

User Manual for PCATS

June 25, 2023

Example 1. Step-by-step instruction of calling PCATS using a simple example of a binary treatment

This simple example provides step-by-step instructions on the use of the PCATS for estimating ATE and PrTE for the non-adaptive treatment setting. The example data are simulated from

- $X \sim N(0, 1)$
- $A|X \sim \text{Bernoulli}(\text{expit}(-0.2 + 3X))$
- $Y|A, X \sim N(X + 5A, 1)$

For the non-adaptive treatment, `staticGP` function is used. With the continuous outcome Y in this simulation data, `outcome.type` is set to “Continuous”. `tr.type` gives the type of the treatment. In this example, the treatment A is a binary indicator variable, so `tr.type` is set to “Discrete”. Since the example is not related to time, we create a “Time” variable with a fixed value 0 in the dataset. And time is set to “Time”. Users do not need to specify `time.value` for the case. The parameter `x.explanatory` specifies the prognostic variables \mathbf{W} and `x.confounding` specifies the confounders \mathbf{V} . The categorical variables in \mathbf{W} and \mathbf{V} should be specified in `x.categorical`. Users can define a link function by `outcome.link` whose default value is “identity”. We use the default settings for the number of burn-in MCMC samples (`burn.num`) and the number of MCMC samples after burn-in (`mcmc.num`), which are 500 for both parameters. By default, estimates of averaged treatment effect and potential outcomes are reported. Users can also request PrTE for one or a list of supplied number(s) specified by `c.margin`. In the example, we set `c.margin` to “0,1,2” to print the corresponding treatment effect $PrTE_0$, $PrTE_1$ and $PrTE_2$, separately. For a continuous outcome, users can choose either GP or BART. For a binary outcome, only BART method is currently available.

1. R code

```

1 library(pcatsAPIclientR)
2
3 download.file("https://github.com/pcats-api/pcats_api_examples
  /raw/main/data/example1.csv", destfile="example1.csv")
4
5 jobid <- pcatsAPIclientR::staticGP(datafile="example1.csv",
6                                   outcome="Y",
7                                   treatment="A",
8                                   time="Time",
9                                   x.explanatory="X",
10                                  x.confounding="X",
11                                  burn.num=500, mcmc.num=500,
12                                  outcome.type="Continuous",
13                                  method="GP",
14                                  tr.type="Discrete",
15                                  c.margin="0")
16
17 cat(paste0("JobID: ", jobid, "\n"))

```

Listing 1: R code of Example 1 to start computation.

```

1 JobID: e2db813e-89a2-4a35-bfa3-d34b5b2b0d0d

```

Listing 2: Output of R code of Example 1.

To make a request from R, we may submit the R code in Listing 1. After submitting the request, the system-generated id will be assigned to the job. Users can extract the job id from the response of the function `staticGP` and use the following code to wait for the job completion. In this example, the job id is "e2db813e-89a2-4a35-bfa3-d34b5b2b0d0d". If the job has been successfully done, the function will return with `status="Done"`.

```

1 status <- pcatsAPIclientR::wait_for_result(jobid)

```

Listing 3: R code of Example 1 to wait for job completion.

To retrieve a job status without waiting for the completion (i.e., polling), one may use the following code.

```

1 status <- pcatsAPIclientR::job_status(jobid)

```

Listing 4: R code of Example 1 to poll job status.

After the job is successfully completed, the analysis results can be extracted from the API response by using the code:

```

1 if (status=="Done") {
2   cat(pcatsAPIclientR::printgp(jobid))
3 }

```

Listing 5: R code of Example 1 to print results.

2. Python code

For Example 1, the following code calls PCATS API from a Python shell.

```

1 import pcats_api_client as pcats_api
2 import requests
3
4 r = requests.get("https://github.com/pcats-api/
   pcats_api_examples/raw/main/data/example1.csv")
5
6 with open("example1.csv", 'wb') as f:
7     f.write(r.content)
8
9 jobid=pcats_api.staticgp(datafile="example1.csv",
10     outcome="Y",
11     treatment="A",
12     time="Time",
13     x_explanatory="X",
14     x_confounding="X",
15     burn_num=500,
16     mcmc_num=500,
17     outcome_type="Continuous",
18     method="GP",
19     tr_type="Discrete",
20     c_margin="0")
21
22 print("JobID: {}".format(jobid))
23
24 status=pcats_api.wait_for_result(jobid)
25
26 if status=="Done":
27     print(pcats_api.printgp(jobid))
28 else:
29     print("Error")

```

Listing 6: Python code of Example 1.

```

1 Average treatment effect:
2 Contrast Estimation SD LB UB PrTE(c=0)
3 A=0 - A=1 -5.048 0.198 -5.415 -4.662 0
4
5 Potential outcomes:
6 A Estimation SD LB UB
7 0 -0.122 0.089 -0.296 0.052
8 1 4.926 0.112 4.708 5.140
9
10 Plot URL: https://pcats.research.cchmc.org/api/job/e2db813e-89a2
   -4a35-bfa3-d34b5b2b0d0d/plot
11
12 Plot Potential URL: https://pcats.research.cchmc.org/api/job/
   e2db813e-89a2-4a35-bfa3-d34b5b2b0d0d/plot/Potential

```

Listing 7: GP result output of Example 1.

The output in Listing 7 presents the result of causal treatment effect estimation. The average treatment effects are generated based on the user-specified treatment values. If the treatment variable is a factor, the sample potential outcomes for each level of the treatment variable are given. The pairwise comparisons of the treatment effects for all treatment groups are calculated. The first table shows the estimated ATE with standard deviation (SD) and the 95%

credible interval. The estimate of ATE is -5.05 with the corresponding 95% equal tail credible interval (CI) of (-5.42, -4.66), where the simulation true ATE is -5. And PrTE for $c.\text{margin}=0$ is provided in the column of $\text{PrTE}(c=0)$. The second table presents the estimated average potential outcomes by treatment groups. It reports that the expected mean and its standard deviation for the potential outcomes are -0.12 and 0.09 for $A = 0$ and 4.93 and 0.11 for $A = 1$. The PCATS also provides several interactive figures. Users can access these figures by using Firefox, Microsoft Edge, Chrome, or Safari. Currently, Internet Explorer does not support these figures. The histograms of MCMC posterior estimates of ATE (Figure 1) and the average of the estimated potential outcomes (Figure 2) are presented in the URLs, which could be copied from Listing 7. The figures are interactive. By hovering the pointers, users are provided with the corresponding posterior estimates of the average treatment effect and potential outcomes. Users can hide any traces in the plot by clicking on the legend on the right side. In each histogram, users could obtain a more detailed report of the results by hovering the mouse over the specific histogram. For example, if a user is interested in finding if the probability of $Y(0) - Y(1)$ is less than -5, then (s)he could click on the pink histogram as in Figure 1, then move the mouse to the bar at $ATE = -5$. The user could also save the entire figure into a PNG file which can be used for sharing the results.

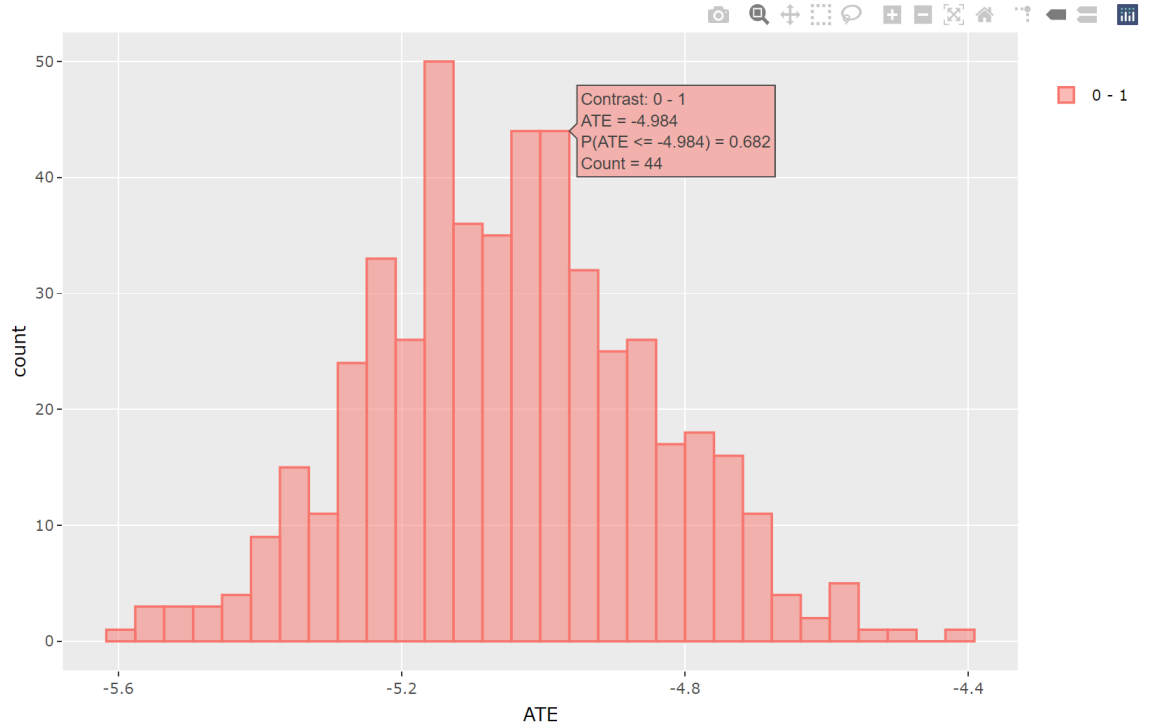


Figure 1: The histogram of the estimated average treatment effect

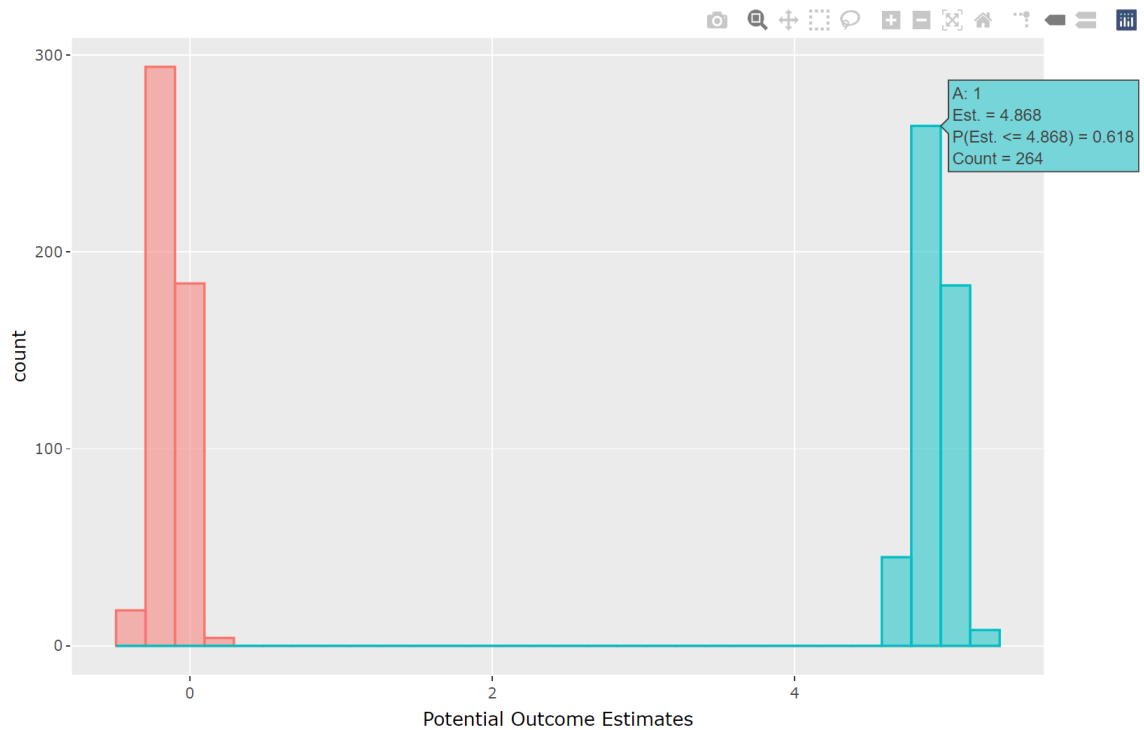


Figure 2: The histogram of the average of the estimated potential outcomes

Users could also set method to “BART” in the staticGP function. The BART results are shown in Listing 8. Comparing to Listing 7, we can see the results of GP and BART methods are quite similar in this case except GP provides more narrow confidence intervals in this example.

```

1 Average treatment effect:
2 Contrast Estimation SD LB UB PrTE(c=0)
3 A=0 - A=1 -5.118 0.286 -5.743 -4.592 0
4
5 Potential outcomes:
6 A Estimation SD LB UB
7 0 -0.166 0.166 -0.464 0.193
8 1 4.951 0.231 4.522 5.438
9
10 Plot URL: https://pcats.research.cchmc.org/api/job/91a699dc-ed89
    -4213-beb3-1ea516f2292c/plot
11
12 Plot Potential URL: https://pcats.research.cchmc.org/api/job/91
    a699dc-ed89-4213-beb3-1ea516f2292c/plot/Potential

```

Listing 8: BART Result output of example 1

Example 2. Estimating CATE for a non-adaptive binary-continuous treatment

The PCATS can estimate the conditional average treatment effect (CATE) which captures heterogeneity of a treatment effect varying by treatment effect modifier(s). In this example, we consider a treatment variable A , and the treatment effect of A varies by gender. The data are generated as the following

- $X \sim N(0, 1)$
- $Z \sim N(0, 1)$
- $\text{Gender} \sim \text{Bernoulli}(0.5)$
- $A|X, Z \sim \text{Bernoulli}(\text{expit}(-0.2 + 3X))$
- $Y|A, X, Z, \text{Gender} \sim N(X + 5A + 3Z + 2A * Z + 2A * \text{Gender}, 1)$

In order to estimate CATE, users can specify the treatment effect modifier(s) with the treatment variable in `tr.hte` in the `staticGP` function. When specified, the corresponding interaction term(s) will be added into mean function of the PCATS model. In this example, by setting `tr.hte = "Gender,Z"`, the interaction terms of Gender by treatment A and Z by treatment A are added into the mean function of the PCATS model.

For continuous treatment, PCATS calculates the sample potential outcomes based on the user-specified values of the variable which are given by `tr.values`. We request PCATS to report the treatment effects at the given values of the continuous treatment. The R code and Python example codes are provided as below.

1. R code

```
1 library(pcatsAPIclientR)
2
3 download.file("https://github.com/pcats-api/pcats_api_examples
4 /raw/main/data/example2.csv", destfile="example2.csv")
5
6 jobid <- pcatsAPIclientR::staticGP(datafile="example2.csv",
7                                   outcome="Y",
8                                   treatment="A",
9                                   time="Time",
10                                  x.explanatory="X,Z",
11                                  x.confounding="X",
12                                  tr.hte="Gender,Z",
13                                  tr.type="Discrete",
14                                  burn.num=500,
15                                  mcmc.num=500,
16                                  outcome.type="Continuous",
17                                  method="GP",
18                                  x.categorical="Gender")
19
20 status <- pcatsAPIclientR::wait_for_result(jobid)
```

```

21 if (status=="Done") {
22   cat(pcatsAPIclientR::printgp(jobid))
23 }

```

Listing 9: R code of Example 2.

The conditional average treatment effects of the sample data are estimated for the user-specified treatment sub-groups. `control.tr` defines the reference group and `treat.tr` defines the treatment group compared to the reference group. Then, the estimated CATE and PrCTE and their 95% confidence intervals at each level of “Gender” will be reported.

```

1 # CATE
2 jobidcate <- pcatsAPIclientR::staticGP.cate(jobid=jobid,
3                                             x="Gender",
4                                             control.tr="1",
5                                             treat.tr="0",
6                                             c.margin="0")
7
8 status <- pcatsAPIclientR::wait_for_result(jobidcate)
9
10 if (status=="Done") {
11   cat(pcatsAPIclientR::printgp(jobidcate))
12 }

```

Listing 10: R code of Example 2 – CATE.

2. Python code

```

1 import pcats_api_client as pcats_api
2 import requests
3
4 r = requests.get("https://github.com/pcats-api/pcats_api_
5                 examples/raw/main/data/example2.csv")
6
7 with open("example2.csv", 'wb') as f:
8     f.write(r.content)
9
10 jobid=pcats_api.staticgp(datafile="example2.csv",
11                          outcome="Y",
12                          treatment="A",
13                          time="Time",
14                          x_explanatory="X,Z",
15                          x_confounding="X",
16                          tr_hte="Gender,Z",
17                          tr_type="Discrete",
18                          burn_num=500,
19                          mcmc_num=500,
20                          outcome_type="Continuous",
21                          method="GP",
22                          x_categorical="Gender")
23
24 print("JobID: {}".format(jobid))
25
26 status=pcats_api.wait_for_result(jobid)
27
28 if status=="Done":

```

```

28     print(pcats_api.printgp(jobid))
29 else:
30     print("Error")
31     exit()
32
33 #CATE
34
35 jobid_cate=pcats_api.staticgp_cate(jobid=jobid,
36     x="Gender",
37     control_tr="1",
38     treat_tr="0",
39     c_margin="0")
40
41 print("CATE JobID: {}".format(jobid_cate))
42
43 status=pcats_api.wait_for_result(jobid_cate)
44
45 if status=="Done":
46     print(pcats_api.printgp(jobid_cate))
47 else:
48     print("Error")

```

Listing 11: Python code of Example 2.

```

1 Average treatment effect:
2   Contrast Estimation    SD    LB    UB
3   A=0 - A=1      -6.243 0.184 -6.611 -5.9
4
5 Potential outcomes:
6   A Estimation    SD    LB    UB
7   0      0.014 0.095 -0.150 0.212
8   1      6.257 0.092  6.086 6.442
9
10 Plot URL:  https://pcats.research.cchmc.org/api/job/1f47521c-427f
    -487a-a671-c4e95253d043/plot
11
12 Plot Potential URL: https://pcats.research.cchmc.org/api/job/1
    f47521c-427f-487a-a671-c4e95253d043/plot/Potential

```

Listing 12: Result output of Example 2.

```

1 Conditional average treatment effect:
2   Constrast Gender Estimation    SD    LB    UB PrCTE(c=0)
3   0 - 1      0      -4.959 0.238 -5.397 -4.471      0
4   0 - 1      1      -7.315 0.217 -7.717 -6.901      0
5
6 Plot URL:  https://pcats.research.cchmc.org/api/job/f45cdbfa
    -7526-4857-bcd8-5757d266a74d/plot

```

Listing 13: Result output of Example 2 – CATE.

The output are presented in Listing 12. The first table shows the estimated ATE by the pairwise comparison for each level of the treatment. The second table presents the estimated potential outcomes for each level of the treatment. The table in Listing 13 shows the estimated CATEs of the user-specified treatment groups for each level of “Gender”. In this example, the simulation true values

of the treatment effect of A given $Z = 0$ were set as -5 for Gender= 0 and -7 for Gender= 1. The estimated PrCTEs for c.margin="0" by "Gender" group are shown in the last column. The interactive figures of the estimated CATEs in Figure 3 can be retrieved through the URL link shown at the bottom of the result in Listing 13. The violin plots show the kernel probability densities of the estimated CATEs, and the box plots show the medians of the estimated CATEs with the boxes indicating the interquartile ranges. The corresponding values are also shown in the hover text.

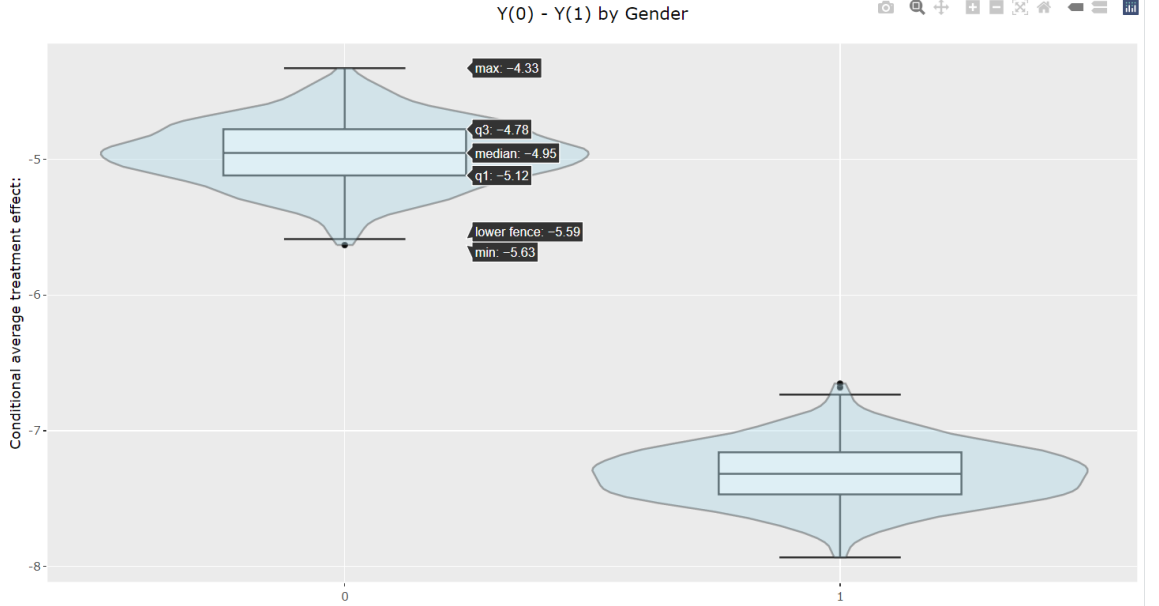


Figure 3: Estimates of the conditional average treatment effect by Gender

Example 3: Estimating ATE for adaptive binary treatment strategies

The dynamicGP function is designed to estimate the ATE for data with two time points. This example shows how to use the dynamicGP function. The simulation data were generated from the following simulation setting:

- $X \sim N(0, 1)$
- $A_1 \sim \text{Bernoulli}(0.5)$
- $L_1|A_1, X \sim N(0.25 + 0.3A_1 - 0.2X, 1)$
- $A_2|L_1, A_1, X \sim \text{Bernoulli}(\text{expit}(-0.2 - 0.38A_1 + L_1))$

- $Y|A_1, X, L_1, A_2 \sim N(-2 + 2.5A_1 + 3.5A_2 + 0.5A_1A_2 - 0.6L_1, sd = 2)$

The expectation of the potential outcomes $Y_i(a_1, a_2)$ can be derived as $E(Y(a_1, a_2)) = -2.15 + 2.32a_1 + 3.5a_2 + 0.5a_1a_2$. The R code and Python code are shown below:

1. R code

```

1 library(pcatsAPIclientR)
2
3 download.file("https://github.com/pcats-api/pcats_api_examples
  /raw/main/data/example3.csv", destfile="example3.csv")
4
5 jobid <- pcatsAPIclientR::dynamicGP(datafile="example3.csv",
6                                     stg1.outcome="L1",
7                                     stg1.treatment="A1",
8                                     stg1.time="Time1",
9                                     stg1.x.explanatory="X",
10                                    stg1.x.confounding="X",
11                                    stg1.outcome.type="
  Continuous",
12
13                                    stg2.outcome="Y",
14                                    stg2.treatment="A2",
15                                    stg2.time="Time2",
16                                    stg2.x.explanatory="X,L1",
17                                    stg2.x.confounding="X,L1",
18                                    stg2.outcome.type="
  Continuous",
19
20                                    burn.num=500,
21                                    mcmc.num=500,
22                                    stg1.tr.type="Discrete",
23                                    stg2.tr.type="Discrete",
24                                    method="GP")
25
26 status <- pcatsAPIclientR::wait_for_result(jobid)
27
28 if (status=="Done") {
29   cat(pcatsAPIclientR::printgp(jobid))
30 }

```

Listing 14: R code of Example 3.

2. Python code

```

1 import pcats_api_client as pcats_api
2 import requests
3
4 r = requests.get("https://github.com/pcats-api/
  pcats_api_examples/raw/main/data/example3.csv")
5
6 with open("example3.csv", 'wb') as f:
7   f.write(r.content)
8
9 jobid=pcats_api.dynamicgp(datafile="example3.csv",
10                            stg1_outcome="L1",
11                            stg1_treatment="A1",
12                            stg1_time="Time1",
13                            stg1_x_explanatory="X",
14                            stg1_x_confounding="X",

```

```

15     stg1_outcome_type="Continuous",
16     stg2_outcome="Y",
17     stg2_treatment="A2",
18     stg2_time="Time2",
19     stg2_x_explanatory="X,L1",
20     stg2_x_confounding="X,L1",
21     stg2_outcome_type="Continuous",
22     burn_num=500,
23     mcmc_num=500,
24     stg1_tr_type="Discrete",
25     stg2_tr_type="Discrete",
26     method="GP")
27
28 print("JobID: {}".format(jobid))
29
30 status=pcats_api.wait_for_result(jobid)
31
32 if status=="Done":
33     print(pcats_api.printgp(jobid))
34 else:
35     print("Error")

```

Listing 15: Python code of Example 3.

The results of GP and BART methods are shown in Listings 16 and 17. Both methods provide similar estimates. The statistical results of ATE and potential outcomes at two time points are shown in the output tables. Stage 1 shows the results of the first time point and Stage 2 shows the second time point. For Stage 1, the interpretation of the result is similar to that in non-adaptive treatment. For Stage 2, the ATE is the total effect. For example, the first row of ATEs given in Listing 16 is an estimate for $E(Y(a_1 = 0, a_2 = 0)) - E(Y(a_1 = 0, a_2 = 1))$. The estimated ATE of -3.8 is closed to the simulation true value of -3.5 and the 95% credible interval, $(-4.64, -2.82)$, covers the true value, as displayed in the histogram of Figure 4. The estimated potential outcomes are also accurate and its histogram is shown in Figure 5.

```

1 Stage 1:
2 Average treatment effect:
3   Contrast Estimation    SD    LB    UB
4   A1=0 - A1=1      -0.435 0.149 -0.743 -0.162
5
6 Potential outcomes:
7   A1 Estimation    SD    LB    UB
8   0      0.068 0.083 -0.106 0.220
9   1      0.502 0.070  0.365 0.637
10
11 Stage 2:
12 Average treatment effect:
13   Contrast Estimation    SD    LB    UB
14   A1=1 & A2=1 - A1=0 & A2=1      3.106 0.472  2.153  3.985
15   A1=1 & A2=1 - A1=0 & A2=0      6.893 0.421  6.075  7.640
16   A1=1 & A2=1 - A1=1 & A2=0      4.191 0.427  3.290  4.976
17   A1=0 & A2=1 - A1=0 & A2=0      3.787 0.468  2.821  4.635
18   A1=0 & A2=1 - A1=1 & A2=0      1.085 0.523  0.127  2.064
19   A1=0 & A2=0 - A1=1 & A2=0     -2.702 0.428 -3.494 -1.849

```

```

20
21 Potential outcomes:
22   A1 A2 Estimation      SD      LB      UB
23   1  1      4.467 0.228  4.043  4.937
24   0  1      1.360 0.358  0.625  2.035
25   0  0     -2.426 0.279 -2.915 -1.826
26   1  0      0.276 0.285 -0.323  0.795
27
28 Plot URL:  https://pcats.research.cchmc.org/api/job/3bf0e733-a381
              -4525-9a5a-5c8552b634e2/plot
29
30 Plot Potential URL: https://pcats.research.cchmc.org/api/job/3
                      bf0e733-a381-4525-9a5a-5c8552b634e2/plot/Potential

```

Listing 16: GP result output of Example 3.

```

1 Stage 1:
2 Average treatment effect:
3   Contrast Estimation      SD      LB      UB
4   A1=0 - A1=1      -0.449 0.17  -0.767 -0.115
5
6 Potential outcomes:
7   A1 Estimation      SD      LB      UB
8   0      0.053 0.118 -0.151 0.277
9   1      0.502 0.117  0.282 0.736
10
11 Stage 2:
12 Average treatment effect:
13   Contrast Estimation      SD      LB      UB
14   A1=1 & A2=1 - A1=0 & A2=1      3.107 0.530  2.063  4.110
15   A1=1 & A2=1 - A1=0 & A2=0      6.585 0.546  5.501  7.692
16   A1=1 & A2=1 - A1=1 & A2=0      4.207 0.519  3.136  5.232
17   A1=0 & A2=1 - A1=0 & A2=0      3.477 0.578  2.281  4.532
18   A1=0 & A2=1 - A1=1 & A2=0      1.099 0.612 -0.080  2.305
19   A1=0 & A2=0 - A1=1 & A2=0     -2.378 0.542 -3.421 -1.321
20
21 Potential outcomes:
22   A1 A2 Estimation      SD      LB      UB
23   1  1      4.389 0.354  3.652  5.076
24   0  1      1.282 0.445  0.424  2.165
25   0  0     -2.196 0.407 -3.014 -1.416
26   1  0      0.182 0.422 -0.683  1.016
27
28 Plot URL:  https://pcats.research.cchmc.org/api/job/5f0e0df5-8d0b-4
              c15-91d2-b2e0d8f193e8/plot
29
30 Plot Potential URL: https://pcats.research.cchmc.org/api/job/5
                      f0e0df5-8d0b-4c15-91d2-b2e0d8f193e8/plot/Potential

```

Listing 17: BART result output of Example 3.

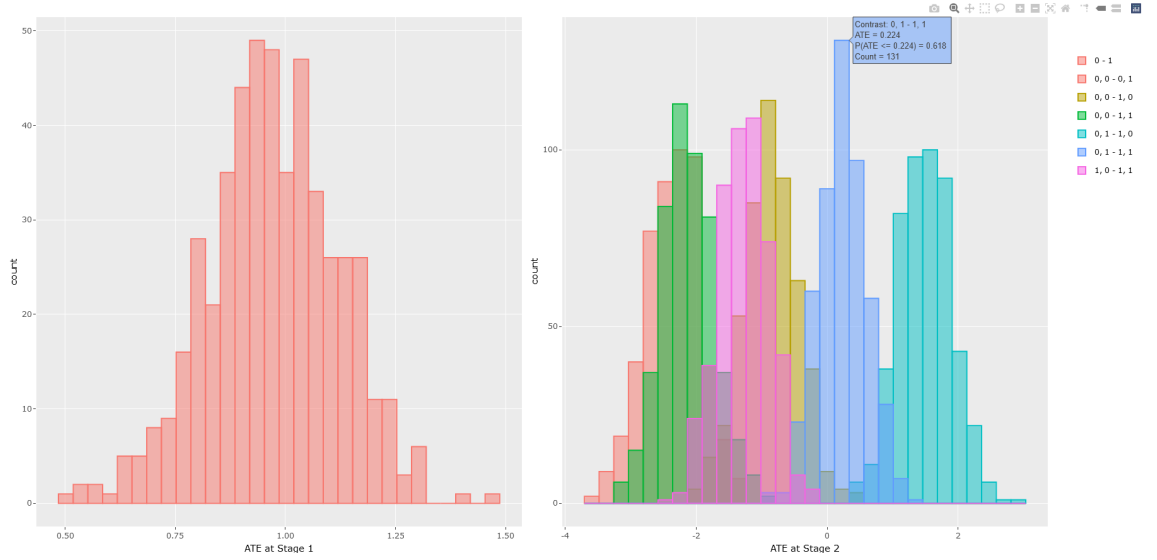


Figure 4: The histogram of the estimated average treatment effect at each stage.

Example 4: Estimating CATE for adaptive binary treatment strategies

This example presents users how to use the API to estimate the condition average treatment effect (CATE) for data with two time points. The simulation setting is as follows:

- $X \sim N(0, 1)$
- $M \sim \text{Bernoulli}(0.4)$
- $A_1 \sim \text{Bernoulli}(\text{expit}(0.3 - 0.5X - 0.4M))$
- $L_1 | A_1, X, M \sim N(0.75X - 0.75A_1 - 0.25A_1M + 0.5M, 1)$
- $A_2 | L_1, A_1, X, M \sim \text{Bernoulli}(\text{expit}(0.5X + 0.2A_1 - 0.05L_1 - 0.1L_1A_1 - 0.01L_1^2 - 0.2M))$
- $Y | A_1, X, M, L_1, A_2 \sim N(3 + 0.5A_1 + 0.4A_2M - L_1 - L_1^2 + 2A_2 - A_1A_2 + M, sd = 2)$

The R code and Python code are shown below:

1. R code

```
1 library(pcatsAPIclientR)
2
3 download.file("https://github.com/pcats-api/pcats_api_examples
  /raw/main/data/example4.csv", destfile="example4.csv")
```

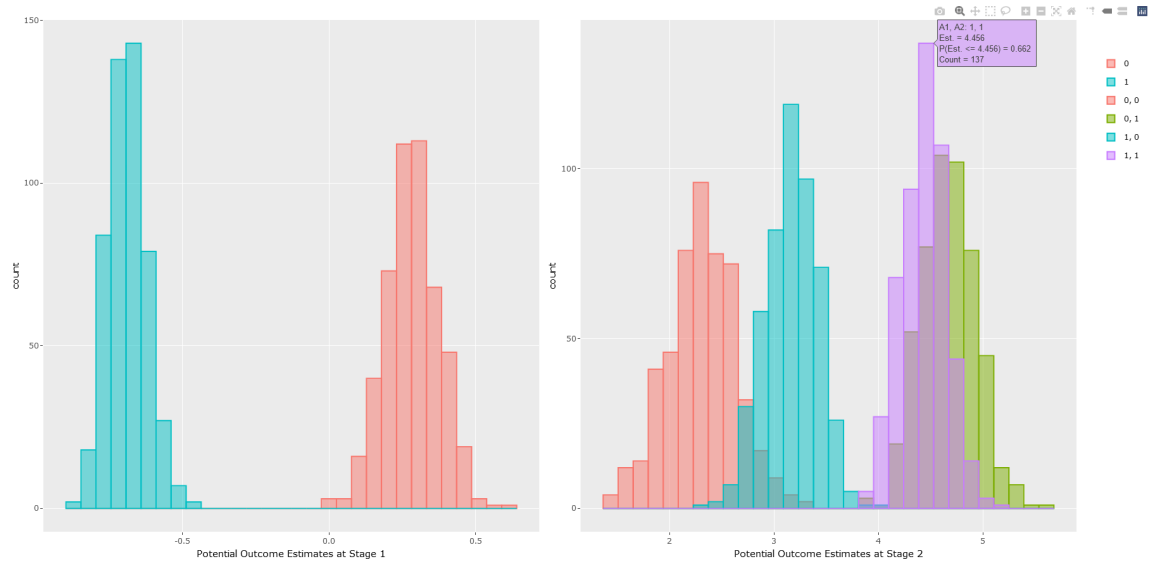


Figure 5: The histogram of the average of the estimated potential outcomes at each stage.

```

4
5 jobid <- pcatsAPIclientR::dynamicGP(datafile="example4.csv",
6                                     stg1.outcome="L1",
7                                     stg1.treatment="A1",
8                                     stg1.time="Time1",
9                                     stg1.x.explanatory="X,M",
10                                    stg1.x.confounding="X,M",
11                                    stg1.outcome.type="
12                                     Continuous",
13                                    stg1.tr.hte="M",
14                                    stg1.tr.type="Discrete",
15                                    stg2.outcome="Y",
16                                    stg2.treatment="A2",
17                                    stg2.time="Time2",
18                                    stg2.x.explanatory="X,L1,M",
19                                    stg2.x.confounding="X,L1,M",
20                                    stg2.outcome.type="
21                                     Continuous",
22                                    stg2.tr2.hte="M",
23                                    stg2.tr.type="Discrete",
24                                    burn.num=500,
25                                    mcmc.num=500,
26                                    x.categorical="M",
27                                    method="GP")
28
29 status <- pcatsAPIclientR::wait_for_result(jobid)
30
31 if (status=="Done") {

```

```

30   cat(pcatsAPIclientR::printgp(jobid))
31 }
32
33 jobidcate <- pcatsAPIclientR::dynamicGP.cate(
34   jobid=jobid,
35   x="M",
36   control.tr="0,0",
37   treat.tr="1,0")
38
39 status <- pcatsAPIclientR::wait_for_result(jobidcate)
40
41 if (status=="Done") {
42   cat(pcatsAPIclientR::printgp(jobidcate))
43 }

```

Listing 18: R code of Example 4.

2. Python code

```

1  import pcats_api_client as pcats_api
2  import requests
3
4  r = requests.get("https://github.com/pcats-api/
   pcats_api_examples/raw/main/data/example4.csv")
5
6  with open("example4.csv", 'wb') as f:
7      f.write(r.content)
8
9  jobid=pcats_api.dynamicgp(datafile="example4.csv",
10     stg1_outcome="L1",
11     stg1_treatment="A1",
12     stg1_time="Time1",
13     stg1_x_explanatory="X,M",
14     stg1_x_confounding="X,M",
15     stg1_outcome_type="Continuous",
16     stg1_tr_hte="M",
17     stg2_outcome="Y",
18     stg2_treatment="A2",
19     stg2_time="Time2",
20     stg2_x_explanatory="X,L1,M",
21     stg2_x_confounding="X,L1,M",
22     stg2_outcome_type="Continuous",
23     stg2_tr2_hte="M",
24     burn_num=500,
25     mcmc_num=500,
26     stg1_tr_type="Discrete",
27     stg2_tr_type="Discrete",
28     x_categorical="M",
29     method="GP")
30
31 print("JobID: {}".format(jobid))
32
33 status=pcats_api.wait_for_result(jobid)
34
35 if status=="Done":
36     print(pcats_api.printgp(jobid))
37 else:
38     print("Error")

```

```

39     exit()
40
41 #CATE
42
43 jobid_cate=pcats_api.dynamicgp_cate(jobid=jobid,
44     x="M",
45     control_tr="0,0",
46     treat_tr="1,0")
47
48 print("CATE JobID: {}".format(jobid_cate))
49
50 status=pcats_api.wait_for_result(jobid_cate)
51
52 if status=="Done":
53     print(pcats_api.printgp(jobid_cate))
54 else:
55     print("Error")

```

Listing 19: Python code of Example 4.

```

1 Stage 1:
2 Average treatment effect:
3     Contrast Estimation    SD    LB    UB
4 A1=0 - A1=1           0.971 0.149 0.663 1.243
5
6 Potential outcomes:
7 A1 Estimation    SD    LB    UB
8 0           0.284 0.091 0.115 0.466
9 1          -0.687 0.066 -0.817 -0.558
10
11 Stage 2:
12 Average treatment effect:
13     Contrast Estimation    SD    LB    UB
14 A1=1 & A2=0 - A1=0 & A2=0           0.851 0.439 0.046 1.736
15 A1=1 & A2=0 - A1=1 & A2=1          -1.271 0.383 -2.045 -0.555
16 A1=1 & A2=0 - A1=0 & A2=1          -1.494 0.405 -2.280 -0.756
17 A1=0 & A2=0 - A1=1 & A2=1          -2.122 0.400 -2.935 -1.316
18 A1=0 & A2=0 - A1=0 & A2=1          -2.346 0.453 -3.187 -1.380
19 A1=1 & A2=1 - A1=0 & A2=1          -0.224 0.392 -0.918 0.594
20
21 Potential outcomes:
22 A1 A2 Estimation    SD    LB    UB
23 1 0           3.164 0.250 2.705 3.661
24 0 0           2.313 0.334 1.650 2.957
25 1 1           4.435 0.213 4.040 4.852
26 0 1           4.659 0.282 4.131 5.250
27
28 Plot URL: https://pcats.research.cchmc.org/api/job/f6eb3598-95c1-4
a4c-8696-70b6e24042d5/plot
29
30 Plot Potential URL: https://pcats.research.cchmc.org/api/job/
f6eb3598-95c1-4a4c-8696-70b6e24042d5/plot/Potential
31
32 Conditional average treatment effect:
33     Constrast M Estimation    SD    LB    UB
34 A1=1 & A2=0 - A1=0 & A2=0 0           0.599 0.449 -0.295 1.439
35 A1=1 & A2=0 - A1=0 & A2=0 1           1.179 0.479 0.358 2.200

```


36 Plot URL: <https://pcats.research.cchmc.org/api/job/ca4bf80a-482c-47f1-afa9-6e8d2977b6d9/plot>
 37

Listing 20: Result output of Example 4.

At Stage 1, the simulation true value of the sample ATE is 0.86, the true stage-2 sample ATEs of $Y(0, 0) - Y(0, 1)$, $Y(0, 0) - Y(1, 0)$, $Y(0, 0) - Y(1, 1)$, $Y(0, 1) - Y(1, 0)$, $Y(0, 1) - Y(1, 1)$, and $Y(1, 0) - Y(1, 1)$ are about -2.17, -1.03, -2.20, 1.14, -0.03, and -1.17 respectively. The 95% credible intervals correctly cover the true expected potential outcomes.

By the simulation design, the true sample CATEs of $(Y(1, 0) - Y(0, 0)|M = 0)$ and $(Y(1, 0) - Y(0, 0)|M = 1)$ are 0.68 and 1.49. The last table shows the estimated CATE for each level of M , and all 95% credible intervals cover the true values.

Example 5: Binary outcome at stage 1 and continuous outcome at stage 2

This example shows how to use the API to estimate ATE for data with two different types of outcome at time points. Since the intermediate outcome L_1 is binary, a BART method is used for this example.

- $U_0 \sim \text{Bernoulli}(0.4)$
- $A_1 \sim \text{Bernoulli}(0.5)$
- $L_1|A_1, U_0 \sim \text{Bernoulli}(\text{expit}(0.25 + 0.3A_1 - 0.2U_0 - 0.05A_1U_0))$
- $A_2|L_1, A_1, U_0 \sim \text{Bernoulli}(\text{expit}(0.4 + 0.5A_1 - 0.3L_1 - 0.4A_1L_1))$
- $Y|A_1, U_0, L_1, A_2 \sim N(2.5 - 0.5A_1 - 0.75A_2 + 0.2A_1A_2 - U_0, sd = 0.2)$

The expectation of Y can be calculated by $E(Y(a_1, a_2)) = 2.1 - 0.5a_1 - 0.75a_2 + 0.2a_1a_2$. The R code and Python code are listed below:

1. R code

```
1 library(pcatsAPIclientR)
2
3 download.file("https://github.com/pcats-api/pcats_api_examples
  /raw/main/data/example5.csv", destfile="example5.csv")
4
5 jobid <- pcatsAPIclientR::dynamicGP(datafile="example5.csv",
6                                     stg1.outcome="L1",
7                                     stg1.treatment="A1",
8                                     stg1.time="Time1",
9                                     stg1.x.explanatory="U0",
10                                    stg1.x.confounding="U0",
11                                    Discrete",
12                                    stg1.tr.type="Discrete",
```

```

13         stg2.outcome="Y",
14         stg2.treatment="A2",
15         stg2.time="Time2",
16         stg2.x.explanatory="U0,L1"
17     ,
18         stg2.x.confounding="U0,L1"
19     ,
20         stg2.outcome.type="
Continuous",
21         stg2.tr.type="Discrete",
22         burn.num=500,
23         mcmc.num=500,
24         method="BART")
25
26 status <- pcatsAPIclientR::wait_for_result(jobid)
27
28 if (status=="Done") {
29     cat(pcatsAPIclientR::printgp(jobid))
30 }

```

Listing 21: R code of Example 5.

2. Python code

```

1 import pcats_api_client as pcats_api
2 import requests
3
4 r = requests.get("https://github.com/pcats-api/
pcats_api_examples/raw/main/data/example5.csv")
5
6 with open("example5.csv", 'wb') as f:
7     f.write(r.content)
8
9 jobid=pcats_api.dynamicgp(datafile="example5.csv",
10     stg1_outcome="L1",
11     stg1_treatment="A1",
12     stg1_time="Time1",
13     stg1_x_explanatory="U0",
14     stg1_x_confounding="U0",
15     stg1_outcome_type="Discrete",
16     stg1_tr_type="Discrete",
17     stg2_outcome="Y",
18     stg2_treatment="A2",
19     stg2_time="Time1",
20     stg2_x_explanatory="U0,L1",
21     stg2_x_confounding="U0,L1",
22     stg2_outcome_type="Continuous",
23     stg2_tr_type="Discrete",
24     burn_num=500,
25     mcmc_num=500,
26     method="BART")
27
28 print("JobID: {}".format(jobid))
29
30 status=pcats_api.wait_for_result(jobid)
31
32 if status=="Done":
33     print(pcats_api.printgp(jobid))

```

```

34 else:
35     print("Error")

```

Listing 22: Python code of Example 5.

```

1 Stage 1:
2 Average treatment effect:
3     Contrast Estimation SD LB UB
4 A1=0 - A1=1      0.064 0.05 -0.03 0.16
5
6 Potential outcomes:
7     A1 Estimation SD LB UB
8 0      0.484 0.037 0.410 0.551
9 1      0.421 0.035 0.347 0.483
10
11 Stage 2:
12 Average treatment effect:
13     Contrast Estimation SD LB UB
14 A1=0 & A2=1 - A1=1 & A2=1      0.281 0.040 0.206 0.357
15 A1=0 & A2=1 - A1=0 & A2=0     -0.777 0.045 -0.864 -0.686
16 A1=0 & A2=1 - A1=1 & A2=0     -0.248 0.048 -0.337 -0.150
17 A1=1 & A2=1 - A1=0 & A2=0     -1.058 0.045 -1.141 -0.970
18 A1=1 & A2=1 - A1=1 & A2=0     -0.529 0.046 -0.621 -0.434
19 A1=0 & A2=0 - A1=1 & A2=0      0.529 0.049 0.436 0.630
20
21 Potential outcomes:
22     A1 A2 Estimation SD LB UB
23 0 1      1.332 0.029 1.276 1.387
24 1 1      1.051 0.029 1.003 1.110
25 0 0      2.109 0.035 2.034 2.174
26 1 0      1.580 0.038 1.507 1.651
27
28 Plot URL: https://pcats.research.cchmc.org/api/job/ea716cc-d14a
           -492f-b6fa-f0cdec3aa8d3/plot
29
30 Plot Potential URL: https://pcats.research.cchmc.org/api/job/
           ea716cc-d14a-492f-b6fa-f0cdec3aa8d3/plot/Potential

```

Listing 23: Result output of Example 5.

The estimated results of $E(Y(a1 = 0, a2 = 0))$, $E(Y(a1 = 0, a2 = 1))$, $E(Y(a1 = 1, a2 = 0))$ and $E(Y(a1 = 1, a2 = 1))$ are 2.11, 1.33, 1.58 and 1.05, which are close to the true values 2.1, 1.35, 1.6 and 1.05.

Example 6. Estimating ATE for bounded data

The PCATS can handle the outcome bounded by the minimum and maximum limits by using the GP method. For this type of outcome, users need to set the option `outcome.bound_censor` as “bounded” and specify the lower and upper bounds in `outcome.lb` and `outcome.ub`. Because the original BART method cannot handle the bounded data, the PCATS will fit the data without considering the bounds if the BART method is chosen. This example is similar to Example 1 except that the outcome Y has the minimum value of zero. For comparison, results are generated using both GP and BART methods.

```

1 library(pcatsAPIclientR)
2
3 download.file("https://github.com/pcats-api/pcats_api_examples/raw/
  main/data/example6.csv", destfile="example6.csv")
4
5 #GP
6 jobid <- pcatsAPIclientR::staticGP(datafile="example6.csv",
7                                   outcome="Y",
8                                   treatment="A",
9                                   time="Time",
10                                  x.explanatory="X",
11                                  x.confounding="X",
12                                  burn.num=500,
13                                  mcmc.num=500,
14                                  outcome.type="Continuous",
15                                  method="GP",
16                                  tr.type="Discrete",
17                                  outcome.lb=0,
18                                  outcome.ub="inf",
19                                  outcome.bound_censor="bounded")
20
21 status <- pcatsAPIclientR::wait_for_result(jobid)
22
23 if (status=="Done") {
24   cat(pcatsAPIclientR::printgp(jobid))
25 }
26
27 #BART
28 jobid <- pcatsAPIclientR::staticGP(datafile="example6.csv",
29                                   outcome="Y",
30                                   treatment="A",
31                                   time="Time",
32                                   x.explanatory="X",
33                                   x.confounding="X",
34                                   burn.num=500,
35                                   mcmc.num=500,
36                                   outcome.type="Continuous",
37                                   method="BART")
38
39 status <- pcatsAPIclientR::wait_for_result(jobid)
40
41 if (status=="Done") {
42   cat(pcatsAPIclientR::printgp(jobid))
43 }

```

Listing 24: R code of Example 6.

```

1 Average treatment effect:
2   Contrast Estimation   SD    LB    UB
3   A=0 - A=1      -4.841 0.166 -5.142 -4.501
4
5 Potential outcomes:
6   A Estimation   SD    LB    UB
7   0      0.426 0.075 0.280 0.572
8   1      5.268 0.095 5.081 5.443
9
10 Plot URL:  https://pcats.research.cchmc.org/api/job/1f44596c
             -2940-4312-8118-91e174bae520/plot

```

```

11
12 Plot Potential URL: https://pcats.research.cchmc.org/api/job/1
    f44596c-2940-4312-8118-91e174bae520/plot/Potential

```

Listing 25: GP result output of Example 6.

```

1 Average treatment effect:
2   Contrast Estimation    SD    LB    UB
3   A=0 - A=1      -4.584 0.216 -5.003 -4.179
4
5 Potential outcomes:
6   A Estimation    SD    LB    UB
7   0      0.501 0.122 0.248 0.719
8   1      5.085 0.168 4.741 5.395
9
10 Plot URL: https://pcats.research.cchmc.org/api/job/db78434a-9842-4
    ef5-ad22-b43c5fee57ce/plot
11
12 Plot Potential URL: https://pcats.research.cchmc.org/api/job/
    db78434a-9842-4ef5-ad22-b43c5fee57ce/plot/Potential

```

Listing 26: BART result output of Example 6.

Listing 25 shows that the GP results in the posterior mean of -4.84 and the credible interval of (-5.14, -4.50), and Listing 26 shows that the BART results in the posterior mean of -4.58 and the credible interval of (-5.00, -4.18). For the estimation of the true ATE of -5, the GP works better than the BART resulting in the underestimation. These results show that reflecting the bounded feature in the outcome is important in making a correct inference.

Example 7. Estimating ATE for data with multiple imputation for missing values

For generating a simulated incomplete dataset, we randomly selected 50% of records of the dataset in Example 1 and let the predictor X be missing from the selected records. We used the MICE package to get five imputed datasets which are saved in the `example8_midata.csv` file. By running the code in Listing 27 or Listing 28, we have the results shown in Listing 29 that contains the estimated ATE and potential outcomes for each of five imputed datasets, and the combined results over all five imputations.

1. R code

```

1 library(pcatsAPIclientR)
2
3 download.file("https://github.com/pcats-api/pcats_api_examples
    /raw/main/data/example7.csv", destfile="example7.csv")
4 download.file("https://github.com/pcats-api/pcats_api_examples
    /raw/main/data/example7_midata.csv", destfile="example7_
    midata.csv")
5
6 jobid <- pcatsAPIclientR::staticGP(datafile="example7.csv",

```

```

7         outcome="Y",
8         treatment="A",
9         time="Time",
10        x.explanatory="X",
11        x.confounding="X",
12        burn.num=500,
13        mcmc.num=500,
14        outcome.type="Continuous",
15        method="GP",
16        tr.type="Discrete",
17        mi.datafile="example7_
          midata.csv")
18
19 status <- pcatsAPIclientR::wait_for_result(jobid)
20
21 if (status=="Done") {
22     cat(pcatsAPIclientR::printgp(jobid))
23 }

```

Listing 27: R code of Example 8.

2. Python code

```

1 import pcats_api_client as pcats_api
2 import requests
3
4 r = requests.get("https://github.com/pcats-api/
5 pcats_api_examples/raw/main/data/example7.csv")
6
7 with open("example7.csv", 'wb') as f:
8     f.write(r.content)
9
10 r = requests.get("https://github.com/pcats-api/
11 pcats_api_examples/raw/main/data/example7_midata.csv")
12
13 with open("example7_midata.csv", 'wb') as f:
14     f.write(r.content)
15
16 jobid=pcats_api.staticgp(datafile="example7.csv",
17 outcome="Y",
18 treatment="A",
19 x_explanatory="X",
20 x_confounding="X",
21 burn_num=500,
22 mcmc_num=500,
23 outcome_type="Continuous",
24 tr_type="Discrete",
25 method="GP",
26 mi_datafile="example7_midata.csv")
27
28 print("JobID: {}".format(jobid))
29
30 status=pcats_api.wait_for_result(jobid)
31
32 if status=="Done":
33     print(pcats_api.printgp(jobid))
34 else:

```

```
33 print("Error")
```

Listing 28: Python code of Example 7.

```

1 Average treatment effect:
2 Imputation Contrast Estimation SD LB UB
3 1 A=0 - A=1 -5.097 0.344 -5.742 -4.437
4 2 A=0 - A=1 -5.098 0.346 -5.743 -4.424
5 3 A=0 - A=1 -5.098 0.344 -5.764 -4.443
6 4 A=0 - A=1 -5.098 0.344 -5.773 -4.452
7 5 A=0 - A=1 -5.097 0.344 -5.729 -4.407
8 Combined A=0 - A=1 -5.098 0.344 -5.743 -4.415
9
10 Potential outcomes:
11 Imputation A Estimation SD LB UB
12 1 0 -0.109 0.152 -0.380 0.216
13 2 0 -0.099 0.152 -0.387 0.217
14 3 0 -0.158 0.144 -0.427 0.144
15 4 0 -0.150 0.148 -0.423 0.156
16 5 0 -0.092 0.154 -0.381 0.230
17 1 1 4.987 0.201 4.613 5.381
18 2 1 4.999 0.202 4.602 5.388
19 3 1 4.940 0.208 4.531 5.350
20 4 1 4.948 0.206 4.558 5.360
21 5 1 5.006 0.200 4.632 5.401
22 Combined 0 -0.122 0.152 -0.411 0.197
23 Combined 1 4.976 0.205 4.584 5.382
24
25 Plot URL: https://pcats.research.cchmc.org/api/job/c0091916-9b9e
-476e-9ca6-1d76d6dece22/plot
26
27 Plot Potential URL: https://pcats.research.cchmc.org/api/job/
c0091916-9b9e-476e-9ca6-1d76d6dece22/plot/Potential

```

Listing 29: Result output of Example 7.