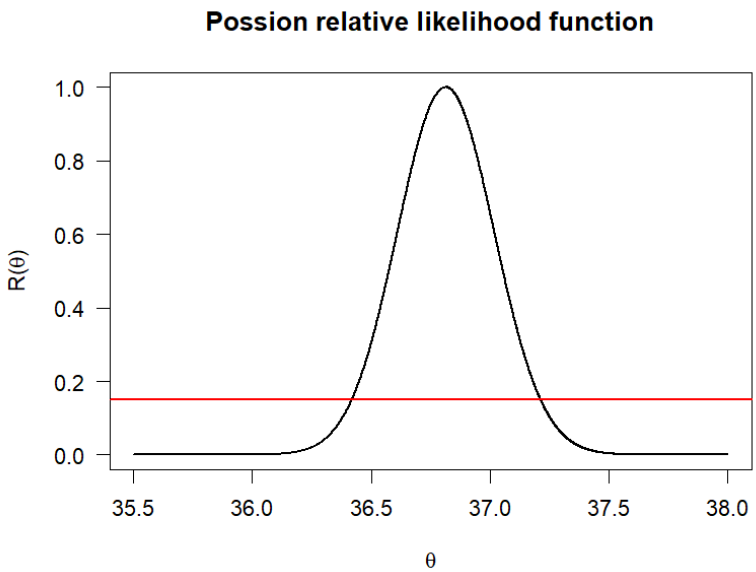
I focus on interval estimation techniques, and expand our analyses of three variates:

- subject.sex: Comparing the distribution of sex of traffic stop subjects between the two cities, and seeing if it is similar, or different, to the broader populations.
- subject.age: Exploring the average age of traffic stop subjects, where we recall that population estimates place the United States median age at approximately 39 years old.
- lat and lng: Examining the location of traffic stops, and seeing if on average they occur near the centre of each city.

The sample size is 883, the maximum likelihood estimate of lambda is 36.814.

Relative likelihood function plot:



The 15% likelihood interval for lambda is [36.42, 37.21].

The approximate 15%, 90% and 95% confidence intervals for lambda are [36.78, 36.85], [36.48, 37.15], and [36.41, 37.21], respectively. These were calculated by:

$\hat{\theta} \pm z_{(\alpha/2)} * \sqrt{(\hat{\theta} / n)}$

Where:

$\hat{\theta}$ is the sample mean, which is equal to λ.

$z_{(\alpha/2)}$ is the critical value of the standard normal distribution corresponding to the desired confidence level $(1 - \alpha) /2$. For a $100q\%$ confidence interval, $\alpha = 1 - q$, so you need to find the z-value for the lower tail probability of $\alpha/2$.

n is the sample size.

```
#1e
## find quantile
a15 <- qnorm((1 + 0.15)/2)
a90 <- qnorm((1 + 0.90)/2)
a95 <- qnorm((1 + 0.95)/2)
## find approximate confidence interval
thetahat - a15*sqrt(thetahat/n) # lower bound
thetahat + a15*sqrt(thetahat/n) # upper bound
thetahat - a90*sqrt(thetahat/n) # lower bound
thetahat + a90*sqrt(thetahat/n) # upper bound
thetahat - a95*sqrt(thetahat/n) # lower bound
thetahat + a95*sqrt(thetahat/n) # upper bound
```

The approximate 95% confidence interval is most similar to the 15% likelihood interval. This is what I would expect, because both intervals capture a specific proportion of the distribution, which in this case is 15%. The equivalent confidence interval for the 95% likelihood interval for lambda is 0.9486, which aligns with the notion that a 95% confidence interval should also have a similar width, capturing roughly 15% of the possible values for the parameter. Therefore, the similarity between the two intervals is consistent with their respective statistical properties and expectations.

```
#1f
### equivalent confidence interval
pchisq(-2 * log(0.15), 1)
```

The interval] 36.41, 37.21 tells us that we can be 95% confident that the true population parameter falls within this range. In other words, there is a high degree of certainty that the true value lies between 36.41 and 37.21 in the context of the study, given the data and methodology employed.

It is possible for lambda = 39, because when θ=39 R(θ)= 1.149788e-24. This implies that there is a little bit support for the parameter value of 39 in the context of the study.

```
> PoisRLF(39, n, thetahat)
[1] 1.149788e-24
```

I analyzed the subject.sex variate for Chicago below.

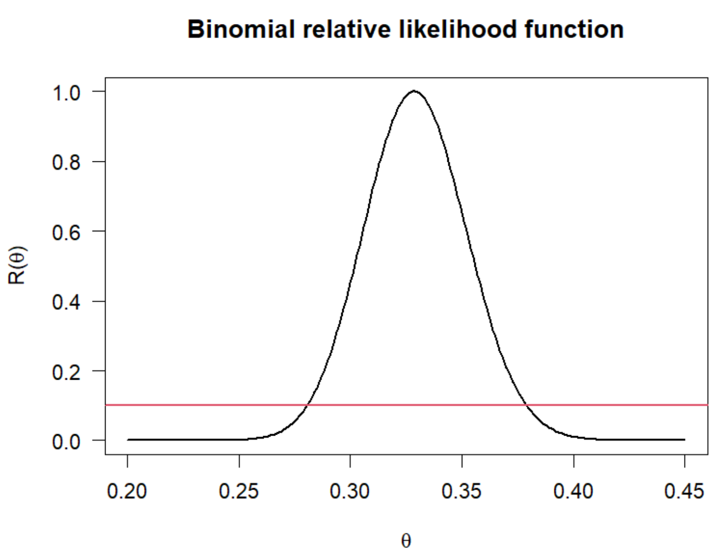
In the sample, for Chicago, the sample size is [417], the number of stops for which subject.sex was `female' is [137], and the maximum likelihood estimate of [θc] is [0.329].

10% likelihood interval for [θc] is [0.281, 0.379], and was calculated by:

```
#2c
BiRLF <- function(theta, n, y, thetahat)
{exp(y*log(theta/thetahat) + (n-y)*log((1-theta)/(1-thetahat)))}
n <- nrow(chicago)
y <- 137
thetahat <- 137/417
theta <- seq(0.2, 0.45, by = 0.001)

plot(theta, BiRLF(theta, n, y, thetahat), main = "Binomial relative likelihood function",
     xlab = expression(theta), ylab = expression(paste("R(", theta, ")")), type = "l",
     lwd = 2, las = 1)
abline(h=0.10, col=2, lwd=2)

uniroot(function(x) BiRLF(x, n, y, thetahat) - 0.1, lower = 0.25, upper = 0.30)$root
uniroot(function(x) BiRLF(x, n, y, thetahat) - 0.1, lower = 0.35, upper = 0.40)$root
```



Based on this interval, I [do not] think [0.513] is a plausible value for [θc].

The [10]% likelihood level is approximately equivalent to a [96.8]% confidence level, because [q <- pchisq(-2 * log(0.1), 1)]. This confidence interval is [0.279, 0.378].

```
#2d
q10 <- pchisq(-2 * log(0.1), 1)
q10
a10 <- qnorm((1 + q)/2)
## find approximate confidence interval
thetahat - a10*sqrt((thetahat*(1-thetahat))/n) # lower bound
thetahat + a10*sqrt((thetahat*(1-thetahat))/n) # upper bound
```

It [is not] possible for [θc] = 0, because [it is impossible that there is no female in chicago, so the true parameter for θc can’t equal 0].

Then I analyzed the lat and lng variates for Chicago.

The sample size is 417. The summary statistics are as follows:

Sample Statistic	lat	lng
Mean	41.850	-87.661
Standard deviation	0.076	0.061
2. 5th percentile	41.700	-87.781
97.5th percentile	41.999	-87.541

A 95% confidence interval for μt is [41.842, 41.857], while for μg is [-87.667, -87.655]. These [are] based on an asymptotic approximation, and were calculated by:

```
#3c
###μt
a_lat <- qnorm((1 + 0.95) / 2)
mean_lat - a_lat * (sd_lat / sqrt(n)) # lower bound
mean_lat + a_lat * (sd_lat / sqrt(n)) # upper bound
###μg
a_lng <- qnorm((1 + 0.95) / 2)
mean_lng - a_lng *(sd_lng / sqrt(n)) # lower bound
mean_lng +a_lng * (sd_lng / sqrt(n)) # upper bound
```

The latitude value given by wiki.openstreetmap.org for my chosen city is [41.8781], while the longitude value is [-87.6298]. Based on my analyses, the latitude value [is not] a plausible value for μt, while the longitude value [is not] a plausible value for μg. These conclusions are based on [41.8781 is not in the 95% confidence interval([41.842, 41.857]) for μt, -87.6298 is not in the 95% confidence interval([-87.667, -87.655]) for μg].

The probability μt lies in the interval is unknown. Since the confidence interval tells us that 95% of intervals constructed from different samples will contain the true population mean latitude. However, it doesn't provide a probability for a specific interval containing the population mean.

A 95% confidence interval for σt is [0.071,0.082], while for σg is [0.057,0.066]. These were calculated by:

```
#3f
###σt
lowert <- qchisq((1 - 0.95) / 2, df = n - 1)
uppert <- qchisq((1 + 0.95) / 2, df = n - 1)
sqrt((n - 1) * sd_lat^2 / uppert) # lower
sqrt((n - 1) * sd_lat^2 / lowert) # upper

###σg
lowerg <- qchisq((1 - 0.95) / 2, df = n - 1)
upperg <- qchisq((1 + 0.95) / 2, df = n - 1)
sqrt((n - 1) * sd_lng^2 / upperg) # lower
sqrt((n - 1) * sd_lng^2 / lowerg) # upper
```