

# 605-HW07-Distributions

*Michael Y.*

*October 13, 2019*

## Contents

<b>HW7 - Distributions</b>	<b>2</b>
Problem 1. Distribution of $\min(\text{Uniform})$	2
$Pr(Y = 1)$	2
$Pr(Y = 2)$	3
$Pr(Y = 3)$	3
General formula for $Pr(Y = y), \forall y \in [1, k]$	3
Problem 2. Failure after 8 years	4
a. Geometric	4
b. Exponential	11
c. Binomial	12
d. Poisson	15

## HW7 - Distributions

### Problem 1. Distribution of min(Uniform)

Let  $X_1, X_2, \dots, X_n$  be  $n$  mutually independent random variables, each of which is uniformly distributed on the integers from 1 to  $k$ .

Let  $Y$  denote the minimum of the  $X_i$ s.

Find the distribution of  $Y$ .

First, note that the variables are *discrete*, taking on *integer* values such that  $X_i \in [1, k]$ .

The question is asking for the distribution of the **order statistic**  $Y = \min(X_i) = X_{(1)}$ .

$$Pr(Y = 1)$$

**First, determine the probability that the minimum is equal to 1:**

Because each of the  $X_i$  is distributed uniformly on  $[1, k]$ , the probability that any individual  $X_i$  equals 1 is  $\frac{1}{k}$ , and thus the probability that any individual  $X_i$  is greater than 1 is  $\frac{k-1}{k}$ .

$$\begin{aligned} Pr(X_i = 1) &= \frac{1}{k}, \quad \text{for each } i, 1 \leq i \leq n \\ Pr(X_i > 1) &= \frac{k-1}{k}, \quad \text{for each } i, 1 \leq i \leq n \end{aligned}$$

In order for the **minimum** of all  $X_i$  to be greater than 1, this requires that **all**  $X_i$  are greater than 1, which would happen with probability  $\left(\frac{k-1}{k}\right)^n$ .

$$Pr(Y > 1) = Pr(\min(X_i) > 1) = Pr(X_1 > 1 ; X_2 > 1 ; \dots ; X_n > 1) = \left(\frac{k-1}{k}\right)^n$$

Therefore,

$$Pr(Y = 1) = 1 - Pr(Y > 1) = 1 - \left(\frac{k-1}{k}\right)^n$$

.

Note that this can be written as

$$\left(\frac{k-0}{k}\right)^n - \left(\frac{k-1}{k}\right)^n$$

$$Pr(Y = 2)$$

**Next, determine the probability that the minimum is equal to 2:**

The probability that any individual  $X_i$  is greater than 2 is  $\frac{k-2}{k}$ , so the probability that  $X_i > 2, \forall i$  is  $\left(\frac{k-2}{k}\right)^n = Pr(\min(X_i) > 2) = Pr(Y > 2)$ .

So, the probability that the minimum is **equal** to 2 is

$$\begin{aligned} Pr(Y = 2) &= 1 - Pr(Y > 2) - Pr(Y = 1) \\ &= 1 - \left(\frac{k-2}{k}\right)^n - \left[1 - \left(\frac{k-1}{k}\right)^n\right] \\ &= \left(\frac{k-1}{k}\right)^n - \left(\frac{k-2}{k}\right)^n \end{aligned}$$

$$Pr(Y = 3)$$

**Next, determine the probability that the minimum is equal to 3:**

$$\begin{aligned} Pr(Y = \min(X_i) = 3) &= 1 - Pr(Y > 3) - Pr(Y = 2) - Pr(Y = 1) \\ &= 1 - \left(\frac{k-3}{k}\right)^n - \left[\left(\frac{k-1}{k}\right)^n - \left(\frac{k-2}{k}\right)^n\right] - \left[1 - \left(\frac{k-1}{k}\right)^n\right] \\ &= 1 - \left(\frac{k-3}{k}\right)^n - \left(\frac{k-1}{k}\right)^n + \left(\frac{k-2}{k}\right)^n - 1 + \left(\frac{k-1}{k}\right)^n \\ &= \left(\frac{k-2}{k}\right)^n - \left(\frac{k-3}{k}\right)^n \end{aligned}$$

**General formula for  $Pr(Y = y), \forall y \in [1, k]$  :**

$$Pr(Y = \min(X_i) = y) = \left(\frac{k-y+1}{k}\right)^n - \left(\frac{k-y}{k}\right)^n$$

## Problem 2. Failure after 8 years

Your organization owns a copier (future lawyers, etc.) or MRI (future doctors).

This machine has a manufacturer's expected lifetime of 10 years.

This means that we expect one failure every ten years.

(Include the probability statements and R Code for each part.).

Probability of 1 failure in 10 years  $\rightarrow$  Probability of failure in 1 year is  $1/10 = 10\% = 0.1$ .

### a. Geometric

What is the probability that the machine will fail after 8 years?.

Provide also the expected value and standard deviation.

Model as a geometric.

(Hint: the probability is equivalent to not failing during the first 8 years..)

```
Annual = 10
pAnnual = 1/Annual      # 0.1
qAnnual = 1-pAnnual     # 0.9
YearsToFail = 8
pgeom(YearsToFail-1,pAnnual) # 0.56953279
```

```
## [1] 0.56953279
```

```
1-qAnnual^YearsToFail      # 0.56953279
```

```
## [1] 0.56953279
```

```
qAnnual~YearsToFail # 0.43046721
```

```
## [1] 0.43046721
```

Annual table - geometric

```
# annual table
cbind(year=0:10,
      Prob_Fail=pgeom(0:10,pAnnual),
      Prob_Not_fail=pgeom(0:10,pAnnual,lower.tail=F)) %>%
kable() %>% kable_styling(c("striped", "bordered"))
```

year	Prob_Fail	Prob_Not_fail
0	0.1000000000	0.9000000000
1	0.1900000000	0.8100000000
2	0.2710000000	0.7290000000
3	0.3439000000	0.6561000000
4	0.4095100000	0.5904900000
5	0.4685590000	0.5314410000
6	0.5217031000	0.4782969000
7	0.5695327900	0.4304672100
8	0.6125795110	0.3874204890
9	0.6513215599	0.3486784401
10	0.6861894039	0.3138105960

$p = 0.1$  ;  $q = 1 - p = 0.9$  ;  $Pr(X = n) = p \cdot q^{(n-1)}$

Probability of failing within the first 8 years (where the first year is enumerated by 0, the second year by 1, ... the eighth year by 7):

$$Pr(X < 8) = \sum_{i=0}^7 p \cdot q^i = p \cdot \sum_{i=0}^7 q^i = p \left[ \frac{1 - q^8}{1 - q} \right] = p \left[ \frac{1 - q^8}{p} \right] = 1 - q^8 = 1 - (0.9)^8 = 1 - .43046721 = 0.56953279$$

This is pgeom(7,1/10):

Probability of failing within the first 8 years

```
pgeom(YearsToFail-1,pAnnual)
```

```
## [1] 0.56953279
```

Therefore, the probability of NOT failing within the first 8 years is

$$Pr(X \geq 8) = 1 - Pr(X < 8) = 1 - (1 - q^8) = q^8 = (0.9)^8 = 0.43046721 .$$

This is  $1 - \text{pgeom}(\text{YearsToFail}-1, \text{pAnnual}) = 1 - \text{pgeom}(\text{YearsToFail}-1, \text{pAnnual}, \text{lower.tail}=\text{FALSE})$  :

```
1-pgeom(YearsToFail-1,pAnnual)
```

```
## [1] 0.43046721
```

```
pgeom(YearsToFail-1,pAnnual,lower.tail=FALSE)
```

```
## [1] 0.43046721
```

The formula for the expected value of a geometric distribution where there are k failures is  $E[x] = \mu = \frac{(1-p)}{p} = \frac{q}{p}$

Expected val (Annual) - units in years

```
expectedval_Annual = qAnnual/pAnnual  
expectedval_Annual
```

```
## [1] 9
```

The formula for the variance of a geometric distribution where there are k failures is  $E[x] = \mu = \frac{(1-p)}{p^2} = \frac{q}{p^2}$

```
# Variance_Annual (units in years^2)  
variance_Annual = qAnnual/(pAnnual^2)  
variance_Annual
```

```
## [1] 90
```

```
# Stdev_Annual (units in years)
stdev_Annual = sqrt(variance_Annual)
stdev_Annual
```

```
## [1] 9.48683298051
```

This result indicates that the expected time to failure is 9 years – however, we have made the assumption that failure could only occur at the beginning of each year. (i.e., when performing the averaging for the expected value, the probability of failure in the first year is multiplied by zero, which implied that such failure occurs IMMEDIATELY, rather than at some random time during the year.)

To be more realistic, it would be better to assume that the failure could occur any time within each year, which averages out to be at the middle of the year, making the expected time to fail closer to 9.5.

However, we expect to obtain a value of 10 because that was the initial time-to-failure included in the problem.

To get this, we need to take smaller timesteps:

### Daily - geometric

Consider daily rather than annual. Then:

```
DaysInYear = 365.25
Daily = Annual * DaysInYear      # 3652.5
pDaily = 1/Daily                  # 0.1 / 365.25 = 0.000273785078713
qDaily = 1-pDaily                 # 0.9 / 365.25 = 0.999726214921
DaysToFail = YearsToFail * DaysInYear # 8 * 365.25 = 2922

pgeom(DaysToFail-1,pDaily) # 0.550720249997 for daily, rather than 0.56953279 for annual
```

```
## [1] 0.550720249997
```

```
1-qDaily^DaysToFail          # 0.550720249997 for daily, rather than 0.56953279 for annual
```

```
## [1] 0.550720249997
```

```
qDaily^DaysToFail           # 0.449279750003 for daily, rather than 0.43046721 for annual
```

```
## [1] 0.449279750003
```

Probability of failing within the first 8 years, when possible failure is assessed daily:

```
pgeom(DaysToFail-1,pDaily)
```

```
## [1] 0.550720249997
```

Probability of NOT failing within the first 8 years, when possible failure is assessed daily, is

$1 - \text{pgeom}(\text{DaysToFail}-1, \text{pDailyAnnual}) = 1 - \text{pgeom}(\text{DaysToFail}-1, \text{pDaily}, \text{lower.tail}=\text{FALSE})$  :

```
1-pgeom(DaysToFail-1,pDaily)
```

```
## [1] 0.449279750003
```



```
pgeom(DaysToFail-1,pDaily,lower.tail=FALSE)
```

```
## [1] 0.449279750003
```

The formula for the expected value of a geometric distribution where there are k failures is  $E[x] = \mu = \frac{(1-p)}{p} = \frac{q}{p}$

**Expected time until failure, when failure can occur daily:**

```
# Expected days until failure:  
expectedval_Daily = qDaily/pDaily  
expectedval_Daily
```

```
## [1] 3651.5
```

```
# (This is one day less than 10 years.)  
  
# Expected Years until failure:  
expectedval_Daily_in_Years = expectedval_Daily / DaysInYear  
expectedval_Daily_in_Years
```

```
## [1] 9.99726214921
```

This is much closer to the 10-year result that we were expecting.

The formula for the variance of a geometric distribution where there are k failures is  $E[x] = \mu = \frac{(1-p)}{p^2} = \frac{q}{p^2}$

Variance - geometric - daily failures possible

```
# Variance_Daily (units in days^2)  
variance_Daily = qDaily/(pDaily^2)  
variance_Daily
```

```
## [1] 13337103.75
```

```
# Stdev_Daily (units in days)  
stdev_Daily = sqrt(variance_Daily)  
stdev_Daily
```

```
## [1] 3651.99996577
```

```
# Stdev Daily (units expressed in years)  
stdev_Daily_in_Years = stdev_Daily / DaysInYear  
stdev_Daily_in_Years
```

```
## [1] 9.9986309809
```

---

## b. Exponential

What is the probability that the machine will fail after 8 years?.

Provide also the expected value and standard deviation.

### Model as an exponential.

The exponential distribution is quite similar to the geometric distribution, except it is continuous, while the geometric distribution is discrete.

The PDF is

$$f(x; \lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

$$\text{and the CDF is } F(x; \lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases}$$

Because the expected life of the device is 10 years,  $\lambda = \frac{1}{10} = 0.1$  .

The probability that the device does **not** fail in the first 8 years is

```
pexp(YearsToFail,rate=pAnnual,lower.tail=TRUE)
```

```
## [1] 0.550671035883
```

so the probability that it fails AFTER 8 years is

```
pexp(YearsToFail,rate=pAnnual,lower.tail=FALSE)
```

```
## [1] 0.449328964117
```

For the exponential distribution, the expected value is  $E[X] = \frac{1}{\lambda} = 10$  and the variance is  $VAR[X] = \frac{1}{\lambda^2} = 100$  .

Thus, the standard deviation of the exponential is  $SD[X] = \sqrt{VAR[X]} = \sqrt{100} = 10$  .

### c. Binomial

What is the probability that the machine will fail after 8 years?.

Provide also the expected value and standard deviation.

Model as a binomial.

(Hint: 0 success in 8 years)

For the Binomial distribution, we again face the granularity problem that became apparent above when looking at the geometric distribution.

The probability mass function for the binomial is  $f(k, n, p) = \Pr(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$

If we are again considering annual results, we have zero successes in 8 years, where the probability is  $p = \frac{1}{10} = 0.1$  .

This gives:  $f(0, 8, 0.10) = \Pr(0; 8, 0.10) = \Pr(X = 0) = \binom{8}{0} (0.10)^0 (1 - 0.10)^{8-0} = (0.9)^8 = 0.43046721$

which is the same result as obtained from the Geometric above (under annual failures.)

using pbinom for annual failures:

```
pbinom(0,8,1/10)
```

```
## [1] 0.43046721
```

When possible failures are only considered on an annual basis,

The expected value of the binomial distribution is  $E[X] = n \cdot p = 8 \cdot \frac{1}{10} = 0.8$

and the variance of the binomial is  $VAR[X] = n \cdot p \cdot (1 - p) = n \cdot p \cdot q = 8 \cdot 0.1 \cdot 0.9 = 0.72$  .

Thus, the standard deviation is  $SD[X] = \sqrt{VAR[X]} = \sqrt{0.72} = 0.848528137424$  .

To obtain greater granularity, we could consider daily, rather than annual, opportunities for failure.

## Binomial - daily

Again, we have

```
DaysInYear = 365.25
Daily = Annual * DaysInYear      # 3652.5
pDaily = 1/Daily                  # 0.1 / 365.25 = 0.000273785078713
qDaily = 1-pDaily                 # 0.9 / 365.25 = 0.999726214921
DaysToFail = YearsToFail * DaysInYear  # 8 * 365.25 = 2922

#### pbinom(k,n,p)
pbinom(0,DaysToFail,pDaily) # 0.449279750003 for daily, rather than 0.43046721 for annual

## [1] 0.449279750003
```

This is the same value as obtained above for the geometric distribution, when possible daily (rather than annual) failures are considered.

As the time interval becomes smaller, we eventually approach continuous time, as measured by the exponential model.

If we instead perform the above calculations on an hourly, minutely, or secondly basis, the results should converge.

### Expected value - binomial - daily

The expected value of the binomial distribution is  $E[X] = n \cdot p$ , which here is  $2922 \cdot 0.000273785079 = 0.8$ .

This is unchanged from the annual value.

```
DaysToFail*pDaily
```

```
## [1] 0.8
```

### Variance and standard deviation - binomial - daily

The Daily variance of the binomial is

$$\begin{aligned} \text{VAR}[X] &= (n * 365.25) \cdot \left(\frac{p}{365.25}\right) \cdot \left(1 - \frac{p}{365.25}\right) \\ &= 2922 \cdot 0.000273785079 \cdot 0.999726214921 \\ &= 0.799780971937 \end{aligned}$$

```
# binomial - variance (daily basis)
binomDailyVAR = DaysToFail*pDaily*qDaily
binomDailyVAR
```

```
## [1] 0.799780971937
```

```
# binomial - standard deviation (daily basis)
binomDailySD = sqrt(binomDailyVAR)
binomDailySD
```

```
## [1] 0.894304742209
```

Thus, the standard deviation is  $SD[X] = \sqrt{\text{VAR}[X]} = \sqrt{0.799780971937} = 0.894304742209$ .

#### d. Poisson

What is the probability that the machine will fail after 8 years?.

Provide also the expected value and standard deviation.

**Model as a Poisson.**

The probability mass function for the Poisson distribution is  $f(k; \lambda) = \Pr(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$

We seek to know whether no failures have occurred during the first 8 years, or  $\Pr(X = 0|t = 8; \lambda = \frac{1}{10} = 0.10)$

The probability mass function for the Poisson distribution across non-unit time is  $f(k; \lambda t) = \Pr(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$

Since here we are fixing  $k = 0$ , the above simply becomes

$$f(0; \lambda t) = \Pr(X = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t} = e^{-(0.1)8} = e^{-0.8} = 0.449328964117$$

This can be computed in r using `dpois()` :

```
YearsToFail
```

```
## [1] 8
```

```
pAnnual
```

```
## [1] 0.1
```

```
lambda = YearsToFail * pAnnual  
lambda
```

```
## [1] 0.8
```

```
dpois(0,lambda)
```

```
## [1] 0.449328964117
```

Note that this result matches that from the exponential calculation above.

### **Poisson mean, variance, standard deviation**

For the poisson distribution, the mean is  $E[X] = \mu = \lambda = 0.8$  and the variance is also  $VAR[X] = \lambda = 0.8$  .

Therefore, the standard deviation is  $SD[X] = \sqrt{VAR[X]} = \sqrt{0.8} = 0.894427191$  .