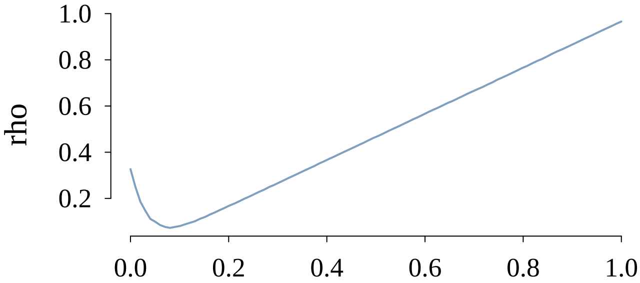
```
post_risk <- Vectorize(function(c, a=0.05, b=1, nsim=5000){
  theta.post = rbeta(nsim, a+1, b+29)
  mean(loss(theta.post, c))
})
```

Task 1

Plot the posterior risk $\rho(c, \{X_i\})$ as a function of c and find where the minimum occurs.

- And we can plot the posterior risk:

```
source(url("https://gist.githubusercontent.com/OlivierBinet

c = seq(0,1, length.out = 100)
rho = post_risk(c)
plot(c, rho, type="l")
```



Task 1

Plot the posterior risk $\rho(c, \{X_i\})$ as a function of c and find where the minimum occurs.

- The minimum of the posterior risk is

```
c[which.min(rho)]
```

```
## [1] 0.08080808
```

Task 2

Perform a sensitivity analysis for the choice of prior.

- Let's define some other prior parameters in a reasonable range:

```
as = c(0.01, 0.05, 0.1, 1)
bs = c(0.5, 1, 3, 10)

prior.params = expand.grid(a=as, b=bs)
head(prior.params)
```

```
##      a   b
## 1 0.01 0.5
## 2 0.05 0.5
## 3 0.10 0.5
## 4 1.00 0.5
## 5 0.01 1.0
## 6 0.05 1.0
```

Task 2

Perform a sensitivity analysis for the choice of prior.

- Let's define some other prior parameters in a reasonable range:

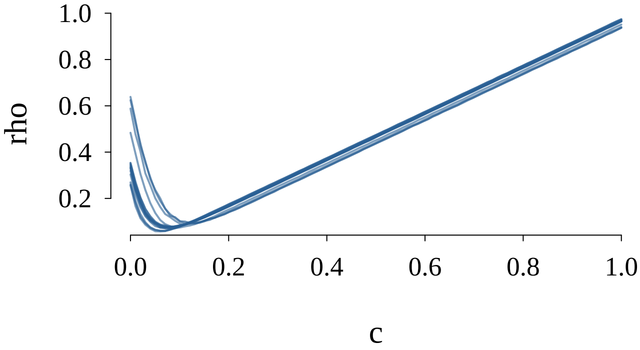
```r
as = c(0.01, 0.05, 0.1, 1)
bs = c(0.5, 1, 3, 10)

prior.params = expand.grid(a=as, b=bs)
head(prior.params)
```

```
##      a   b
## 1 0.01 0.5
## 2 0.05 0.5
## 3 0.10 0.5
## 4 1.00 0.5
## 5 0.01 1.0
## 6 0.05 1.0
```

Task 2

Perform a sensitivity analysis for the choice of prior.

```
rho = post_risk(c, a=prior.params[1,1], b=prior.params[1,2])
plot(c, rho, type="l")

for (i in 2:nrow(prior.params)) {
  rho = post_risk(c,  a=prior.params[i,1], b=prior.params[i,2])
  lines(c, rho)
}
```



Task 3

Plot the Bayes decision rule, the empirical mean and the constant decision $c = 0.1$ as a function of the number of observed cases.

- First let's rewrite the posterior risk function to take the number of observed cases as an argument.

```
post_risk <- Vectorize(function(c, X, n=30, a=0.05, b=1, ns
  theta.post = rbeta(nsim, a+X, b+n-X)
  mean(loss(theta.post, c))
})
```

- Now let's define the Bayes rule:

```
bayes_rule = Vectorize(function(X, n=30, nsim=400) {
  c = seq(0,1, length.out=100)
  c[which.min(post_risk(c, X, nsim=nsim))]
})
```

Task 3

Plot the Bayes decision rule, the empirical mean and the constant decision $c = 0.1$ as a function of the number of observed cases.

- First let's rewrite the posterior risk function to take the number of observed cases as an argument.

```
post_risk <- Vectorize(function(c, X, n=30, a=0.05, b=1, ns
  theta.post = rbeta(nsim, a+X, b+n-X)
  mean(loss(theta.post, c))
})
```

- Now let's define the Bayes rule:

```
bayes_rule = Vectorize(function(X, n=30, nsim=400) {
  c = seq(0,1, length.out=100)
  c[which.min(post_risk(c, X, nsim=nsim))]
})
```

Task 3

Plot the Bayes decision rule, the empirical mean and the constant decision $c = 0.1$ as a function of the number of observed cases.

- First let's rewrite the posterior risk function to take the number of observed cases as an argument.

```
post_risk <- Vectorize(function(c, X, n=30, a=0.05, b=1, ns
  theta.post = rbeta(nsim, a+X, b+n-X)
  mean(loss(theta.post, c))
})
```

- Now let's define the Bayes rule:

```
bayes_rule = Vectorize(function(X, n=30, nsim=400) {
  c = seq(0,1, length.out=100)
  c[which.min(post_risk(c, X, nsim=nsim))]
})
```

Task 3

Plot the Bayes decision rule, the empirical mean and the constant decision $c = 0.1$ as a function of the number of observed cases.

- Let's also define the sample mean rule and the constant rule:

```
mean_rule <- Vectorize(function(X, n=30) {
  return(X/n)
})

constant_rule <- Vectorize(function(X, n=30) {
  return(0.1)
})
```

Task 3

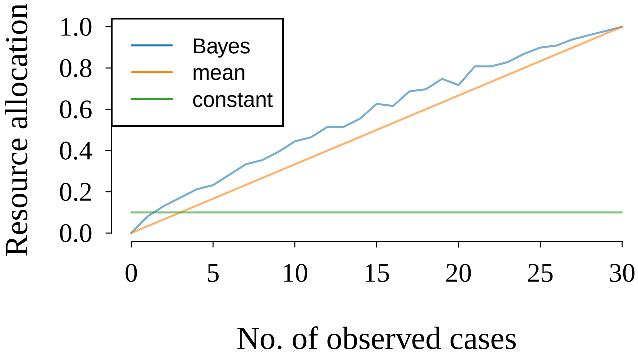
Plot the Bayes decision rule, the empirical mean and the constant decision $c = 0.1$ as a function of the number of observed cases.

- Finally we can plot the different rules:

```
X = 0:30

# Bayes rule
plot(X, bayes_rule(X), type="l", col=1)

# Empirical mean
lines(X, X/30, col=2)

# Constant rule
lines(X, rep(0.1, length(X)), col=3)

legend("bottomright", legend=c("Bayes", "mean", "constant"),
       lty=1, col=1:3)
```

Task 3

Plot the Bayes decision rule, the empirical mean and the constant decision $c = 0.1$ as a function of the number of observed cases.



Task 4

Plot the frequentist risk $R(\theta, \delta)$ as a function of θ for the three procedures in the previous task. Report your findings.

- Let's define the frequentist risk:

```
freq_risk = Vectorize(function(rule, theta, nsim=400) {
  X.s = rbinom(nsim, 30, theta)
  mean(loss(theta, rule(X.s)))
}, vectorize.args = "theta")
```

Task 4

Plot the frequentist risk $R(\theta, \delta)$ as a function of θ for the three procedures in the previous task. Report your findings.

- Let's define the frequentist risk:

```
freq_risk = Vectorize(function(rule, theta, nsim=400) {
  X.s = rbinom(nsim, 30, theta)
  mean(loss(theta, rule(X.s)))
}, vectorize.args = "theta")
```

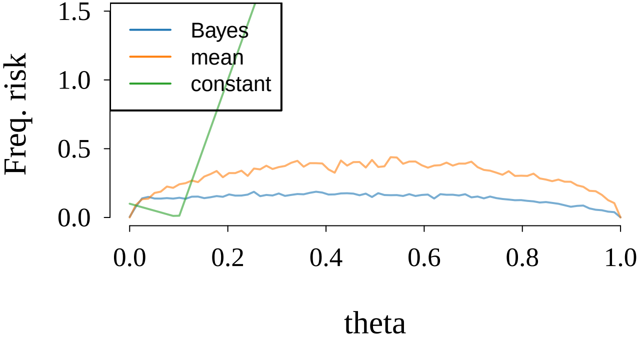
Task 4

Plot the frequentist risk $R(\theta, \delta)$ as a function of θ for the three procedures in the previous task. Report your findings.

```
theta = seq(0,1, length.out=20)

# Bayes rule
plot(theta, freq_risk(bayes_rule, theta), type="l", ylim=c(0,1),
     xlab="theta", ylab="Freq. risk",
     col=cmap.seaborn(1))

# Sample mean
lines(theta, freq_risk(mean_rule, theta), col=cmap.seaborn(2))

# Constant
lines(theta, freq_risk(constant_rule, theta), col=cmap.seaborn(3

legend("topleft", legend=c("Bayes", "mean", "constant"),
       lty=1, col=cmap.seaborn(c(1,2,3)),
       cex=0.7)
```

Task 4

Plot the frequentist risk $R(\theta, \delta)$ as a function of θ for the three procedures in the previous task.



Task 5

Based on your plot of the frequentist risk, consider the three estimators—the constant, the mean, and the Bayes estimators. Which estimators are admissible? Be sure to explain why or why they are not admissible.