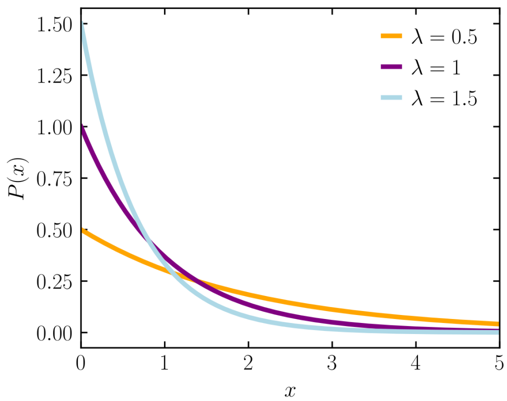
Homework exercise

To be solved at home before the exercise session.

1. a. Let $x_1, \ldots, x_n$ be a random sample (iid) from some distribution F_θ with the unknown parameter θ. Which of the three one-sample tests (t-test, sign test or signed rank test) would you use (and why!) to test whether the location (expected value/median) of the data is equal to 1 if we know for certain that the distribution F_θ is
   i. an exponential distribution with unknown rate parameter θ,
   ii. a normal distribution with variance 2 and unknown expected value θ,
   iii. a Laplace distirbution with known scale parameter 5 and unknown location parameter θ,
   iv. a Poisson distirbution with unknown parameter θ?

a) Three test types: t-test, sign test, signed rank test
i. an exponential distribution with unknown rate parameter

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



where

$\lambda =$ Rate Parameter

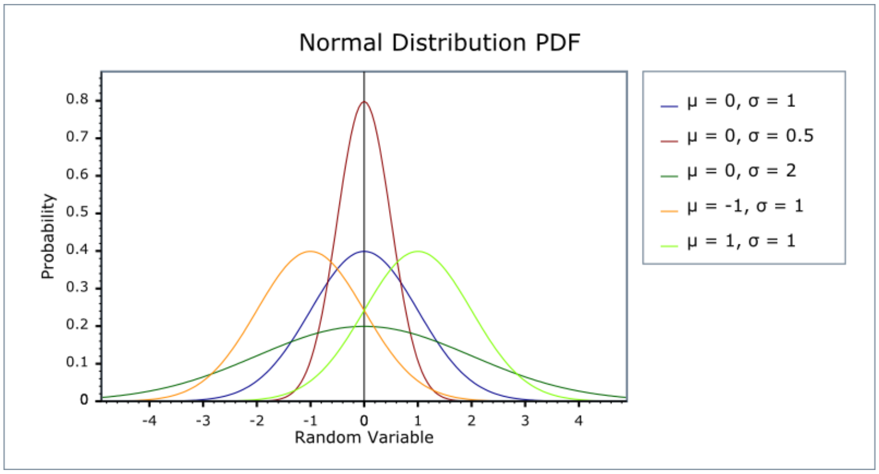
Unknown parameter theta = lambda (rate parameter)
sign test should be used as test statistics for this distribution
Reason: sign test is used when the distribution assumption is continuous distribution. Since exponential distribution is a continuous distribution but not necessarily symmetric, the best test statistics would be sign test

ii. a normal distribution with variance 2 and unknown expected value

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$\begin{aligned} \mu &= \text{ Mean} \\ \sigma &= \text{ Standard Deviation} \\ \pi &\approx 3.14159\cdots \\ e &\approx 2.71828\cdots \end{aligned}$$

Unknown parameter theta = mu (expected value)



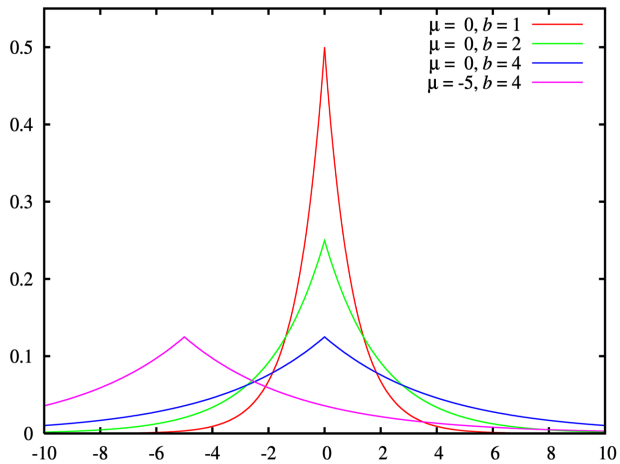
t-test should be used as test statistics for this distribution
Reason: t-test is used when normality is assumed. Since this is normal distribution, the best test statistics would be t-test

iii. a Laplace distirbution with known scale parameter 5 and unknown location parameter

Laplace Distribution: Definition

- Introduced by Pierre-Simon Laplace
- It is a continuous probability distribution
- Also known as *double exponential distribution or Gumbel distribution*
- Govern the difference between the two identical distributed variable
- Increase of Laplace motion or a variance gamma process calculated over the time represents Laplace distribution
- A variable shows Laplace(μ, b) distribution when the density function is defined as

$$f(x \mid \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right) \quad = \frac{1}{2b} \begin{cases} \exp\left(-\frac{\mu - x}{b}\right) & \text{if } x < \mu \\ \exp\left(-\frac{x-\mu}{b}\right) & \text{if } x \geq \mu \end{cases}$$



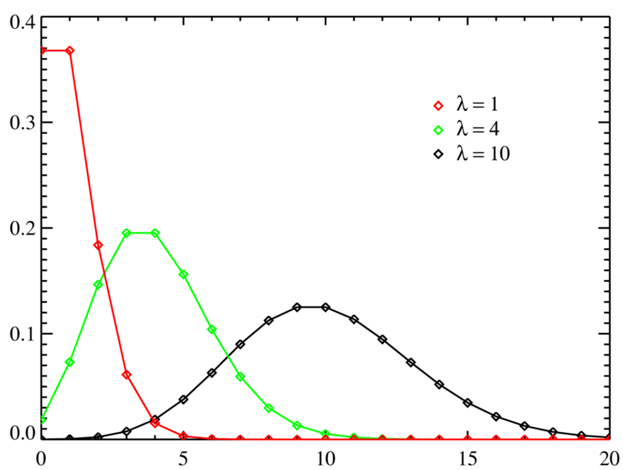
Unknown parameter theta = b
signed rank test should be used as test statistics for this distribution
Reason: signed rank test is used when the distribution assumption is continuous symmetric distribution. Since laplace distribution is a continuous symmetric distribution around the mean, the best test statistics would be signed-rank test

iv. a Poisson distirbution with unknown parameter

Poisson Probability Distribution

$$P(x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

λ- mean number of successes over a given interval
$Var(X) = \lambda$



Unknown parameter theta = lambda
sign test should be used as test statistics for this distribution
Reason: sign test is used when the distribution assumption is continuous distribution. Since lambda distribution is a continuous distribution but not necessarily symmetric, the best test statistics would be sign test

b. The data set `airmiles` lists the passenger miles flown by commercial airlines in the United States for each year from 1937 to 1960. To inspect whether the yearly passenger miles equal 10000 on average, a researcher performed a sign test to test the null hypothesis $med_x = 10000$ on significance level 5% with the results shown below and concluded that there is no evidence against the null hypothesis. Do you agree with the researcher's conclusion?

```
airmiles
```

```
## Time Series:
## Start = 1937
## End = 1960
## Frequency = 1
##  [1]   412   480   683  1052  1385  1418  1634  2178  3362  5948  6109
## [12]  5981  6753  8003 10566 12528 14760 16769 19819 22362 25340 25343
## [23] 29269 30514
```

```
# Sign test
binom.test(sum(airmiles > 10000), length(airmiles))
```

```
##
##  Exact binomial test
##
## data:  sum(airmiles > 10000) and length(airmiles)
## number of successes = 10, number of trials = 24, p-value = 0.5413
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2210969 0.6335694
## sample estimates:
## probability of success
##              0.4166667
```

b) Yes, I agree with the researchers' conclusion
1st, this time series has only 24 samples, which is quite small and there is no normality assumed, so it is correct to use the sign test for this data.
2nd, the obtained p-value is 0.5, which is much larger than alpha = 0.05. A p-value higher than 0.05 (> 0.05) is not statistically significant and indicates strong evidence for the null hypothesis
3rd, 95 percent CI contains probability of success = 0.5 and for the sample estimates, it calculates as 0.41, which is quite close to 0.5 probability, which could be the real value of the chance that mean number of miles is indeed about 10000