

LatTest Mathematica Review

This is my Mathematica review notebook.
For the source code on GitHub, [click here](#).

1. Mathematica Basics

创建和访问list:

```
In[ ]:= x = Range[10]
```

```
Out[ ]:= {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}
```

```
In[ ]:= x[[3]]
```

```
Out[ ]:= 3
```

```
In[ ]:= x[[1;;3]]
```

```
Out[ ]:= {1, 2, 3}
```

迭代器:

```
In[ ]:= Table[x^2, {x, 10}]
```

```
Out[ ]:= {1, 4, 9, 16, 25, 36, 49, 64, 81, 100}
```

赋值:

```
In[ ]:= x = 7
```

```
Out[ ]:= 7
```

2. Section2

■ Question 1.2.1 : Sampling without replacement.

10000行代码中有30行可以被改进，抽取10行代码

```
In[ ]:= (*1行代码需要被改进的概率*)
Binomial[10000, 10]
Binomial[30, 1] * Binomial[10000 - 30, 9] / Binomial[10000, 10]
N[%, 18]
```

```
Out[ ]:= 2 743 355 077 591 282 538 231 819 720 749 000
```

```
Out[ ]:= 2 143 751 464 028 247 883 152 007 617
73 351 740 042 547 661 450 048 655 635
```

```
Out[ ]:= 0.0292256388571663780
```

```
In[ ]:= (*0行代码需要被改进的概率*)
Binomial[30, 0] * Binomial[10000 - 30, 10] / Binomial[10000, 10]
N[%]
(*使用%将结果转化为分数*)
```

```
Out[ ]:= 1 016 852 777 770 732 245 908 435 612 997
1 047 882 000 607 823 735 000 695 080 500
```

```
Out[ ]:= 0.970389
```

也可以使用内置的HypergeometricDistribution函数直接计算

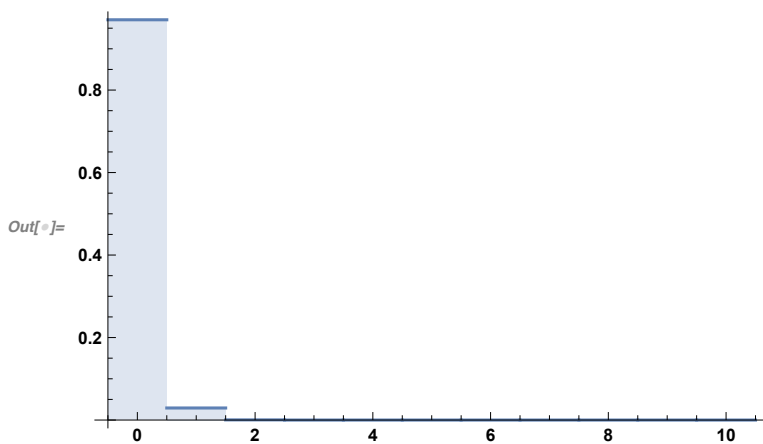
```
PDF[HypergeometricDistribution[10, 30, 10000], 1]
N[%]
```

```
Out[ ]:= 2 143 751 464 028 247 883 152 007 617
73 351 740 042 547 661 450 048 655 635
```

```
Out[ ]:= 0.0292256
```

使用DiscretePlot命令来打印Hypergeometric的概率

```
DiscretePlot[PDF[HypergeometricDistribution[10, 30, 10000], k], {k, 0, 10}, PlotRange -> F
```

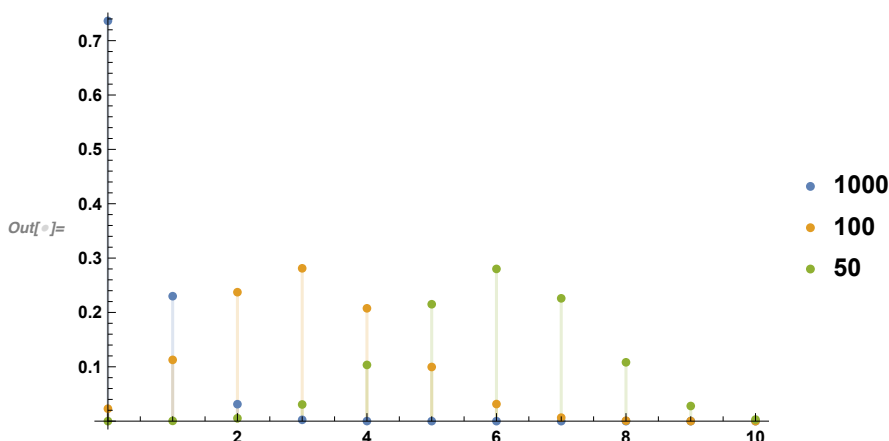


Now suppose that total number of lines of code from which the sample of size 10 is taken is (a) 1000, (b) 100 or (c) 50, whilst the total number of improvable lines of code remains the same at 30. Alter your command to permit, the probabilities for each of (a), (b) and (c) to be plotted on the one plot. Comment on the differences between the probabilities.

(*从N个中选10个，其中有k行代码需要被改进的概率的PDF，这里同时生成3个超几何分布，分别是N为1000，500和50*)

`Table[PDF[HypergeometricDistribution[10, 30, N], k], {N, {1000, 100, 50}}]`
$$Out[*]= \left\{ \begin{array}{ll} \frac{\text{Binomial}[30,k] \text{Binomial}[970,10-k]}{263\,409\,560\,461\,970\,212\,832\,400} & 0 \leq k \leq 10 \\ 0 & \text{True} \end{array} \right., \left\{ \begin{array}{ll} \frac{\text{Binomial}[30,k] \text{Binomial}[70,10-k]}{17\,310\,309\,456\,440} & 0 \leq k \leq 10 \\ 0 & \text{True} \end{array} \right., \left\{ \begin{array}{ll} \frac{\text{Binomial}[20,10-k] \text{Binomial}[30,k]}{10\,272\,278\,170} & 0 \leq k \leq 10 \\ 0 & \text{True} \end{array} \right\}$$

```
DiscretePlot[Evaluate[%], {k, 0, 10}, PlotRange→Full, PlotLegends→{1000, 100, 50}]
```



- Question 1.2.2: Bayes Theorem - Sampling Without Replacement

Given an observed number of 5 lines of code that could be improved in the 20 samples, the total number of lines of code that could be improved out of 1000 was 0, 1, 2, ..., 1000. Plot these probabilities.

(* prior 先验概率, Join命令将两个list合并, 先验概率的和需要相加为1, 有1001个值. Joint是先验概率和likelihood的乘积*)

```
Short[prior = Join[Table[0.9/101, 101], Table[0.1/900, 900]]]
```

```
Short[joint = prior * Evaluate[Table[PDF[HypergeometricDistribution[20, i, 1000], 5],
```

```
Out[ ]/Short= {0.00891089, 0.00891089, 0.00891089, 0.00891089,
<<993>>, 0.000111111, 0.000111111, 0.000111111, 0.000111111}
```

$$\text{Out[5]/Short} = \{0., 0., 0., 0., 0., 1.67454 \times 10^{-11}, 9.89578 \times 10^{-11}, 3.41126 \times 10^{-10}, \\ 8.95927 \times 10^{-10}, 1.98535 \times 10^{-9}, \ll 981 \gg, 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.\}$$

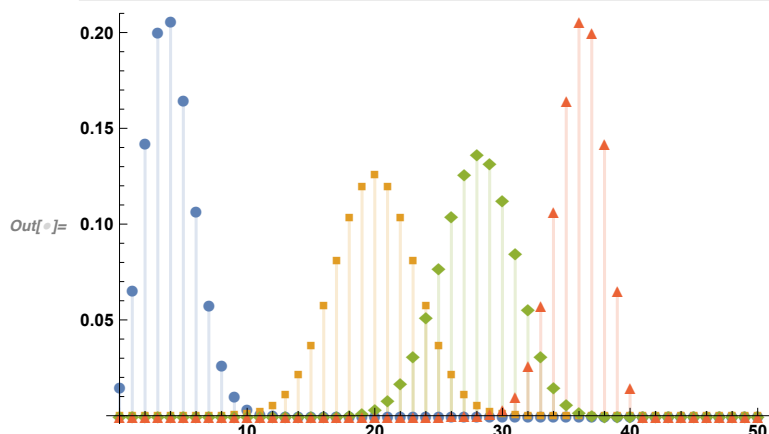
$\ln[\bullet] :=$ (*conditional , 根据贝叶斯定理得到的条件概率*)

```
Short[conditional = joint / Total[joint]]
```

$$Out[5]/Short = \{0., 0., 0., 0., 0., 1.49743 \times 10^{-9}, 8.84913 \times 10^{-9}, 3.05046 \times 10^{-8}, \\ 8.01167 \times 10^{-8}, 1.77537 \times 10^{-7}, \langle\langle 981 \rangle\rangle, 0., 0., 0., 0., 0., 0., 0., 0., 0., 0.\}$$

■ Question 1.3 Binomial Probabilities

```
Table[PDF[BinomialDistribution[40, p], k], {p, {0.1, 0.5, 0.7, 0.9}}];
DiscretePlot[Evaluate[%], {k, 0, 50}, PlotMarkers -> Automatic, PlotRange -> Full]
```



Section3

■ Question 1

17个球标着输3个球标赢，轮流拿球，不放回，拿到第三个标记为赢的球的获胜.

a. 如果你先拿，求你第二次拿就赢了的概率（前三个拿的全是赢）

```
In[ ]:= PDF[HypergeometricDistribution[3, 3, 20], 3]
```

Out[]:= $\frac{1}{1140}$

b. 如果你先拿，求你对家第二次拿球赢了的概率

```
In[ ]:= PDF[HypergeometricDistribution[3, 3, 20], 2] * 1 / 17
```

Out[]:= $\frac{1}{380}$

c. 如果你先拿，你赢的概率

```
In[ ]:= Sum[PDF[HypergeometricDistribution[2k, 3, 20], 2] * 1 / (20 - 2k), {k, 1, 9}]
```

Out[]:= $\frac{35}{76}$

d. 如果你第二个拿，你赢的概率

```
In[ ]:= Sum[PDF[HypergeometricDistribution[2k+1, 3, 20], 2] * 1 / (19 - 2k), {k, 1, 9}]
```

Out[]:= $\frac{41}{76}$

■ Question 2

100 fuses there are 20 defective ones. A sample of 5 fuses are randomly selected from the lot without replacement. Let X be the number of defective fuses found in the example

- Find $P(X = 0)$.
- Cumulative probability $P(X \leq 3)$
- Mean of X , $E(X)$
- Second Moment of X , $E(X^2)$
- Variance of X , $\text{Var}(X)$
- Probability bar graph for PMF of X

```
(*a solution*)
PDF[HypergeometricDistribution[5, 20, 100], 0];
N[%]
```

Out[191]= 0.319309

```
(*b solution*)
CDF[HypergeometricDistribution[5, 20, 100], 3];
N[%]
```

Out[193]= 0.994646

```
In[194]:= (*c 两种方法*)
mu = Mean[HypergeometricDistribution[5, 20, 100]]
Expectation[X, X \[Distributed] HypergeometricDistribution[5, 20, 100]]
```

Out[194]= 1

Out[195]= 1

```
(*d solution*)
ex2 = Expectation[X^2, X \[Distributed] HypergeometricDistribution[5, 20, 100]]
```

Out[196]= $\frac{175}{99}$

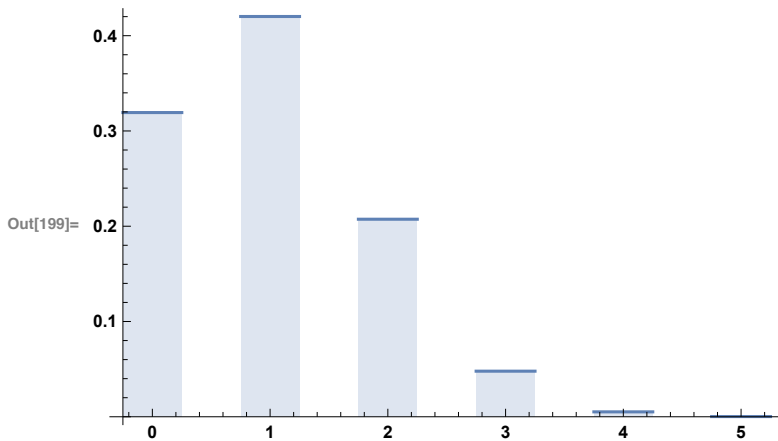
```
In[197]:= (*e 两种方法*)
ex2 - mu^2
Variance[HypergeometricDistribution[5, 20, 100]]
```

Out[197]= $\frac{76}{99}$

Out[198]= $\frac{76}{99}$

```
(*f solution*)
```

```
DiscretePlot[PDF[HypergeometricDistribution[5, 20, 100], k], {k, 0, 5}, ExtentSize->0.5]
```



Section 6

■ Question 1

Suppose $X_1 \sim N(3, 4)$, $X_2 \sim N(3, 4)$, X_1 and X_2 are independent.

a. Find MGF of X_1 and X_2

b. Let $Y = 5X_1 - 2X_2 + 6$, Find the MGF of Y

c. Name the distribution Y and give the values of the associated parameters. Answer: $Y \sim N(15, 116)$

In[225]:=

```
(*注意：在这个垃圾软件中，正态分布使用标准差做输入，而不是公式中的方差*)
```

```
(*a*)
```

```
MomentGeneratingFunction[NormalDistribution[3,2],t]
```

```
(*b*)
```

```
Y = TransformedDistribution[5X1-2X2+6,{X1~NormalDistribution[3,2],X2~NormalDistribution[3,2]}
MomentGeneratingFunction[Y,t]
```

Out[225]= e^{3t+2t^2}

Out[227]= $e^{15t+58t^2}$

■ Question 2

Let \bar{X} be the mean of a random sample of size 12 from the uniform distribution on the interval (0, 1) (which has the mean $1/2$ and variance $1/12$) Use Mathematica to do the following:

a. Find pdf of \bar{X}

b. Find the probability $P(1/2 \leq \bar{X} \leq 2/3) = P(6 \leq X \leq 8) = F_X(8) - F_X(6)$

c. By the CLT \bar{X} approximately has a normal distribution with mean $1/2$ and variance $1/144$.

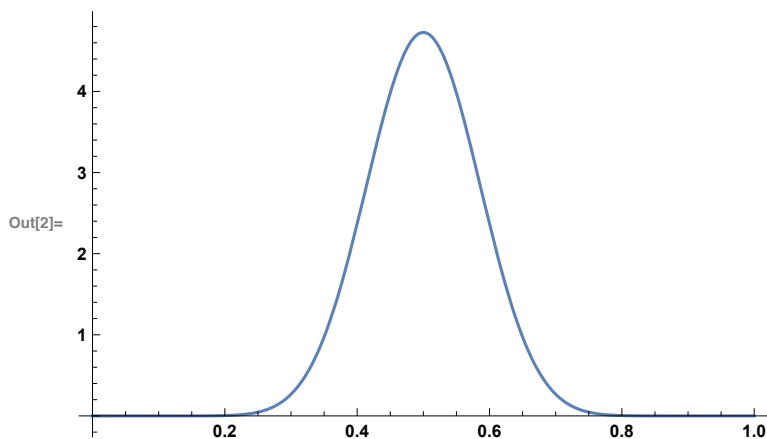
Now approximate the probability $P(1/2 \leq \bar{X} \leq 2/3)$ based on this normal distribution. Compare the result with that in (b)

(*a*)

12*PDF[UniformSumDistribution[12],x]/.{x→12y} (*生成12个UniformSumDistribution*)

Plot[12*Evaluate[PDF[UniformSumDistribution[12],x] /. {x → y*12}], {y,0,1}]

Out[1]= 12

$$\begin{aligned}
 & \begin{cases} 0 & 12y < 0 \\ \frac{35831808y^{11}}{1925} & 0 \leq 12y < 1 \\ \frac{743008370688y^{11}-12(-1+12y)^{11}}{39916800} & 1 \leq 12y < 2 \\ \frac{743008370688y^{11}+66(-2+12y)^{11}-12(-1+12y)^{11}}{39916800} & 2 \leq 12y < 3 \\ \frac{1}{39916800} \left(743008370688y^{11} - \right. & 3 \leq 12y < 4 \\ \quad \left. 220(-3+12y)^{11} + 66(-2+12y)^{11} - 12(-1+12y)^{11} \right) & \\ \frac{1}{39916800} \left(743008370688y^{11} + 495(-4+12y)^{11} - \right. & 4 \leq 12y < 5 \\ \quad \left. 220(-3+12y)^{11} + 66(-2+12y)^{11} - 12(-1+12y)^{11} \right) & \\ \frac{1}{39916800} \left(743008370688y^{11} - 792(-5+12y)^{11} + 495(-4+12y)^{11} - \right. & 5 \leq 12y < 6 \\ \quad \left. 220(-3+12y)^{11} + 66(-2+12y)^{11} - 12(-1+12y)^{11} \right) & \\ \frac{1}{39916800} \left(-792(7-12y)^{11} + 495(8-12y)^{11} - \right. & 6 \leq 12y < 7 \\ \quad \left. 220(9-12y)^{11} + 66(10-12y)^{11} - 12(11-12y)^{11} + (12-12y)^{11} \right) & \\ \frac{1}{39916800} \left(495(8-12y)^{11} - 220(9-12y)^{11} + \right. & 7 \leq 12y < 8 \\ \quad \left. 66(10-12y)^{11} - 12(11-12y)^{11} + (12-12y)^{11} \right) & \\ \frac{1}{39916800} & 8 \leq 12y < 9 \\ \left(-220(9-12y)^{11} + 66(10-12y)^{11} - 12(11-12y)^{11} + (12-12y)^{11} \right) & \\ \frac{66(10-12y)^{11}-12(11-12y)^{11}+(12-12y)^{11}}{39916800} & 9 \leq 12y < 10 \\ \frac{-12(11-12y)^{11}+(12-12y)^{11}}{39916800} & 10 \leq 12y < 11 \\ \frac{(12-12y)^{11}}{39916800} & 11 \leq 12y \leq 12 \\ 0 & \text{True} \end{cases}
 \end{aligned}$$


```
(*b , 两种方法*)
N[Probability[6 ≤ x ≤ 8, x ≈ UniformSumDistribution[12]]]
N[CDF[UniformSumDistribution[12], 8] - CDF[UniformSumDistribution[12], 6]]
```

Out[3]= 0.477724

Out[4]= 0.477724

```
In[5]:= (*c , 两种方法*)
N[Probability[1/2 ≤ x ≤ 2/3, x ≈ NormalDistribution[1/2, 1/12]]]
N[CDF[NormalDistribution[1/2, 1/12], 2/3] - CDF[NormalDistribution[1/2, 1/12], 1/2]]
```

Out[5]= 0.47725

Out[6]= 0.47725

■ Question 3

Let $Y = A + B + C + D$ be the sum of a random sample of size 4 from the distribution whose pdf is that of $U^{1/3}$ where $U \sim U(0, 1)$.

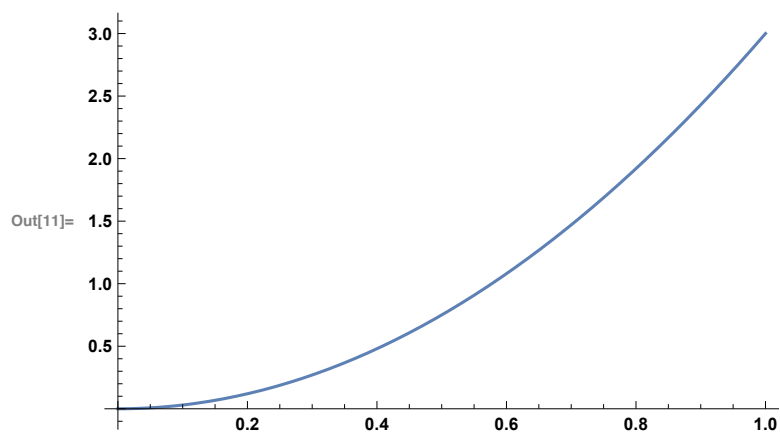
- The function t1 in the notebook Lab6.nb sets up the pdf of $U^{1/3}$. Plot it and find the mean and variance of A.
- Find the pdf of Y, and plot the pdf of Y.
- Calculate $P(0.3 \leq Y \leq 1.5)$ based on the pdf of Y.
- Approximate $P(0.3 \leq Y \leq 1.5)$ based on the normal distribution, and see how close it is to the result of (c).


```

In[10]:= (*a. solution*)
(*function t1 in notebook Lab6.nb*)
t1 = TransformedDistribution[U^(1/3), U  $\approx$  UniformDistribution[]]
(*plot pdf graph*)
Plot[PDF[t1,x],{x,0,1}]
(*mean and variance*)
Mean[t1]
Variance[t1]

```

Out[10]= BetaDistribution[3, 1]



Out[12]= $\frac{3}{4}$

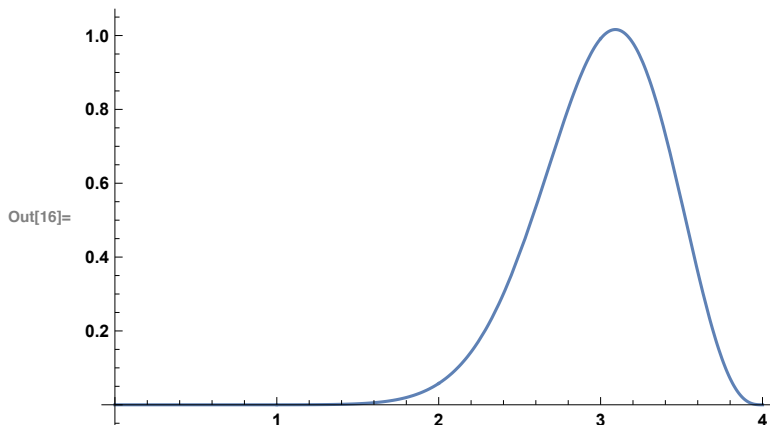
Out[13]= $\frac{3}{80}$

```
In[14]:= (*b. solution*)
(*construct distribution of Y*)
tf=TransformedDistribution[A + B + C + D,{A<=t1,B<=t1,C <= t1,D <= t1}]
ee= PDF[tf]
Plot[ee[x],{x,0,4}]
```

```
Out[14]= TransformedDistribution[x1 + x2 + x3 + x4,
  {x1 <= BetaDistribution[3, 1], x2 <= BetaDistribution[3, 1],
   x3 <= BetaDistribution[3, 1], x4 <= BetaDistribution[3, 1]}]
```

```
Out[15]= Function[x,
  
$$\frac{1}{61600} \left( 415800 (-4+x)^3 \text{Sign}[-4+x] + 831600 (-4+x)^4 \text{Sign}[-4+x] + 665280 (-4+x)^5 \right. \\ \left. \text{Sign}[-4+x] + 277200 (-4+x)^6 \text{Sign}[-4+x] + 67320 (-4+x)^7 \text{Sign}[-4+x] + \right. \\ 9900 (-4+x)^8 \text{Sign}[-4+x] + 880 (-4+x)^9 \text{Sign}[-4+x] + 44 (-4+x)^{10} \text{Sign}[-4+x] + \\ (-4+x)^{11} \text{Sign}[-4+x] - 166320 (-3+x)^5 \text{Sign}[-3+x] - \\ 166320 (-3+x)^6 \text{Sign}[-3+x] - 71280 (-3+x)^7 \text{Sign}[-3+x] - \\ 15840 (-3+x)^8 \text{Sign}[-3+x] - 1980 (-3+x)^9 \text{Sign}[-3+x] - \\ 132 (-3+x)^{10} \text{Sign}[-3+x] - 4 (-3+x)^{11} \text{Sign}[-3+x] + 11880 (-2+x)^7 \text{Sign}[-2+x] + \\ 5940 (-2+x)^8 \text{Sign}[-2+x] + 1320 (-2+x)^9 \text{Sign}[-2+x] + \\ 132 (-2+x)^{10} \text{Sign}[-2+x] + 6 (-2+x)^{11} \text{Sign}[-2+x] - 220 (-1+x)^9 \text{Sign}[-1+x] - \\ \left. 44 (-1+x)^{10} \text{Sign}[-1+x] - 4 (-1+x)^{11} \text{Sign}[-1+x] + x^{11} \text{Sign}[x] \right)$$

```



```
(*c. solution*)
N[CDF[tf,1.5]- CDF[tf,0.3]]
```

```
Out[17]= 0.000350282
```

```

In[24]:= (*d. solution*)
Mean[tf]
Variance[tf]
N[CDF[NormalDistribution[3,Sqrt[0.15]],1.5]-CDF[NormalDistribution[3,Sqrt[0.15]],0.3]]

(*跟正态分布结果相比, 结果不是很好, 因为区间[0.3, 1.5]距离中心点3太远了, 所以最后得结果值很小*)

Out[24]= 3

Out[25]=  $\frac{3}{20}$ 

Out[26]= 0.0000537556

```

Section 9

■ Question 4 - MLE & FisherInformation, Cramer-Rao lower bound

Let X_1, \dots, X_n be a random sample from a gamma distribution with $\alpha = 4$ so that

$$f(x; \theta) = \frac{e^{-\frac{x}{\theta}} x^3}{6 \theta^4}, \quad 0 < x < \infty, \quad 0 < \theta$$

Use Mathematica to calculate appropriate derivatives and means.

- Find the maximum likelihood estimator of θ
- Is the MLE an efficient estimator of θ ?
- Give an approximate $100(1 - \alpha)\%$ confidence interval for θ

```
(*Solutions*)
dist=ProbabilityDistribution[1/(6* $\theta^4$ )* $x^3$ *Exp[- $x/\theta$ ],{ $x,0,\infty$ },Assumptions→ $\theta>0$ ]; (*构建
logpdfProductDerivative = D[Log[PDF[dist, x]],  $\theta$ ]; (*logpdf的乘积对 $\theta$ 求偏导*)
h[x_] := Simplify[D[Log[PDF[dist, x]],  $\theta$ ], Assumptions→ $x > 0$ ] (*构建一个函数*)
fisherInformation = Expectation[h[x] ^ 2, x  $\approx$  dist]

mle[n_] := Solve[Sum[h[x[i]], {i, 1, n}] == 0,  $\theta$ ]
estimates = {mle[2], mle[3], mle[4]} (*三个极大似然估计量*)
mu = Expectation[ $\theta /. estimates$ , {x[1]  $\approx$  dist, x[2]  $\approx$  dist, x[3]  $\approx$  dist, x[4]  $\approx$  dist}]
LowerBound = 1 / ({2, 3, 4} * fisherInformation)
varianceMLE = (Expectation[( $\theta /. estimates$ ) ^ 2, {x[1]  $\approx$  dist, x[2]  $\approx$  dist, x[3]  $\approx$  dist, x[4]  $\approx$  dist}])
```

$$\text{Out[171]} = \frac{4}{\theta^2}$$

$$\text{Out[173]} = \left\{ \left\{ \theta \rightarrow \frac{1}{8} (x[1] + x[2]) \right\} \right\}, \left\{ \left\{ \theta \rightarrow \frac{1}{12} (x[1] + x[2] + x[3]) \right\} \right\}, \\ \left\{ \left\{ \theta \rightarrow \frac{1}{16} (x[1] + x[2] + x[3] + x[4]) \right\} \right\}$$

$$\text{Out[174]} = \{\{\theta\}, \{\theta\}, \{\theta\}\}$$

$$\text{Out[175]} = \left\{ \frac{\theta^2}{8}, \frac{\theta^2}{12}, \frac{\theta^2}{16} \right\}$$

$$\text{Out[176]} = \left\{ \left\{ \frac{\theta^2}{8} \right\}, \left\{ \frac{\theta^2}{12} \right\}, \left\{ \frac{\theta^2}{16} \right\} \right\}$$

So, the MLE is $\bar{X}/4$

The Fisher information is $4/\theta^2$

The Cramer-Rao lower bound is $\theta^2/(4n)$

The MLE is unbiased and achieves the lower bound so is efficient.

Sample examples

■ Question 1

Given a discrete distribution:

$$P(X = x) = C \{200 - x\} \{100\} * 2^{-2\theta\theta+x}, \quad x = 0, 1, 2, \dots, 100$$

a). Find the probability $P(X \geq 10)$

b). Find the mean $E(X)$

c). Find $E(X^2)$

d). Find $E\left(\frac{1}{(1+X)^2}\right)$

```
In[56]:= (*a. solution*)
Q1 = ProbabilityDistribution[Binomial[200-x, 100] * 2^(x - 200), {x, 0, 100, 1}]
N[Probability[X ≥ 10, X ≈ Q1]]
```

```
Out[56]= ProbabilityDistribution[2-200+x Binomial[200 - x, 100], {x, 0, 100, 1}]
```

```
Out[57]= 0.469421
```

```
(*b. solution*)
N[Mean[Q1]]
```

```
Out[59]= 10.326
```

```
In[61]:= (*c. solution*)
N[Expectation[X^2, X ≈ Q1]]
```

```
Out[61]= 169.022
```

```
In[66]:= (*d. solution*)
N[Expectation[(X+1)^-2, X ≈ Q1]]
```

```
Out[66]= 0.0886201
```

■ Question 2

Continuous random variable X have following pdf:

$$f(x) = \frac{2}{9} (2 - x) (1 + x), -1 < x < 2.$$

- Find the cdf $F(x)$ of X .
- Find the probability $P(-2 < X < \frac{9}{5})$.
- Find the mean $E(X)$.
- Find the mgf $M(t) = E[e^{tx}]$.
- Find the third moment $E(X^3)$.
- Let $Y = X^2$.
 - What are the possible values of Y ?
 - Find the pdf $g(y)$ of Y .

```
In[33]:= (*a. solution*)
Q2 = ProbabilityDistribution[2/9*(x+1)*(2-x), {x, -1, 2}]
CDF[Q2]
```

```
Out[33]= ProbabilityDistribution[ $\frac{2}{9} (2 - x) (1 + x)$ , {x, -1, 2}]
```

```
Out[34]= Function[x, { $\frac{1}{27} (7 + 12 x + 3 x^2 - 2 x^3)$  -1 < x < 2, Listable}
True
```

```
(*b. solution*)
CDF[Q2, 9/5] - CDF[Q2, -2]
N[%]
```

Out[37]= $\frac{3332}{3375}$

Out[38]= 0.987259

```
In[39]:= (*c. solution*)
Mean[Q2]
```

Out[39]= $\frac{1}{2}$

```
In[40]:= (*d. solution*)
MomentGeneratingFunction[Q2, t]
```

Out[40]=
$$\frac{e^{-t} (4 + 6 t + e^{3 t} (-4 + 6 t))}{9 t^3}$$

```
In[41]:= (*e. solution*)
Expectation[X^3, X ≈ Q2]
```

Out[41]= $\frac{4}{5}$

```
In[43]:= (*f - ii solution*)
Y = TransformedDistribution[X^2, {X ≈ Q2}]
PDF[Y]
```

Out[43]= TransformedDistribution[X^2 ,
 $\{X \approx \text{ProbabilityDistribution}[\frac{2}{9} (2 - \check{x}) (1 + \check{x}), \{\check{x}, -1, 2\}]\}$]

Out[44]= Function[\check{x} , $\begin{cases} \frac{2 + \sqrt{\check{x}} - \check{x}}{9 \sqrt{\check{x}}} & 1 \leq \check{x} < 4 \\ -\frac{2(-2 + \check{x})}{9 \sqrt{\check{x}}} & 0 < \check{x} < 1 \\ 0 & \check{x} \geq 4 \text{ || } \check{x} < 0 \\ \text{Indeterminate} & \text{True} \end{cases}$, Listable]