

```

1  #!/usr/bin/env lua
2  --
3  --
4  --
5  --
6  --
7  --
8  --
9  --
10 --
11
12 local your, our={}, {b4={}, help=[[
13 peek.lua (OPTIONS)
14 (c)2022 Tim Menzies, MIT license
15 Understand N items after log(N) probes, or less.
16
17 -file      ../..data/auto93.csv
18 -best      .5
19 -help      false
20 -dull      .35
21 -rest      3
22 -seed      10019
23 -rnd       1.2f
24 -task      ~
25 -p         2]]}
26
27 for k, _ in pairs(_ENV) do our.b4[k] = k end
28 local any, asserts, cells, copy, fmt, go, id, main, many, map, new, o, push
29 local rand, randi, rnd, rogues, rows, same, settings, slots, sort, thing, things
30 local COLS, EG, EGS, NUM, RANGE, SYM
31 local class= function(t, new)
32   t.__index=t
33   return setmetatable(t, {_call=function(_, ...) return t.new(...) end}) end
34
35 -- Copyright (c) 2022, Tim Menzies
36 --
37 -- Redistribution and use in source and binary forms, with or without
38 -- modification, are permitted provided that the following conditions are met:
39 --
40 --   1. Redistributions of source code must retain the above copyright notice,
41 --   this list of conditions and the following disclaimer.
42 --   2. Redistributions in binary form must reproduce the above copyright
43 --   notice, this list of conditions and the following disclaimer in the
44 --   documentation and/or other materials provided with the distribution.
45 --
46 -- THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS
47 -- IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO,
48 -- THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR
49 -- PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR
50 -- CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL,
51 -- EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO,
52 -- PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR
53 -- PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF
54 -- LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING
55 -- NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS
56 -- SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE
57 -----
58
59
60
61
62
63
64 COLS=class{}
65 function COLS.new(t, i, where, now)
66   print("colsnew", o(t))
67   i = new({all={}, x={}, y={}}, COLS)
68   for at, s in pairs(t) do
69     print("cadd", at, s)
70     now = push(i.all, (s:find"^[A-Z]" and NUM or SYM) (at, s))
71     if not s:find"." then
72       push((s:find"-" or s:find"+") and i.y or i.x, now) end end
73   return i end
74
75 function COLS.__tostring(i, txt)
76   function txt(c) return c.txt end
77   return fmt("COLS:[all %s x %s y %s]", o(i.all, txt), o(i.x, txt), o(i.y, txt)) end
78
79 function COLS.add(i, t, add)
80   return map(i.all, function(col) col:add(t[i.at]) return x end) end
81
82 function COLS.better(i, row1, row2)
83   local s1, s2, e, n, a, b = 0, 0, 10, #i.y
84   for _, col in pairs(i.y) do
85     a = col:norm(row1.has[col.at])
86     b = col:norm(row2.has[col.at])
87     s1 = s1 - e^(col.w * (a-b)/n)
88     s2 = s2 - e^(col.w * (b-a)/n) end
89   return s1/n < s2/n end
90
91 -----
92 EG=class{}
93 function EG.new(t) return new({has=t, id=id()}, EG) end
94 function EG.__tostring(i) return fmt("EG%s%", i.id, o(i.has)) end
95
96 EGS=class{}
97 function EGS.new() return new({rows={}, cols=nil}, EGS) end
98
99 function EGS.__tostring(i) return fmt("EGS[#rows %s cols %s]", #i.rows, i.cols) end
100
101 function EGS.add(i, row)
102   print("egadd", o(row))
103   row = row.has and row.has or row
104   if i.cols then push(i.rows, EG(i.cols:add(row))) else i.cols=COLS(row) end end
105
106 function EGS.bestRest(i)
107   local best, rest, tmp, bests, restsFraction = {}, {}, {}
108   i.rows = sort(i.rows, function(a, b) return i.cols:better(a, b) end)
109   bests = (#i.rows)^your.best
110   restsFraction = (bests * your.rest)/(#i.rows - bests)
111   for j, x in pairs(i.rows) do
112     if j <= bests then push(best, x)
113     elseif rand() < restsFraction then push(rest, x) end end
114   return best, rest end
115
116 function EGS.clone(i, inits, j)
117   j = EGS()
118   print("clone", o(map(i.cols.all, function(col) return col.txt end)))
119   j:add(map(i.cols.all, function(col) return col.txt end))
120   for _, x in pairs(inits or {}) do j:add(x) end
121   return j end
122
123 function EGS.file(i, f) for row in rows(f) do i:add(row) end; return i end
124
125 function EGS.mid(cols)
126   return map(cols or i.cols.all,
127     function(col) return col:mid() end) end
128
129 -----
130 NUM=class{}
131 function NUM.new(at, s, big)
132   big = math.huge
133   return new({lo=big, hi=-big, at=at or 0, txt=s or "",
134     n=0, mu=0, m2=0, sd=0,
135     w=(s or ""):find"-" and -1 or 1}, NUM) end
136
137 function NUM.__tostring(i)
138   return fmt("NUM[{:at %s :txt %s :lo %s :hi %s :mu %s :sd %s}",
139     i.at, i.txt, i.lo, i.hi, rnd(i.mu), rnd(i.sd)) end
140
141 function NUM.add(i, x, d)
142   if x==">" then
143     i.n = i.n+1
144     d = x - i.mu
145     i.mu = i.mu + d/i.n
146     i.m2 = i.m2 + d*(x-i.mu)
147     i.lo = math.min(x, i.lo); i.hi = math.max(x, i.hi) end
148   return x end
149
150 function NUM.div(i) return i.n < 2 and 0 or (i.m2/(i.n-1))^0.5 end
151
152 function NUM.mid(i) return i.mu end
153
154 function NUM.norm(i, x) return i.hi-i.lo < 1E-9 and 0 or (x-i.lo)/(i.hi-i.lo) end
155
156 function NUM.ranges(i, j, bests, rests)
157   local ranges, x, lo, hi, gap, tmp = {}
158   hi = math.max(i.hi, j.hi)
159   lo = math.min(i.lo, j.lo)
160   gap = (hi - lo)/your.bins
161   tmp = lo
162   for j=lo, hi, gap do push(ranges, RANGE(i, tmp, tmp+gap)); tmp= tmp+gap end
163   ranges[1].lo = -math.huge
164   ranges[#ranges].hi = math.huge
165   for _, pair in pairs({bests, "bests"}, {rests, "rests"}) do
166     for _, row in pairs(pair[1]) do
167       x = row.has[i.at]
168       if x=="?" then
169         ranges[(x - lo)//gap].stats:add(pair[2]) end end end end
170
171 -----
172 RANGE=class{}
173 function RANGE.new(col, lo, hi, stats)
174   return new({col=col, lo=lo, hi=hi or lo, ys=stats or SYM(), all={}, RANGE) end
175
176 function RANGE.__tostring(i)
177   return fmt("RANGE[{:col %s :lo %s :hi %s :ys %s}", i.col, i.lo, i.hi, o(i.ys)) end
178
179 -----
180 SYM=class{}
181 function SYM.new(at, s)
182   return new({at=at or 0, txt=s or "", has={}, n=0, most=0, mode=nil}, SYM) end
183
184 function SYM.__tostring(i)
185   return fmt("SYM[{:at %s :txt %s :mode %s :has %s}",
186     i.at, i.txt, i.mode, o(i.has)) end
187
188 function SYM.add(i, x)
189   if x=="?" then
190     i.n = i.n+1
191     i.has[x] = 1 + (i.has[x] or 0) end
192   return x end
193
194 function SYM.div(i)
195   e=0; for _, v in pairs(i.has) do e=e - v/i.n*math.log(v/i.n, 2) end; return e end
196
197 function SYM.mid(i, most, out)
198   most=-1
199   for x, n in pairs(i.has) do if n>most then out, most=x, n end end; return out end
200
201 function SYM.ranges(i, j, bests, rests)
202   local tmp, out, x = {}, {}
203   for _, pair in pairs({bests, "bests"}, {rests, "rests"}) do
204     for _, row in pairs(pair[1]) do
205       x = row.has[i.at]
206       if x=="?" then
207         tmp[x] = tmp[x] or SYM()
208         tmp[x]:add(pair[2]) end end end
209   for x, stats in pairs(tmp) do push(out, RANGE(i, x, x, stats)) end
210   return out end

```

```

208 -----
209 --
210 --
211 --
212 --
213 --
214 --
215 fmt = string.format
216 new = setmetatable
217 same = function(x,...) return x end
218
219 function asserts(test,msg)
220 msg=msg or ""
221 if test then return print("PASS:".msg) end
222 our.fails = our.fails+1
223 print("FAIL:".msg)
224 if your.Debug then assert(test,msg) end end
225
226 function copy(t, u)
227 if type(t)~="table" then return t end
228 u={};for k,v in pairs(t) do u[k]=copy(v) end;return new(u,getmetatable(t)) end
229
230 function id() our.id = 1+(our.id or 0); return our.id end
231
232 function many(t,n, u) u={};for j=1,n do push(u,any(t)) end; return u end
233
234 function map(t,f, u)
235 u={};for _,v in pairs(t) do u[1+#u]=(f or same)(v) end; return u end
236
237 function o(t,f, u,key)
238 key= function(k)
239 if t[k] then return fmt(":%s%s", k, rnd((f or same)(t[k]))) end end
240 u = #t>0 and map(map(t,f),rnd) or map(slots(t),key)
241 return {"..table.concat(u, " ").."}" end
242
243 function rand(lo,hi)
244 your.seed = (16807 * your.seed) % 2147483647
245 return (lo or 0) + ((hi or 1) - (lo or 0)) * your.seed / 2147483647 end
246
247 function randi(lo,hi) return math.floor(0.5 + rand(lo,hi)) end
248
249 function push(t,x) table.insert(t,x); return x end
250
251 function rnd(x)
252 return fmt(type(x)=="number" and x~=#x//1 and your.rnd or"%s",x) end
253
254 function rows(file, x)
255 file = io.input(file)
256 return function()
257 x=io.read(); if x then return things(x) else io.close(file) end end end
258
259 function main( defaults,tasks)
260 tasks = your.task=="all" and slots(go) or {your.task}
261 defaults=copy(your)
262 our.failures=0
263 for _,x in pairs(tasks) do
264 if type(our.go[x]) == "function" then our.go[x]() else print("?", x) end
265 your = copy(defaults) end
266 rogues()
267 return our.failures end
268
269 function rogues()
270 for k,_ in pairs(_ENV) do if not our.b4[k] then print("?",k) end end end
271
272 function settings(help, t)
273 t={}
274 help:gsub("\n [-](^%s+)[^n]*%s(^%s+)", function(slot, x)
275 for n,flag in ipairs(arg) do
276 if flag:sub(1,1)=="-" and slot:match("^"..flag:sub(2).."%.*")
277 then x=x=="false" and "true" or x=="true" and "false" or arg[n+1] end end
278 t[slot] = thing(x) end
279 if t.help then print(t.help) end
280 return t end
281
282 function slots(t,u) u={};for x,_ in pairs(t) do u[1+#u]=x end;return sort(u) end
283
284 function sort(t,f) table.sort(t,f); return t end
285
286 function thing(x)
287 x = x:match"^%s*(-)%s*$"
288 if x=="true" then return true elseif x=="false" then return false end
289 return tonumber(x) or x end
290
291 function things(x,sep, t)
292 t={};for y in x:gmatch(sep or"([^\+])") do t[1+#t]=thing(y) end; return t end
293

```

```

293 -----
294 --
295 --
296 --
297 --
298 --
299 --
300 our.go, our.no = {},{}; go=our.go
301 function go.settings() print("our",o(our)); print("your",o(your)) end
302
303 function go.sample() print(EGS():file(your.file)) end
304
305 function go.clone()
306 a= EGS():file(your.file); print(a)
307 b= a:clone() end
308
309 function go.sort( i,a,b)
310 i= EGS():file(your.file)
311 a,b=i:bestRest()
312 print(#a, #b)
313 end
314
315 your = settings(our.help)
316 os.exit( main() )

```