Wk3: PRob Distributions

D. ODay

2022-06-27

Probability and Probability Distributions

Intro to Probability

Definitions

- **Probability** is the chance that an event will or will not occur.the terms are typically expressed in fractions or decimals
- An **event** is one or more of the possible outcomes of a situation or experiment.
- **experiment** is an activity which produces an event
- **sample space** is the set of all possible outcomes from an experiment
- Events are termed **mutually exclusive** when one and only one can take place at the same time
- **Collectivelly Exhaustive** refers to lists containing all of the possible events which may result from an experiment
- The probability an event will occur is $P = \frac{\#ofeventoutcomes}{TotalPossibleOutcomes}$

Rules and conditions

- Concern
  - the case where one event or another will occur
    - Also known as marginal or unconditional probability
    - P(A) = the probability P of event A occurring
    - Where a single probability is involved, only one event can take place
    - Ex) what is the probability of selecting a part out of 100 of them
    - P = 1%
  - The next situation with two or more events where they both may occur

Mutually Exclusive

- For mutually exclusive events
  - P(A or B) = P(A) + P(B)
- For non-mutually exclusive events
  - P(A or B) = P(A) + P(B) - P(A+B)
  - Ex question
    - With two vendors A and B and with some defective parts for each vendor, what is the probability of selecting vendor A and a defective part?

Independent Conditions

- Marginal Probability
  - P(A) Independent Event (e.g. coin toss)
- Joint Probability
  - The probability of two or more events occurring together (or in succession) is the product of their marginal probabilities
  - P(AB) = P(A) x P(B)
  - Ex) The probability of a machine operator producing a defective part at any point in time is 0.05. What is the probability that three bad parts will be produced in succession?
    * P(ABC) = P(A) x P(B) x P(C)
    * P(3 Defectives) = P(Def) x P(Def) x P(Def)
    * P(3 def) = 0.05 x 0.05 x 0.05
- Conditional Probability
  - P(B|A)
    * The probability event B will occur given event A has occurred
    * P(B|A) = P(B) because A and B are independent

Dependent Conditions

- Conditional Probability
  - $P(B|A) = \frac{P(B \cap A)}{A}$

Joint Probabilities Under Statistical Dependence

- The formula for joint probabilities under statistical dependence is a variation of the conditional probability formula
- The joint probability of B and A is the following
  - P(BA) = P(B|A) x P(A)
  - Where P(BA) is the probability of events B and A happening together in succession

Probability Distributions

- The theoretical frequency distributions which are collectively exhaustive

```
require(lolcat)
```

Distribution in R example

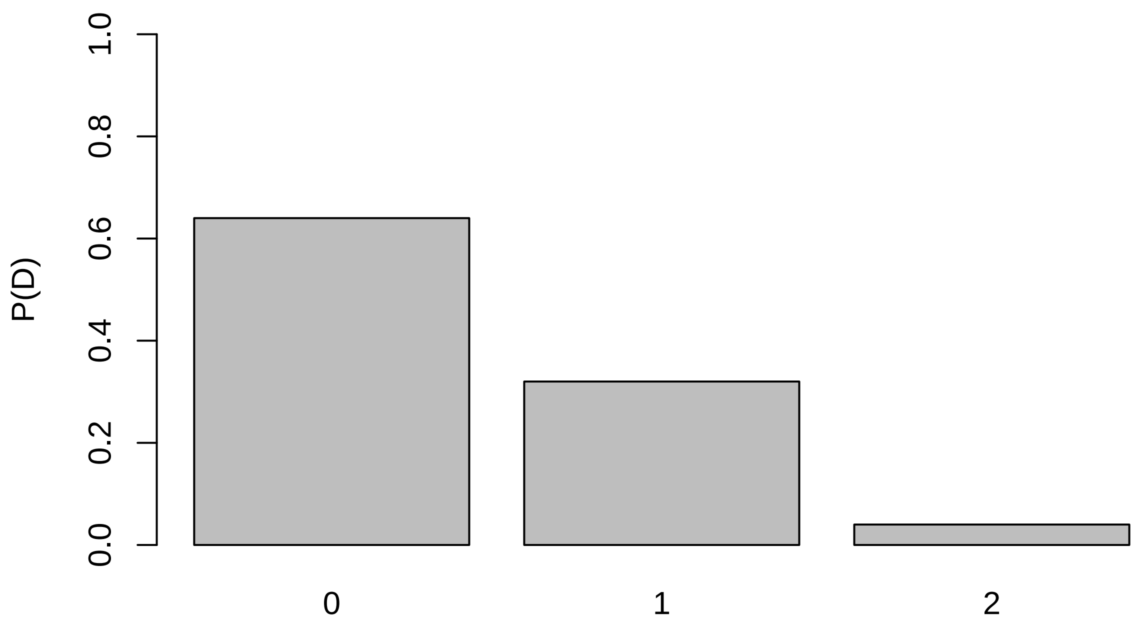
```
## Loading required package: lolcat
```

```
## lolcat 2.0.0
```

```r
#Get distribution
table.dist.binomial(n=2, p=0.2)
```

```
##   x p.at.x eq.and.above eq.and.below
## 0 0   0.64         1.00         0.64
## 1 1   0.32         0.36         0.96
## 2 2   0.04         0.04         1.00
```

```r
#Barplot of Distribution
n = 2
P = 0.2
data=dbinom(x=0:n, size=n, prob=P)
names(data) = 0:n
barplot(data, ylab="P(D)", ylim=c(0,1))
```



Types of Probability Distributions

- Discrete
    - There are a limited number of possible values
- Continuous
    - A continuous probability distribution has relatively unlimited possibilities for variable value
- A **random variable** is one which can take on different values as a result of the outcomes of a random experiment. Can be either discrete or continuous

```r
Daily.Production <- read.table("~/Documents/GitHub/school_cu/school_cu/methods for quality improvement/I

(freqdistdp<-round.object(frequency.dist.grouped(Daily.Production$V1),3))
```

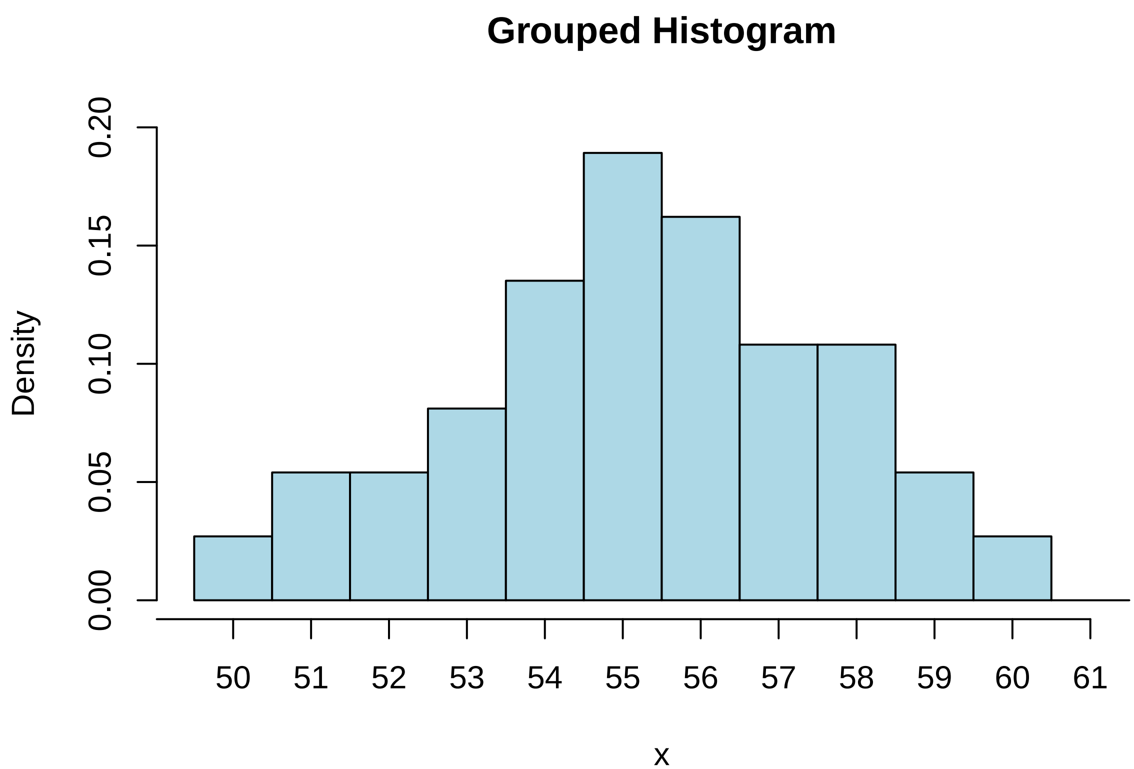
```
##     l min midpoint max u freq rel.freq cum.up cum.down
## 1  [  50     50.5  51 )    1    0.027  0.027    1.000
## 2  [  51     51.5  52 )    2    0.054  0.081    0.973
## 3  [  52     52.5  53 )    2    0.054  0.135    0.919
## 4  [  53     53.5  54 )    3    0.081  0.216    0.865
## 5  [  54     54.5  55 )    5    0.135  0.351    0.784
## 6  [  55     55.5  56 )    7    0.189  0.541    0.649
## 7  [  56     56.5  57 )    6    0.162  0.703    0.459
## 8  [  57     57.5  58 )    4    0.108  0.811    0.297
## 9  [  58     58.5  59 )    4    0.108  0.919    0.189
## 10 [  59     59.5  60 )    2    0.054  0.973    0.081
## 11 [  60     60.5  61 )    1    0.027  1.000    0.027
```

```
(probdistdp<-freqdistdp[,c("min","freq","rel.freq")])
```

```
##    min freq rel.freq
## 1   50    1    0.027
## 2   51    2    0.054
## 3   52    2    0.054
## 4   53    3    0.081
## 5   54    5    0.135
## 6   55    7    0.189
## 7   56    6    0.162
## 8   57    4    0.108
## 9   58    4    0.108
## 10  59    2    0.054
## 11  60    1    0.027
```

```
colnames(probdistdp)<-c("Daily Production", "#of Days", "P(DP)")

# Probability Distribution (Histogram)
hist.grouped(Daily.Production$V1, freq = F, anchor.value=50, ylim=c(0,0.20))
```



Expected Value of a Discrete Variable

- One of the most important factors related to any probability distribution is the ability to define the expected value of a random variable

```
# Expected Value of a Discrete Random Variable
x<-probdistdp$`Daily Production`
y<-probdistdp$`P(DP)`
weighted.mean(x,y)
```

```
## [1] 55.24324
```

```
mean(Daily.Production$V1)
```

```
## [1] 55.24324
```

Common Probability Distributions

- Discrete
  - Binomial
  - Poisson
  - Hypergeometric
  - Geometric
- Continuous
  - Normal

  - Exponential
  - Weibull Family
  - Johnson Family
  - Other Distributions

Discrete Distributions

The binomial distribution

- Basically either/or between two probabilities

```
# Get density at exactly that x-value
dbinom(x=45, size=50, prob=0.8)
```

```
## [1] 0.0295312
```

```
#binomial table (lolcat)
#gives desnsity at x for each X
round.object(table.dist.binomial(n=50, p=0.8), 5)
```

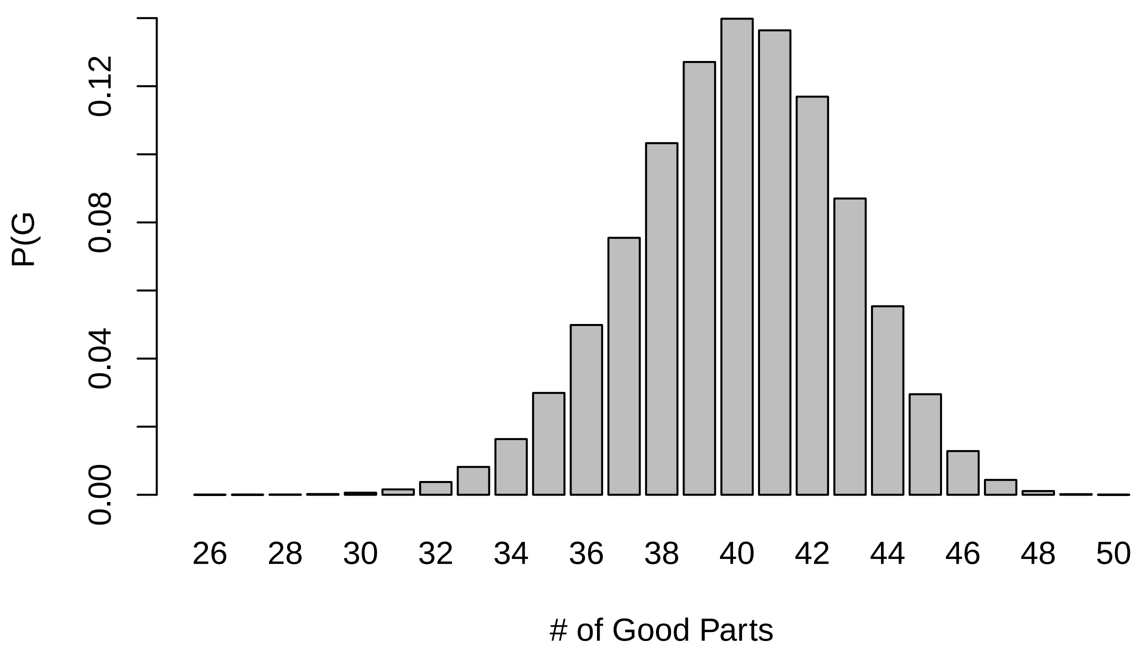
```
##      x  p.at.x eq.and.above eq.and.below
## 0    0 0.00000      1.00000      0.00000
## 1    1 0.00000      1.00000      0.00000
## 2    2 0.00000      1.00000      0.00000
## 3    3 0.00000      1.00000      0.00000
## 4    4 0.00000      1.00000      0.00000
## 5    5 0.00000      1.00000      0.00000
## 6    6 0.00000      1.00000      0.00000
## 7    7 0.00000      1.00000      0.00000
## 8    8 0.00000      1.00000      0.00000
## 9    9 0.00000      1.00000      0.00000
## 10  10 0.00000      1.00000      0.00000
## 11  11 0.00000      1.00000      0.00000
## 12  12 0.00000      1.00000      0.00000
## 13  13 0.00000      1.00000      0.00000
## 14  14 0.00000      1.00000      0.00000
## 15  15 0.00000      1.00000      0.00000
## 16  16 0.00000      1.00000      0.00000
## 17  17 0.00000      1.00000      0.00000
## 18  18 0.00000      1.00000      0.00000
## 19  19 0.00000      1.00000      0.00000
## 20  20 0.00000      1.00000      0.00000
## 21  21 0.00000      1.00000      0.00000
## 22  22 0.00000      1.00000      0.00000
## 23  23 0.00000      1.00000      0.00000
## 24  24 0.00000      1.00000      0.00000
## 25  25 0.00000      1.00000      0.00000
## 26  26 0.00001      1.00000      0.00001
## 27  27 0.00002      0.99999      0.00003
## 28  28 0.00007      0.99997      0.00010
## 29  29 0.00022      0.99990      0.00032
```

```
## 30 30 0.00061       0.99968       0.00093
## 31 31 0.00158       0.99907       0.00251
## 32 32 0.00375       0.99749       0.00626
## 33 33 0.00818       0.99374       0.01444
## 34 34 0.01636       0.98556       0.03080
## 35 35 0.02992       0.96920       0.06072
## 36 36 0.04986       0.93928       0.11059
## 37 37 0.07547       0.88941       0.18606
## 38 38 0.10328       0.81394       0.28933
## 39 39 0.12711       0.71067       0.41644
## 40 40 0.13982       0.58356       0.55626
## 41 41 0.13641       0.44374       0.69267
## 42 42 0.11692       0.30733       0.80959
## 43 43 0.08701       0.19041       0.89660
## 44 44 0.05537       0.10340       0.95197
## 45 45 0.02953       0.04803       0.98150
## 46 46 0.01284       0.01850       0.99434
## 47 47 0.00437       0.00566       0.99871
## 48 48 0.00109       0.00129       0.99981
## 49 49 0.00018       0.00019       0.99999
## 50 50 0.00001       0.00001       1.00000
```

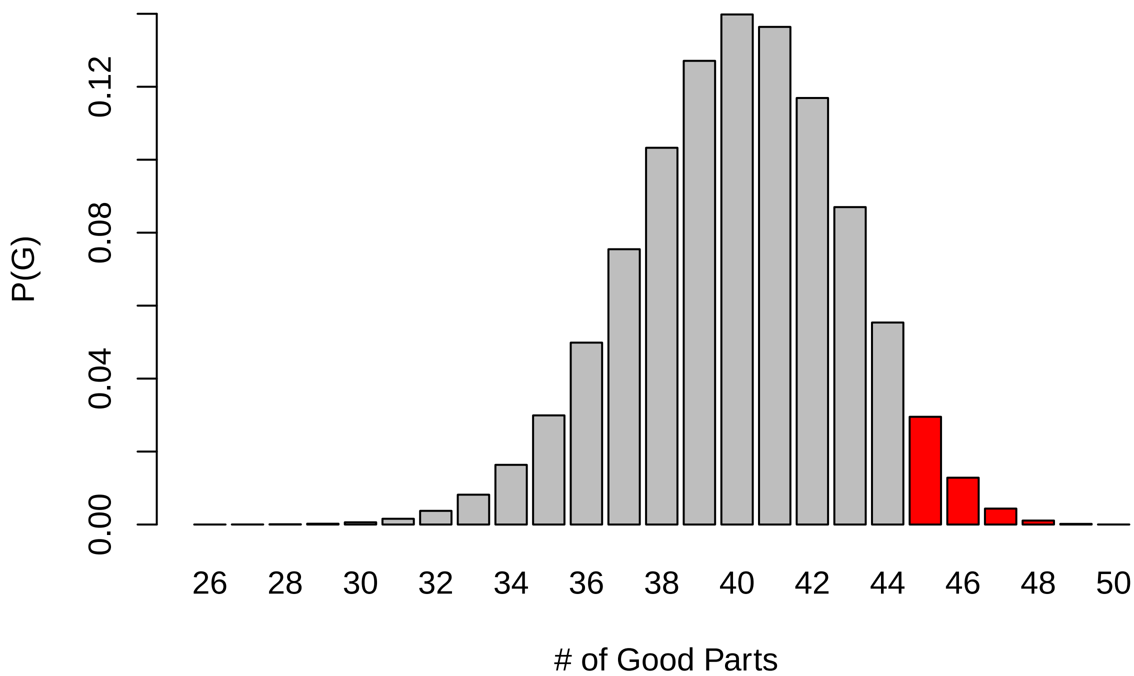
```r
#Barplot of Binomial Prob Distribution

n=50
P=0.8

data = dbinom(x=26:n, size=n, prob=P)
names(data) = 26:n
barplot(data, xlab="# of Good Parts", ylab="P(G", ylim=c(0, 0.15))
```



```r
cols = rep("grey", n+1)
cols[20:25] = "red"
barplot(data, col=cols, xlab=" # of Good Parts", ylab="P(G)", ylim=c(0, 0.14))
```




```r
#Probability of >=45
#Not that pbinom gives P(X>x) for upper tail probs
pbinom(q=44, size=50, prob=0.8, lower.tail=F)
```

```
## [1] 0.04802722
```

The Poisson distribution

- The number of occurences within a time frame

```r
# Get density at exactly that x-value
dpois(x=10, lambda=25)
```

```
## [1] 0.000364985
```

```r
#dpoisson table (lolcat)
#gives density at x for each X
round.object(table.dist.poisson(lambda = 25), 5)
```

```
##      x  p.at.x eq.and.above eq.and.below
## 0    0 0.00000      1.00000      0.00000
## 1    1 0.00000      1.00000      0.00000
## 2    2 0.00000      1.00000      0.00000
## 3    3 0.00000      1.00000      0.00000
```

```
## 4      4 0.00000       1.00000       0.00000
## 5      5 0.00000       1.00000       0.00000
## 6      6 0.00000       1.00000       0.00001
## 7      7 0.00002       0.99999       0.00002
## 8      8 0.00005       0.99998       0.00008
## 9      9 0.00015       0.99992       0.00022
## 10    10 0.00036       0.99978       0.00059
## 11    11 0.00083       0.99941       0.00142
## 12    12 0.00173       0.99858       0.00314
## 13    13 0.00332       0.99686       0.00647
## 14    14 0.00593       0.99353       0.01240
## 15    15 0.00989       0.98760       0.02229
## 16    16 0.01545       0.97771       0.03775
## 17    17 0.02273       0.96225       0.06048
## 18    18 0.03157       0.93952       0.09204
## 19    19 0.04153       0.90796       0.13357
## 20    20 0.05192       0.86643       0.18549
## 21    21 0.06181       0.81451       0.24730
## 22    22 0.07023       0.75270       0.31753
## 23    23 0.07634       0.68247       0.39388
## 24    24 0.07952       0.60612       0.47340
## 25    25 0.07952       0.52660       0.55292
## 26    26 0.07646       0.44708       0.62939
## 27    27 0.07080       0.37061       0.70019
## 28    28 0.06321       0.29981       0.76340
## 29    29 0.05450       0.23660       0.81790
## 30    30 0.04541       0.18210       0.86331
## 31    31 0.03662       0.13669       0.89993
## 32    32 0.02861       0.10007       0.92854
## 33    33 0.02168       0.07146       0.95022
## 34    34 0.01594       0.04978       0.96616
## 35    35 0.01138       0.03384       0.97754
## 36    36 0.00791       0.02246       0.98545
## 37    37 0.00534       0.01455       0.99079
## 38    38 0.00351       0.00921       0.99430
## 39    39 0.00225       0.00570       0.99656
## 40    40 0.00141       0.00344       0.99796
## 41    41 0.00086       0.00204       0.99882
## 42    42 0.00051       0.00118       0.99933
## 43    43 0.00030       0.00067       0.99963
## 44    44 0.00017       0.00037       0.99980
## 45    45 0.00009       0.00020       0.99989
## 46    46 0.00005       0.00011       0.99994
## 47    47 0.00003       0.00006       0.99997
## 48    48 0.00001       0.00003       0.99999
## 49    49 0.00001       0.00001       0.99999
## 50    50 0.00000       0.00001       1.00000
## 51    51 0.00000       0.00000       1.00000
## 52    52 0.00000       0.00000       1.00000
## 53    53 0.00000       0.00000       1.00000
## 54    54 0.00000       0.00000       1.00000
## 55    55 0.00000       0.00000       1.00000
## 56    56 0.00000       0.00000       1.00000
## 57    57 0.00000       0.00000       1.00000
```

```
## 58   58 0.00000      0.00000      1.00000
## 59   59 0.00000      0.00000      1.00000
## 60   60 0.00000      0.00000      1.00000
## 61   61 0.00000      0.00000      1.00000
## 62   62 0.00000      0.00000      1.00000
## 63   63 0.00000      0.00000      1.00000
## 64   64 0.00000      0.00000      1.00000
## 65   65 0.00000      0.00000      1.00000
## 66   66 0.00000      0.00000      1.00000
## 67   67 0.00000      0.00000      1.00000
## 68   68 0.00000      0.00000      1.00000
## 69   69 0.00000      0.00000      1.00000
## 70   70 0.00000      0.00000      1.00000
## 71   71 0.00000      0.00000      1.00000
## 72   72 0.00000      0.00000      1.00000
## 73   73 0.00000      0.00000      1.00000
## 74   74 0.00000      0.00000      1.00000
## 75   75 0.00000      0.00000      1.00000
## 76   76 0.00000      0.00000      1.00000
## 77   77 0.00000      0.00000      1.00000
## 78   78 0.00000      0.00000      1.00000
## 79   79 0.00000      0.00000      1.00000
## 80   80 0.00000      0.00000      1.00000
## 81   81 0.00000      0.00000      1.00000
## 82   82 0.00000      0.00000      1.00000
## 83   83 0.00000      0.00000      1.00000
## 84   84 0.00000      0.00000      1.00000
## 85   85 0.00000      0.00000      1.00000
## 86   86 0.00000      0.00000      1.00000
## 87   87 0.00000      0.00000      1.00000
## 88   88 0.00000      0.00000      1.00000
## 89   89 0.00000      0.00000      1.00000
## 90   90 0.00000      0.00000      1.00000
## 91   91 0.00000      0.00000      1.00000
## 92   92 0.00000      0.00000      1.00000
## 93   93 0.00000      0.00000      1.00000
## 94   94 0.00000      0.00000      1.00000
## 95   95 0.00000      0.00000      1.00000
## 96   96 0.00000      0.00000      1.00000
## 97   97 0.00000      0.00000      1.00000
## 98   98 0.00000      0.00000      1.00000
## 99   99 0.00000      0.00000      1.00000
## 100 100 0.00000      0.00000      1.00000
## 101 101 0.00000      0.00000      1.00000
## 102 102 0.00000      0.00000      1.00000
## 103 103 0.00000      0.00000      1.00000
## 104 104 0.00000      0.00000      1.00000
## 105 105 0.00000      0.00000      1.00000
## 106 106 0.00000      0.00000      1.00000
## 107 107 0.00000      0.00000      1.00000
## 108 108 0.00000      0.00000      1.00000
## 109 109 0.00000      0.00000      1.00000
## 110 110 0.00000      0.00000      1.00000
## 111 111 0.00000      0.00000      1.00000
```

```
## 112 112 0.00000      0.00000      1.00000
## 113 113 0.00000      0.00000      1.00000
## 114 114 0.00000      0.00000      1.00000
## 115 115 0.00000      0.00000      1.00000
## 116 116 0.00000      0.00000      1.00000
## 117 117 0.00000      0.00000      1.00000
## 118 118 0.00000      0.00000      1.00000
## 119 119 0.00000      0.00000      1.00000
## 120 120 0.00000      0.00000      1.00000
## 121 121 0.00000      0.00000      1.00000
## 122 122 0.00000      0.00000      1.00000
## 123 123 0.00000      0.00000      1.00000
## 124 124 0.00000      0.00000      1.00000
## 125 125 0.00000      0.00000      1.00000
```

```
round.object(table.dist.poisson(lambda = 25)[7:51,], 5)
```

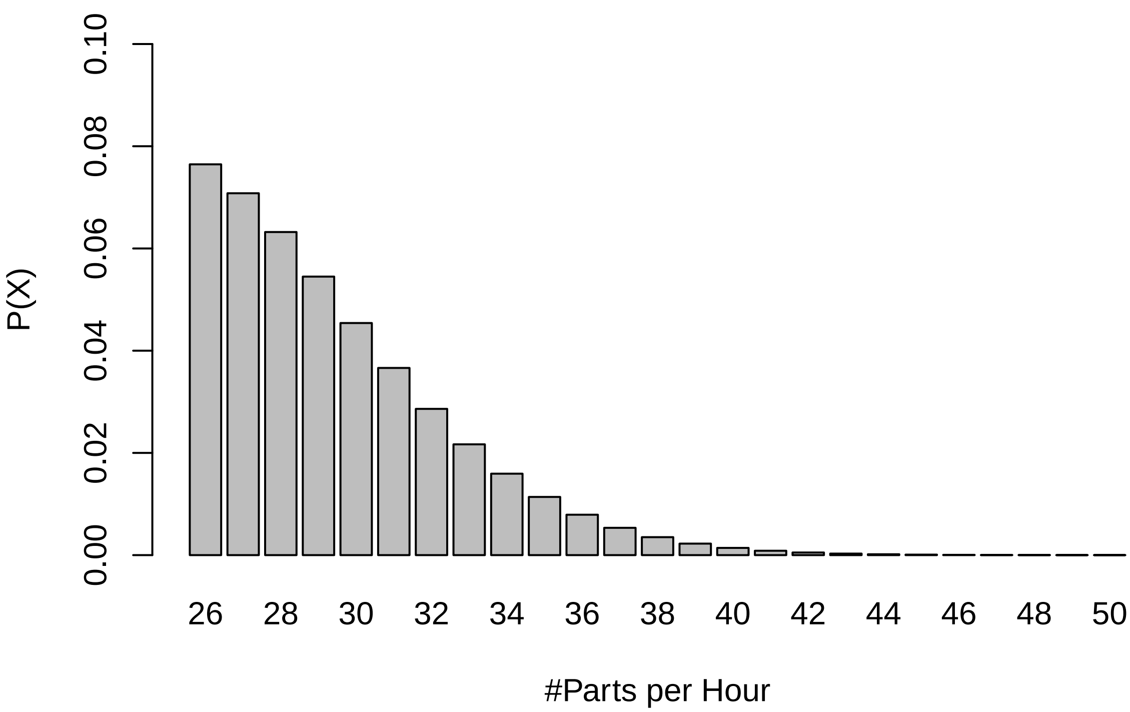
```
##     x  p.at.x eq.and.above eq.and.below
## 6   6 0.00000      1.00000      0.00001
## 7   7 0.00002      0.99999      0.00002
## 8   8 0.00005      0.99998      0.00008
## 9   9 0.00015      0.99992      0.00022
## 10 10 0.00036      0.99978      0.00059
## 11 11 0.00083      0.99941      0.00142
## 12 12 0.00173      0.99858      0.00314
## 13 13 0.00332      0.99686      0.00647
## 14 14 0.00593      0.99353      0.01240
## 15 15 0.00989      0.98760      0.02229
## 16 16 0.01545      0.97771      0.03775
## 17 17 0.02273      0.96225      0.06048
## 18 18 0.03157      0.93952      0.09204
## 19 19 0.04153      0.90796      0.13357
## 20 20 0.05192      0.86643      0.18549
## 21 21 0.06181      0.81451      0.24730
## 22 22 0.07023      0.75270      0.31753
## 23 23 0.07634      0.68247      0.39388
## 24 24 0.07952      0.60612      0.47340
## 25 25 0.07952      0.52660      0.55292
## 26 26 0.07646      0.44708      0.62939
## 27 27 0.07080      0.37061      0.70019
## 28 28 0.06321      0.29981      0.76340
## 29 29 0.05450      0.23660      0.81790
## 30 30 0.04541      0.18210      0.86331
## 31 31 0.03662      0.13669      0.89993
## 32 32 0.02861      0.10007      0.92854
## 33 33 0.02168      0.07146      0.95022
## 34 34 0.01594      0.04978      0.96616
## 35 35 0.01138      0.03384      0.97754
## 36 36 0.00791      0.02246      0.98545
## 37 37 0.00534      0.01455      0.99079
## 38 38 0.00351      0.00921      0.99430
## 39 39 0.00225      0.00570      0.99656
## 40 40 0.00141      0.00344      0.99796
## 41 41 0.00086      0.00204      0.99882
```

```
## 42 42 0.00051       0.00118       0.99933
## 43 43 0.00030       0.00067       0.99963
## 44 44 0.00017       0.00037       0.99980
## 45 45 0.00009       0.00020       0.99989
## 46 46 0.00005       0.00011       0.99994
## 47 47 0.00003       0.00006       0.99997
## 48 48 0.00001       0.00003       0.99999
## 49 49 0.00001       0.00001       0.99999
## 50 50 0.00000       0.00001       1.00000
```

```r
#Barplot of Poisson Prob Distribution

lambda=25
x = 10

data = dpois(x=26:50, lambda=lambda)
names(data) = 26:50
barplot(data, xlab="#Parts per Hour", ylab="P(X)", ylim=c(0, 0.10))
```




```r
#Probability of >=45
#Note gives P(X>x) for upper tail probs
pbinom(q=44, size=50, prob=0.8, lower.tail=F)
```

```
## [1] 0.04802722
```