

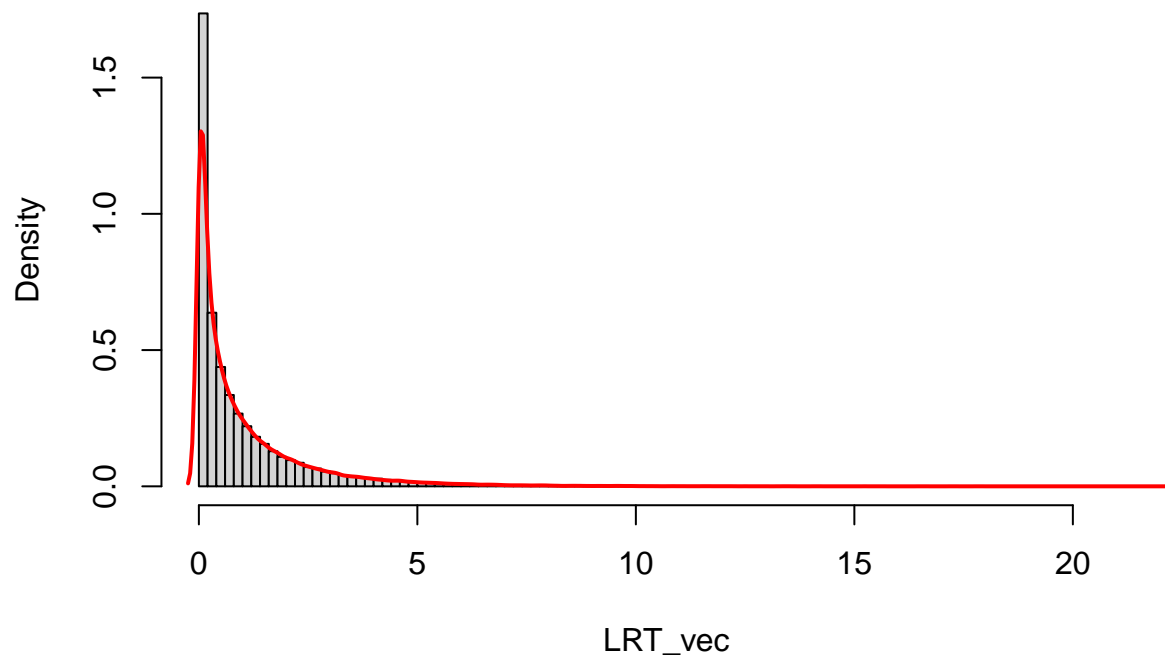
# Demonstration of the chi-squared approximation using Likelihood Ratio Test

## Part One

Let  $X_1, X_2, \dots, X_5 \stackrel{iid}{\sim} N(\mu, \sigma^2 = 10)$  be an independent, identically distributed samples from a Normal Distribution where  $\sigma^2 = 10$  is a *known* constant.

We want to test the null hypothesis  $H_0 : \mu = 10$  vs  $H_1 : \mu \neq 10$  at level of significance,  $\alpha$ .

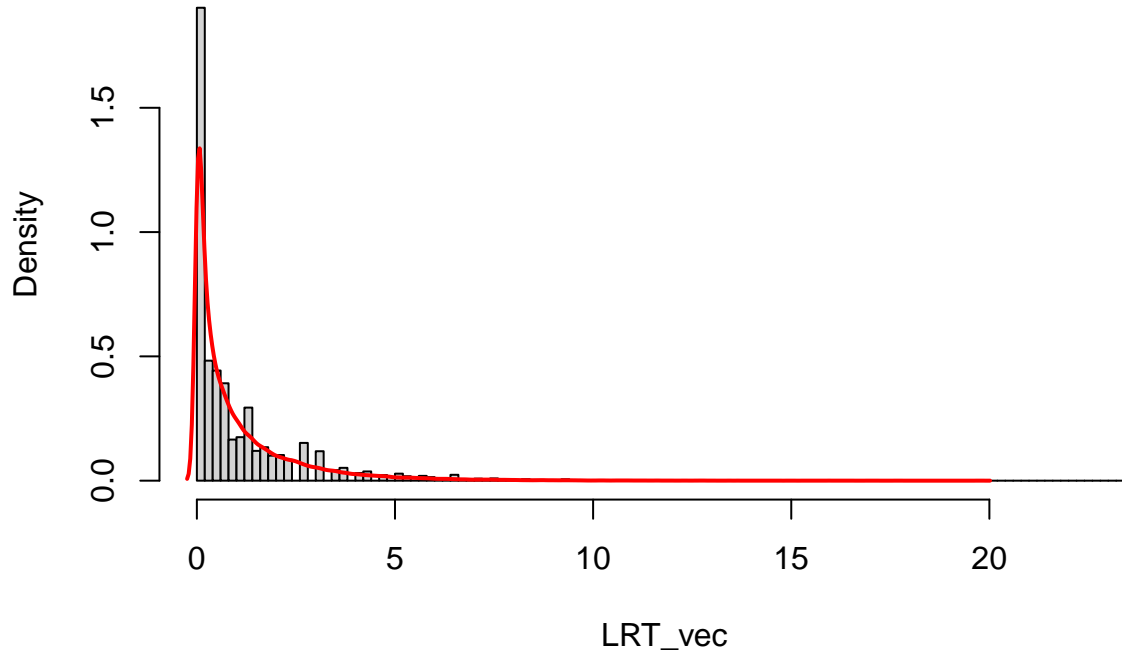
**Histogram of LRT\_vec**



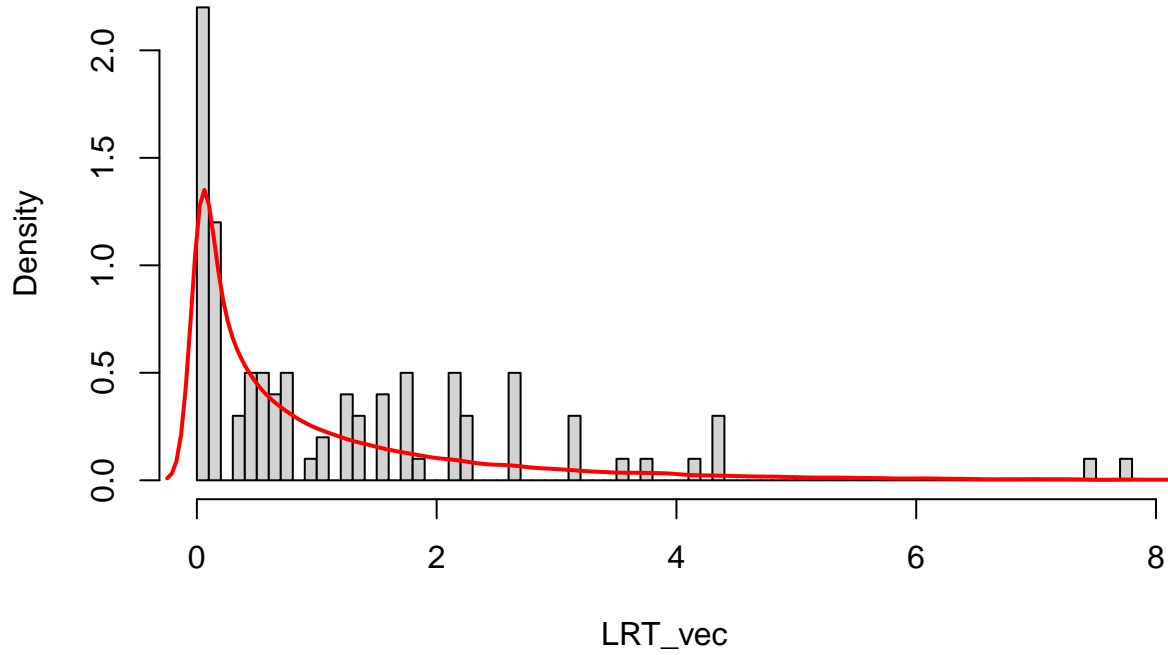
### Part Two

Let  $X_1, X_2, \dots, X_5 \stackrel{iid}{\sim} Pois(\lambda)$  be an independent, identically distributed samples from a Poisson Distribution  
We want to test the null hypothesis  $H_o : \lambda = 10$  vs  $H_1 : \lambda \neq 10$  at level of significance,  $\alpha$ .

### Large sample



## Small sample



### Part Three

When samples are drawn from Normal distribution, the test statistic follows  $\chi^2_{(df=1)}$  and therefore we see no difference between the chi-square line and the histogram. However this not true for samples with non-normal distributions, those samples are **approximated** and **converge** to the test statistic  $\chi^2_{(df=1)}$ , this is observed by the plots in part two. As we increase  $n$ , both the histogram and the density curve become more smooth as shown on the plot above. When  $n$  was small, the density does not match the histogram, it was only when we started to increase  $n$ , the density started to smooth out and match the histogram, the same observation was made for when  $n$  was large for the density curve but small for the histogram. In conclusion,  $n$  does play a role on how close the histogram and the density curve are to each other but only if the samples come from a non-normal distribution.

```

#code for plot in part One
LRT_normal=function(){
x=rnorm(5,mean=10,sd=sqrt(10))
L_theta0=prod(dnorm(x,mean=10,sd=sqrt(10)))
L_theta1=prod(dnorm(x,mean=mean(x),sd=sqrt(10)))
return(-2*log(L_theta0/L_theta1))
}

LRT_vec=replicate(100000, LRT_normal())

hist(LRT_vec,breaks=100,freq = FALSE)
lines(density(rchisq(100000,df=1)), type = "l", col="red", lwd=2.0)

```

```

#code for first plot in part Two
LRT_pois=function(){
x=rpois(5,lambda=10)
L_theta0=prod(dpois(x,lambda=10))
L_theta1=prod(dpois(x,lambda=mean(x)))
return(-2*log(L_theta0/L_theta1))
}

LRT_vec=replicate(100000, LRT_pois())
hist(LRT_vec,breaks=100,freq = FALSE,main="Large sample")
X = rchisq(100000, df=1)
lines(density(X), type = "l", col="red", lwd=2.0)

```

```

#code for second plot in part Two
LRT_pois=function(){
x=rpois(5,lambda=10)
L_theta0=prod(dpois(x,lambda=10))
L_theta1=prod(dpois(x,lambda=mean(x)))
return(-2*log(L_theta0/L_theta1))
}

LRT_vec=replicate(100, LRT_pois())
hist(LRT_vec,breaks=100,freq = FALSE,main="Small sample")
X = rchisq(100000, df=1)
lines(density(X), type = "l", col="red", lwd=2.0)

```