STAT529
Applied Bayesian Decision Theory
Homework 1

Fangda Li
li1208@purdue.edu

March 10, 2017

1 Paper

The chosen paper is:

Kiesel, Scott, and Wheeler Ruml. "A bayesian effort bias for sampling-based motion planning." 26th International Conference on Automated Planning and Scheduling: Planning and Robotics Workshop. 2016.

Motion planning problem in the context of robotics aims to find a path between the start state and the goal state, while avoiding known obstacles. Moreover, in the study of *sampling-based* motion planning algorithms, an efficient algorithm tends to sample less random states before a path is found. In this work, the authors used a Bayesian approach to estimate the "effort" that it takes for each existing state in the search graph to grow into the goal region. As a result, in order to expand the search graph, their algorithm simply selects the state with the least amount of "effort".

More specifically, the authors used a Beta distribution as the prior to estimate the "effort" for a given state. Then, the observation comes from an attempt to expand that state, of which the outcome is modeled as a Bernoulli trial. A successful attempt will update the posterior with a higher α and a lower β, thus shifting the distribution toward 1, which indicates less "effort".

2 C-scan

The decision model is setup as follows:

- State of nature Θ: $\Theta = \{\theta_1, \theta_2\}$, where θ_1 refers to *baby is okay*, θ_2 refers to *baby is not okay*.
- Action space A: $A = \{a_1, a_2\}$, where a_1 refers to *do C-section*, θ_2 refers to *not do C-section*.

Table 1: Loss function $L(a, \theta)$ for Question 2

	θ_1	θ_2
a_1	0	c
a_2	d	0

- Loss function $L(a, \theta)$, as shown in Table 1.
- Observation space X: $X = \{x_1, x_2\}$, where x_1 refers to *baby looks okay from C-scan*, x_2 refers to *baby looks not okay from C-scan.*
- Data distribution $f(x|\theta)$, as shown in Table 2.

Table 2: Data distribution $f(x|\theta)$ for Question 2

	θ_1	θ_2
x_1	0.7	0.1
x_2	0.3	0.9

- The prior $\Pi(\theta)$:

$$P(\Theta = \theta_1) = \pi, P(\Theta = \theta_2) = 1 - \pi.$$

- The procedure space T, as shown in Table 3.

Table 3: All possible procedures for Question 2

	x_1	x_2
t_1	a_1	a_1
t_2	a_1	a_2
t_2	a_2	a_1
t_2	a_2	a_2

- The Frequentist risk $r(t(x), \theta)$:

$$r(t(x), \theta) = E[L(t(x), \theta)|\theta] = \sum_x L(t(x)|\theta) f(x|\theta)$$

.

- The Bayes risk $R(t(x), \pi)$:

$$R(t(x), \pi) = \sum_\theta r(t(x), \theta)\Pi(\theta).$$

Now we answer the questions based on the decision model:

Table 4: The Frequentist risks for Q2 a). Bold represents the minimax risk.

	θ_1	θ_2
$r(t_1(x), \theta)$	0	4
$r(t_2(x), \theta)$	**3**	0.4
$r(t_3(x), \theta)$	7	3.6
$r(t_4(x), \theta)$	10	0

Table 5: The Bayes risks for Q2 a). Bold represents the smallest Bayes risk.

$R(t_1(x), \pi)$	0.4
$R(t_1(x), \pi)$	**0.276**
$R(t_1(x), \pi)$	6.66
$R(t_1(x), \pi)$	9.99

a) A procedure $t_n(x) \in T$ is Bayes when $n = \arg\min_n R(t_n(x), \pi)$ and is minimax when $n = \arg\min_n \max_\theta(r(t_n(x), \theta))$. Now we choose $c = 4, d = 10$ and $\pi = 0.9$. Then the Frequentist risks and the Bayes risks $\forall t(x) \in T$ are computed, as shown in Table 4 and Table 5. As a result, t_1 is the Bayes procedure and t_2 is the minimax procedure.

b) Similar to step a), we compute the Frequentist risks and the Bayes risks, where $c = 10, d = 10$ and $\pi = 0.7$. The results are shown in Table 4 and Table 7.

Table 6: The Frequentist risks for Q2 b). Bold represents the minimax risk.

	θ_1	θ_2
$r(t_1(x), \theta)$	0	10
$r(t_2(x), \theta)$	**3**	1
$r(t_3(x), \theta)$	7	9
$r(t_4(x), \theta)$	10	0

Table 7: The Bayes risks for Q2 b). Bold represents the smallest Bayes risk.

$R(t_1(x), \pi)$	3
$R(t_2(x), \pi)$	**2.4**
$R(t_3(x), \pi)$	7.6
$R(t_4(x), \pi)$	7

c) Similar to step a) and b), we compute the Frequentist risks, where $c = 5, d = 10$ and $\pi = 0.9$, as shown in Table 8. Since the risks of t_2 under both states of nature are respectively smaller than those of t_3, t_3 is inadmissible because of t_2.

d) It is impossible to make t_4 inadmissible because the Frequentist risk of t_4 at $\theta = 2$ is

Table 8: The Frequentist risks for Q2 c).

	θ_1	θ_2
$r(t_1(x), \theta)$	0	5
$r(t_2(x), \theta)$	3	0.5
$r(t_3(x), \theta)$	7	4.5
$r(t_4(x), \theta)$	10	0

always the smallest risk possible (zero):

$$\begin{aligned} r(t_4(x), \theta = 2) &= L(t_4(1), \theta = 2)P(x = 1|\theta = 2) + L(t_4(2), \theta = 2)P(x = 2|\theta = 2) \\ &= L(a_2, \theta = 2)P(x = 1|\theta = 2) + L(a_2, \theta = 2)P(x = 2|\theta = 2) \\ &= 0 \times 0.1 + 0 \times 0.9 \\ &= 0. \end{aligned}$$

3 Monty Hall

The door I choose to begin with is *door 1.* Then the decision model is setup as follows:

- State of nature Θ: $\Theta = \{\theta_1, \theta_2\}$, where θ_1 refers to *price behind the same door*, θ_2 refers to *price behind the other two doors.*
- Action space A: $A = \{a_1, a_2\}$, where a_1 refers to *choose the same door*, a_2 refers to *choose the other door.*
- Loss function $L(a, \theta)$, as shown in Table 9.

Table 9: Loss function $L(a, \theta)$ for Question 3

	θ_1	θ_2
a_1	0	10
a_2	10	0

- Observation space X: $X = \{x_1, x_2\}$, where x_1 refers to *host opens door 2*, x_2 refers to *host opens door 3.*
- Data distribution $f(x|\theta)$, as shown in Table 10.

Table 10: Data distribution $f(x|\theta)$ for Question 3

	θ_1	θ_2
x_1	0.5	0.5
x_2	0.5	0.5

- The prior $\Pi(\theta)$:

$$P(\Theta = \theta_1) = \pi = 0.4, P(\Theta = \theta_2) = 1 - \pi = 0.6.$$

- The procedure space T, as shown in Table 11.

Table 11: All possible procedures for Question 3

	x_1	x_2
t_1	a_1	a_1
t_2	a_1	a_2
t_2	a_2	a_1
t_2	a_2	a_2

- The Frequentist risk $r(t(x), \theta)$:

$$r(t(x), \theta) = E[L(t(x), \theta)|\theta] = \sum_x L(t(x)|\theta) f(x|\theta)$$

.

- The Bayes risk $R(t(x), \pi)$:

$$R(t(x), \pi) = \sum_\theta r(t(x), \theta)\Pi(\theta).$$

Based on the decision model established above, the Frequentist and Bayesian risks are derived in Table 12 and 13, respectively. Always switching doors, a.k.a. t_4, achieves the lowest Bayesian risk. Therefore, the Bayes procedure is to choose *door 1* and then always switch to the other door.

Table 12: The Frequentist risks for Q3.

	θ_1	θ_2
$r(t_1(x), \theta)$	0	10
$r(t_2(x), \theta)$	5	5
$r(t_3(x), \theta)$	5	5
$r(t_4(x), \theta)$	10	0

Table 13: The Bayes risks for Q3. Bold represents the smallest Bayes risk.

$R(t_1(x), \pi)$	6
$R(t_2(x), \pi)$	4
$R(t_3(x), \pi)$	4
$R(t_4(x), \pi)$	**3**

4 Cell Phone

a) The prior is given as a Beta distribution, $\Pi(\theta) \sim Beta(\alpha, \beta)$, where $\alpha = 10$ and $\beta = 10$. The outcome of the random experiment (the likelihood) can be modeled as a Binomial distribution:

$$f(x|\theta) = \binom{20}{x}\theta^x(1-\theta)^{20-x}.$$

Then, the posterior distribution will be proportional to a Beta distribution: $\Pi(\theta|x) \sim Beta(x+\alpha, 20-x+\beta)$:

$$\begin{aligned}\Pi(\theta|x) &\propto \binom{20}{x}\theta^x(1-\theta)^{20-x}\frac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\\ &\propto C\frac{1}{B(x+\alpha, 20-x+\beta)}\theta^{(x+\alpha)-1}(1-\theta)^{(20-x+\beta)-1}\\ &\propto Beta(x+\alpha, 20-x+\beta).\end{aligned}$$

Now the posterior expected loss can be computed using the following formula:

$$E[L(t(x),\theta)|x] = \int L(t(x),\theta)\Pi(\theta|x)d\theta$$

For the two actions, their posterior expected losses are:

$$E[L(a_1,\theta)|x] = \int_{0.5}^{1} 10\Pi(\theta|x)d\theta = 10P(\theta > 0.5|x),$$

$$E[L(a_2,\theta)|x] = \int_{0}^{0.5} 5\Pi(\theta|x)d\theta = 5P(\theta \le 0.5|x),$$

where $\Pi(\theta|x) \sim Beta(x+10, 30-x)$. Note that the value of $P(\theta \le 0.5|x)$ is equal to the cdf of the posterior at $\theta = 0.5$. As a result, the cut-off x value that satisfies $P(\theta \le 0.5|x) > \frac{10}{15}$ is calculated to be 8. Finally, we conclude that the Bayes procedure is: *choose a_1 when $x \le 8$, otherwise choose a_2.*

b) Frequentist hypothesis testing is carried out, assuming the null hypothesis to be $H_0 : \theta \le 0.5$:

$$\begin{aligned}P(reject\ H_0|H_0\ is\ valid) &= P(X = C|\theta \le 0.5)\\ &= \binom{20}{C}\theta^C(1-\theta)^{20-C}\\ &= \binom{20}{C}0.5^C(1-0.5)^{20-C}\\ &= 0.05.\end{aligned}$$

By solving the above equation, the smallest integer C that makes $P(reject\ H_0|H_0\ is\ valid) < 0.05$ is 14. As a result, the Frequentist procedure is: *choose a_2 when $C \ge 14$, otherwise choose a_1.*

c) In order to mimic the Frequentist procedure in b), we have to find the α, β of a Beta distribution prior such that $P(\theta \leq 0.5|C = 14) = \frac{10}{15}$ is satisfied. In other words, the cdf value at $\theta = 0.5$ of distribution $Beta(C + \alpha, 20 - C + \beta)$ should be as close as possible to $\frac{10}{15}$. As a result, by iterating through possible values for α and β, the best parameters I get are $\alpha = 8$ and $\beta = 19$.

d) Given a uniform prior, our posterior distribution becomes a Binomial distribution:

$$\Pi(\theta|x) \propto \binom{20}{x}\theta^x(1-\theta)^{20-x} \times 1 \sim Binom(20, \theta).$$

Assuming the two loss functions have the form:

$$L(a_1, \theta) = \begin{cases} 0, & \text{if } \theta \leq 0.5 \\ c_1, & \text{if } 0.5 < \theta \leq 1, \end{cases}$$

$$L(a_2, \theta) = \begin{cases} c_2, & \text{if } \theta \leq 0.5 \\ 0, & \text{if } 0.5 < \theta \leq 1. \end{cases}$$

For the two actions, their posterior expected losses are:

$$E[L(a_1, \theta)|x] = \int_{0.5}^{1} c_1\Pi(\theta|x)d\theta = c_1P(\theta > 0.5|x),$$

$$E[L(a_2, \theta)|x] = \int_{0}^{0.5} c_2\Pi(\theta|x)d\theta = c_2P(\theta \leq 0.5|x).$$

Note that the value of $P(\theta \leq 0.5|x)$ is equal to the cdf of the posterior at $\theta = 0.5$. As a result, the cut-off x value that satisfies $P(\theta \leq 0.5|x) > \frac{c_1}{c_1+c_2}$ will be chosen to be C as in the Frequentist test. Since we already know $C = 14$ and $P(\theta \leq 0.5|x = 14) = 0.9793$, by fixing $c_2 = 1$ for convenience, the c_1 value that makes $\frac{c_1}{c_1+c_2} \leq 0.9793$ as close as possible is 47. Finally, we conclude that the loss function that makes b) a Bayes procedure is $c_1 = 47, c_2 = 1$.

e) The linear loss functions are:

$$L(a_1, \theta) = \begin{cases} 0, & \text{if } \theta \leq 0.5 \\ \theta - 0.5, & \text{if } 0.5 < \theta \leq 1, \end{cases}$$

$$L(a_2, \theta) = \begin{cases} 0.5 - \theta, & \text{if } \theta \leq 0.5 \\ 0, & \text{if } 0.5 < \theta \leq 1. \end{cases}$$

The general Bayes procedure is to find the action that minimizes the expected posterior loss, which is defined as:

$$E[L(a_1, \theta)|x] = \int_{0.5}^{1} (\theta - 0.5)\Pi(\theta|x)d\theta = \int_{0.5}^{1} \theta\Pi(\theta|x)d\theta - 0.5P(\theta > 0.5|x),$$

$$E[L(a_2,\theta)|x] = \int_0^{0.5} (0.5-\theta)\Pi(\theta|x)d\theta = 0.5P(\theta \le 0.5|x) - \int_0^{0.5} \theta\Pi(\theta|x)d\theta.$$

Now, the Bayes procedure will choose a_1 as long as the following inequality is satisfied:

$$\begin{gathered}
E[L(a_1,\theta)|x] \le E[L(a_2,\theta)|x] \\
\int_{0.5}^{1} \theta\Pi(\theta|x)d\theta - 0.5P(\theta > 0.5|x) \le 0.5P(\theta \le 0.5|x) - \int_0^{0.5} \theta\Pi(\theta|x)d\theta \\
\int_0^{0.5} \theta\Pi(\theta|x)d\theta + \int_{0.5}^{1} \theta\Pi(\theta|x)d\theta \le 0.5P(\theta \le 0.5|x) + 0.5P(\theta > 0.5|x) \\
\int_0^{1} \theta\Pi(\theta|x)d\theta \le 0.5,
\end{gathered}$$

where $\int_0^1 \theta\Pi(\theta|x)d\theta$ is the posterior mean. To conclude, the general form of Bayes procedure for this problem is to choose a_1 when the posterior mean is smaller than 0.5 and choose a_2 otherwise.

5 HDL Cholesterol

The decision model is setup as follows:

- State of nature: $\omega = \theta_W - \theta_M$.
- Observation/data:

$$\begin{aligned}
Y = \overline{X}_W - \overline{X}_M &\sim N(\theta_W - \theta_M, \frac{\sigma_W^2}{n_W} + \frac{\sigma_M^2}{n_M}) \\
&\sim N(\omega, 4).
\end{aligned}$$

- Action space $A = \{a_1, a_2, a_3, a_4, a_5\}$, where a_1 refers to *say* $-5 \le \omega < -3$, a_2 refers to *say* $-3 \le \omega < -1$, a_3 refers to *say* $-1 \le \omega < 1$, a_4 refers to *say* $1 \le \omega < 3$, and a_5 refers to *say* $3 \le \omega < 5$.
- Loss function for action a_i:

$$L(a_i,\omega) = \begin{cases} c_{lower} - \omega, & \text{if } \omega \le c_{lower} \\ 0, & \text{if } c_{lower} < \theta \le c_{upper} \\ \omega - c_{upper}, & \text{if } \omega > c_{upper}. \end{cases}$$

- Posterior distribution: $\Pi(\omega|y) \sim N(m, v^2)$.

a) **Prior 1**, $\Pi(\omega) \sim N(0, (2.5)^2)$:

The mean and variance for the posterior distribution can be computed:

$$m = \frac{(2.5)^2 \times 4}{(2.5)^2 + 4} = 2.44$$

$$v^2 = \frac{(2.5)^2 \times 4}{(2.5)^2 + 4} = 2.44.$$

As a result, the posterior distribution $\Pi(\omega|y)$ is $N(2.44, (1.56)^2)$. In order to determine the Bayes procedure, we use the following simulation procedure:

1. Draw n random samples of ω according to the posterior distribution $\Pi(\omega|y)$;
2. For each action $a \in A$, compute the average loss across all random samples of ω with respect to the loss function $L(a_i, \omega)$;
3. Choose the action with the lowest average loss.

Now, let's assume that the state of nature space is all the possible real numbers $\mathbb{R}$. Then the simulation result of the average losses (a.k.a. the *estimated* expected posterior loss) for all actions is shown in Table 14. Finally, we conclude that the Bayes procedure for

Table 14: The *estimated* average loss for prior 1. Bold represents the smallest loss.

$\tilde{E}(L(a_1,\omega)\|y)$	5.454
$\tilde{E}(L(a_2,\omega)\|y)$	3.460
$\tilde{E}(L(a_3,\omega)\|y)$	1.600
$\tilde{E}(L(a_4,\omega)\|y)$	**0.512**
$\tilde{E}(L(a_5,\omega)\|y)$	0.945

prior 1 is a_4: *say* $1 \leq \omega < 3$.

b) **Prior 2**, $\Pi(\omega) \sim N(0, (10)^2)$:

According to the problem statement, the posterior for this prior is $N(y, \sigma^2) = N(4, 4)$. Now, we can compute the Bayes procedure based on the simulation procedure in a). The results can be found in Table 15. Finally, we conclude that the Bayes procedure for prior

Table 15: The *estimated* average loss for prior 2. Bold represents the smallest loss.

$\tilde{E}(L(a_1,\omega)\|y)$	7.015
$\tilde{E}(L(a_2,\omega)\|y)$	5.018
$\tilde{E}(L(a_3,\omega)\|y)$	3.075
$\tilde{E}(L(a_4,\omega)\|y)$	1.461
$\tilde{E}(L(a_5,\omega)\|y)$	**0.790**

2 is a_5: *say* $3 \leq \omega < 5$. This result makes sense since an uninformative prior makes the Bayes procedure devolve into a Frequentist procedure, which makes the decision solely based on data: if the data says $\omega = 4$, then choose the interval that contains $\omega = 4$.

6 Code

```
from pylab import *
from scipy.stats import beta, binom
```

```
fact = math.factorial

def q3():
  loss = array([[0, 10], [10, 0]]) # action * theta
  X = array([[0.5, 0.5], [0.5, 0.5]]) # x * theta
  prior = array([0.4, 0.6])
  r = zeros((4,2))
  R = zeros(4)

  for i, t in enumerate([(0,0), (0,1), (1,0), (1,1)]):
    a1, a2 = t
    for theta in arange(0,2):
      risk = loss[a1,theta] * X[a1,theta] + loss[a2,theta] * X[a2,theta]
      r[i,theta] = risk
      R[i] += risk * prior[theta]

  print 'Frequentist risks:'
  print r
  print 'Min:', argmin(r,axis=0)

  print 'Bayesian risks:'
  print R
  print 'Min:', argmin(R,axis=0)

def q4a():
  X = range(21)
  a_prior, b_prior = 10, 10
  for x in X:
    a_posterior, b_posterior = x+10, 30-x
    beta_posterior = beta.cdf(0.5, a_posterior, b_posterior, loc=0, scale=1)
    print 'x = %d, cdf(0.5) = %f, diff = %f' % (x, beta_posterior,
     ↪ beta_posterior-2./3)

def q4b():
  C = range(21)
  f = zeros(21)
  for i,c in enumerate(C):
    f[i] = fact(20) / (fact(c) * fact(20-c)) * (0.5)**c * (1-0.5)**(20-c)
  plot(C,f,'o')
  show()
  print C
  print f

def q4c():
  x = 14
```

```
  diffs = zeros((20,20))
  for a_prior in arange(1,21):
    for b_prior in arange(1,21):
      a_posterior, b_posterior = x+a_prior, 20-x+b_prior
      beta_posterior = beta.cdf(0.5, a_posterior, b_posterior, loc=0,
       ↪  scale=1)
      diff = beta_posterior-2./3
      diffs[a_prior-1, b_prior-1] = diff
      # print 'x = %d, cdf(0.5) = %f, diff = %f' % (x, beta_posterior,
       ↪  diff)
  best_index = squeeze(argwhere(diffs == amin(abs(diffs))))
  print amin(abs(diffs))
  print 'Best a_prior = %d, b_prior = %d' % (best_index[0], best_index[1])

def q4d():
  # We want the ratio a/b knowing that x = 14
  c1, c2 = 47, 1
  print "CDF at 14 =", binom.cdf(14, 20, 0.5)
  print "c1/(c1+c2) =", c1*1.0/(c2+c1)

  # xx = linspace(0,1,1000)
  # beta_prior = beta.cdf(xx, a_prior, b_prior, loc=0, scale=1)
  # beta_posterior = beta.cdf(xx, a_posterior, b_posterior, loc=0,
   ↪  scale=1)
  # plot(xx, beta_prior)
  # title('prior, a = %d, b = %d'%(a_prior, b_prior))
  # figure()
  # plot(xx, beta_posterior)
  # title('posterior, a = %d, b = %d'%(a_posterior, b_posterior))
  # show()

def q5():
  # Generate 10000 w based on posterior distribution
  # Calculate the loss for each action and do histogram
  mean = 4 # 2.44
  std = 2 # 1.56
  n = 10000
  ws = normal(mean, std, n)
  avgloss = zeros(5)
  for i,(l,u) in enumerate([(-5,-3), (-3,-1), (-1,1), (1,3), (3,5)]):
    # Find the non-zeros loss region samples
    loss = 0.0
    loss += sum(abs(ws[ws<l] - l))
    loss += sum(abs(ws[ws>=u] - u))
    avgloss[i] = loss / n
```

```
  print avgloss
  plot(avgloss,'o')
  show()

def main():
  q5()

if __name__ == '__main__':
  main()
```