Homework_5

January 3, 2020

1 Homework 5 - STAT 7730

1.1 Random Sampling

1.1.1 Problem 1

Consider the pmf $p(x)$ defined by $p(x) = c\sqrt(x)\not\Vdash(x \in \{1, 2, \ldots, 50\})$

(a) What is the numerical value of c?

[2]:
```
import math
import numpy as np
import matplotlib.pyplot as plt
```

[31]:
```
x = np.array(range(0,51))**(.5)
c = x.sum()**(-1)

print("The numerical value of c is %f" % c)
```

```
The numerical value of c is 0.004183
```

(b) For $n = 10^3$, generate $X_1, \ldots, X_n$ iid with common pmf p. Let

$$\hat{p}(x) = \#\{k : X_k = x\}/n$$

be the *empirical pmf* of $X_1, \ldots, X_n$. Make a graph comparing $p(x)$ and $\hat{p}(x)$. Repeat with $n = 10^6$. Interpret your results.

[32]:
```
prob = x*c;

n = 1000
sample = np.zeros(n)

for i in range(0,n):
    rnd = np.random.uniform()
    sample[i] = sum(x <= rnd for x in np.cumsum(prob))

p_hat = np.zeros(50)

for i in range(0,50):
    p_hat[i] = sum(sample == i+1)
```

```
p_hat = p_hat/n
```

[36]:
```
fig, ax = plt.subplots()
ax.plot(list(range(1,51)), prob[1:51], 'r.', label='Real pmf')
ax.plot(list(range(1,51)),p_hat,'b.', label = 'Empirical pmf')
plt.ylabel('probability')
plt.xlabel('x')
ax.legend()
plt.show()
```

The real pmf and the empirical pmf differ signifficantly in the first case when $n = 1000$. Nevertheless, in the next case ($n = 1000000$) we can observe that both pmfs coincide.

[7]:
```
n = 1000000
sample = np.zeros(n)
cmsm = np.cumsum(prob)
for i in range(0,n):
    rnd = np.random.uniform()
    sample[i] = sum(cmsm <= rnd)

p_hat2 = np.zeros(50)

for i in range(0,50):
    p_hat2[i] = sum(sample == i+1)
```

```
p_hat2 = p_hat2/n
```

[26]:
```
fig, ax = plt.subplots()
ax.plot(list(range(1,51)), prob[1:51], 'yo', label='Real pmf')
ax.plot(list(range(1,51)),p_hat2,'b.', label = 'Empirical pmf')
plt.ylabel('probability')
plt.xlabel('x')
ax.legend()
plt.show()
```

2. Consider the pdf

$$f(x) = \frac{1}{2\sqrt{(x)}} \nVdash (0 < x < 1)$$

(a) How would you convert a uniform (0,1) random variable U into a random variable X that had pdf f?

Take $X = U^2$, then if $0 \leq x \leq 1$ we have

$$F_X(x) = P[X \geq x] = P[U^2 \geq x] = P[U \geq \sqrt{x}] = \sqrt{x}$$

and

$$f_X(x) = \frac{1}{2\sqrt{x}} \nVdash (0 < x < 1)$$

(b) Generate a sample of size 10^3 from f and compaere a plot of f to a histogram of your samples. Repeat with a sample of size 10^6.

As in the previous problem, the empirical distribution seems to converge only in the second case, with $n = 10^6$. The plot of f and this last histogram have similar shapes.

[55]:
```
n = 1000

u = np.random.uniform(size = n)

sample_f = u**2
weights = np.ones_like(sample_f)/len(sample_f)
plt.hist(sample_f, 20,weights=weights, alpha=0.8)
plt.show()
plt.plot([x/20 for x in range(1,21)],\
            [0.5*math.sqrt(x/20)**(-1) for x in range(1,21)],'r')
plt.show()
```

[56]:

```
n = 1000000

u = np.random.uniform(size = n)

sample_f = u**2
weights = np.ones_like(sample_f)/len(sample_f)
plt.hist(sample_f, 20,weights=weights, alpha=0.8)
plt.show()


plt.plot([x/20 for x in range(1,21)],\
            [0.5*math.sqrt(x/20)**(-1) for x in range(1,21)],'r')
plt.show()sum = lambda arg1, arg2: arg1 + arg2
```

1.2 Problem 3

Consider the function

$$g(x) = (sin(10x))^2|x^3 + 2x - 3|\mathbb{1}(x \in (-1,0) \cup (1/2,1))$$

and the pdf $f(x) = \frac{1}{c}g(x)$ that is the same shape as g. Use rejection sampling to generate a sample of size 10^3 from f and compare a plot of g to a histogram of your samples. Repeat with a sample of size 10^6.

Note that $x^3 + 2x - 3 = (x-1)(x^2 + x + 3) = ([x+0.5]^2 + 2.75) < 0$ when $x < 1$.

Hence, $|x^3 + 2x - 3| = 3 - 2x - x^3$, and

$$\frac{d}{dx}(3 - 2x - x^3) = -3x^2 - 2 < 0,$$

so that $3 - 2x - x^3$ is decreasing and attains its maximum when $x = -1$. Because $(sin(10x))^2 \leq 1$, we have that

$$g(x) \leq (3 - 2x - x^3)\mathbb{1}(x \in (-1,0) \cup (1/2,1)) \leq 3 - 2(-1) - (-1)^3 = 6.$$

Therefore, we can use rejection sampling on the rectangle $[0,6] \times [-1,1]$. [Figure] shows how with $n = 6$, the empirical distribution shows a very similar shape to the one of g.

[21]:
```
g = lambda x: (math.sin(10*x)**2)*abs(x**3 + 2*x- 3)*(((x > -1)and (x < 0)) +␣
↪((x > 1/2) and (x < 1)));

n = 1000
i = 0
sample = np.zeros(n)

while (i < n):
    x = np.random.uniform()*2 - 1
    y = np.random.uniform()*6
    if (y <= g(x)):
        sample[i] = x
        i = i + 1


weights = np.ones_like(sample)/len(sample)
plt.hist(sample, 30,weights=weights, alpha=0.8)
plt.show()

plt.plot([float(x)/40-1 for x in range(0,81)],\
            [g(float(x)/40-1) for x in range(0,81)],'r')
plt.show()
```

[37]:
```
n = 1000000

i = 0
sample = np.zeros(n)
```

```
while (i < n):
    x = np.random.uniform()*2 - 1
    y = np.random.uniform()*6
    if (y <= g(x)):
        sample[i] = x
        i = i + 1


weights = np.ones_like(sample)/len(sample)
plt.hist(sample, 30,weights=weights, alpha=0.8)
plt.show()

plt.plot([float(x)/40-1 for x in range(0,81)],\
            [g(float(x)/40-1) for x in range(0,81)],'r')
plt.show()
```