HW5

Will Tirone

```
Warning in rgl.init(initValue, onlyNULL): RGL: unable to open X11 display

Warning: 'rgl.init' failed, running with 'rgl.useNULL = TRUE'.
```

2.3)

a)

handwritten

b)

By the factorization theorem, we have $g(T(x), \alpha, \beta)$ and since $T(x)$ is a sufficient statistic, this contains all the information about the data with respect to α and β. Plotting the log-likelihood below, we can see the maximum likelihood is achieved at $\alpha, \beta = (7.06, 1.0042)$. From some (crude) manual calculations it looks like the range of α, β that is within 15% of the maximum likelihood is $\alpha \in (1.12, 22.3)$ and $\beta \in (0.972, 1.0225)$

```
n=100
alpha_range = linspace(4,10,n)
beta_range = linspace(1,1.01,n)
likelihood = matrix(data=NA, nrow=n, ncol=n)

vals = function(alpha, beta){
  sum_beta = sum(beta^(0:(n-1)))
  calc = (877 * log(alpha)) + ((47370-877) * log(beta)) -alpha*sum_beta
  return(calc)
}
```
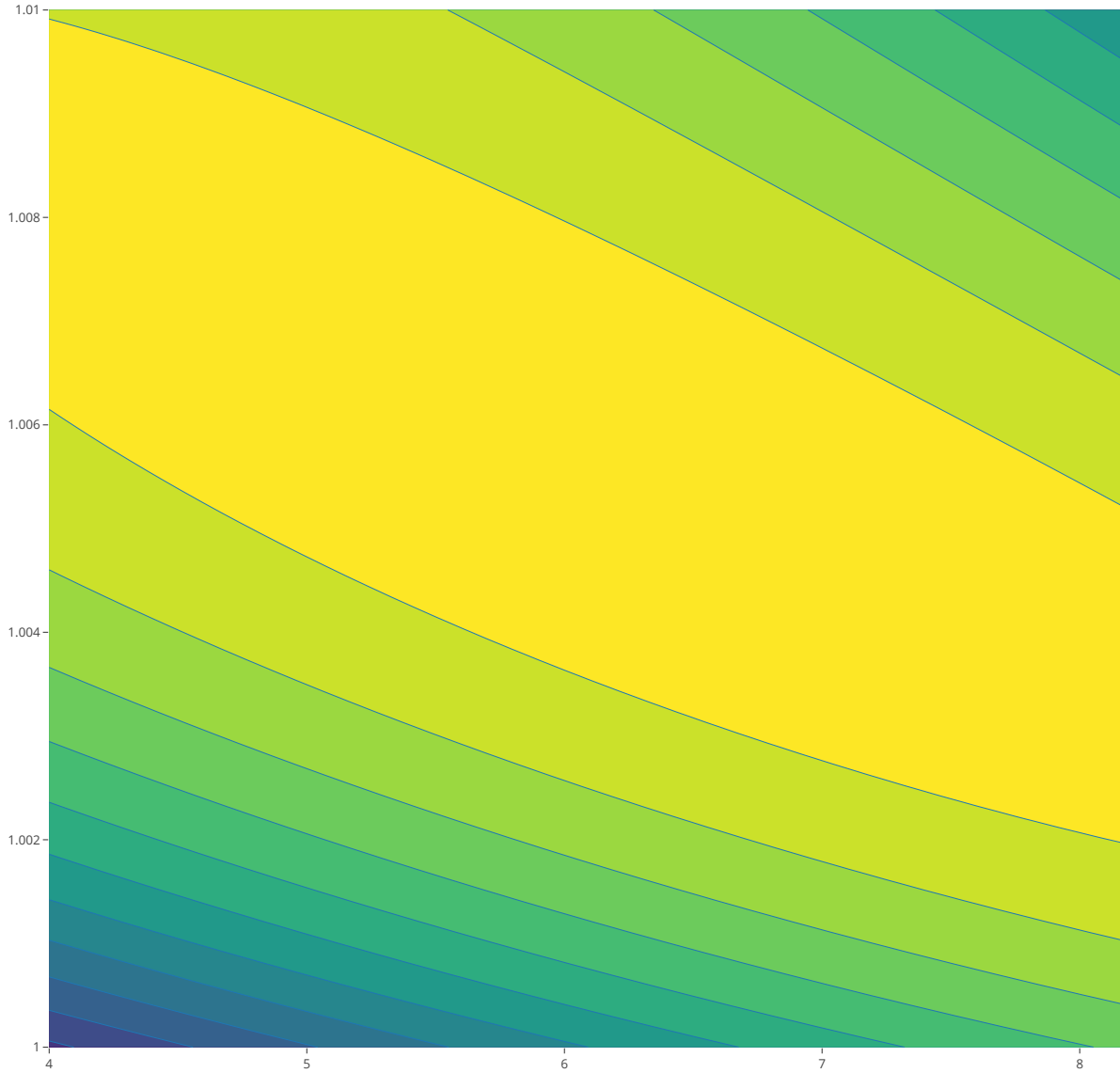
```
for (i in alpha_range){
  for (j in beta_range){
    alpha_index = which(alpha_range == i)
    beta_index = which(beta_range == j)
    likelihood[alpha_index,beta_index] = vals(i,j) # alpha, beta
  }
}

fig = plot_ly(
  x=alpha_range,
  y=beta_range,
  z = likelihood,
  type='contour'
)

fig
```

PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed

c)

From Ch. 2, the parameter $\mu_t = \alpha\beta^{t-1}$ says "the expected annual count grows by a factor of β every year." With a hypothesis test, again from the text, this is structured as as $H_0 : \beta = 1$ (no trend) against $H_A : \beta \neq 1$. I think the inference task here highly depends on who the audience is. I imagine a good statistician could justify a variety of inference tasks, but it depends on the context. For example, if the study was being done to convince climate change skeptics that there is an increase in Tropical Cyclones year after year, a binary conclusion of "Yes" or "No" corresponding to the null and alternative hypotheses might be useful.

However, imagine a scenario where federal funding or budgeting for coastal disaster management organizations was dependent on our modeling. In that case, the government would want to know the severity of the change in β. So, for example, if we constructed a confidence interval of $\beta \in (0.75, 1.25)$ we may be able to suggest that budgeting roughly the same amount year to year would be adequate. However, if we came up with an interval of $\beta \in (10, 12)$ we could very strongly recommend a large increase in budget to handle the increased effects of TC's.

Something not mentioned is the *severity* of the TC's, so I imagine we would want to incorporate that as well. If the count increased but they were extremely weak, we probably wouldn't be very concerned. That could be an alternative (or joint) inferential task we would want to conduct.

d)

The model is probably useful. We could conduct a chi-squared goodness-of-fit test and check for any obvious violations of assumptions. Since it seems like there are no obviously violated assumptions for our data, we can try a goodness-of-fit test. However, knowing the statistic alone can't help us understand the model fit - our statistic really could have come from any distribution and without seeing the actual data we can't assess fit. It just helps us estimate the parameter.

3.2)

a)

handwritten

b)

handwritten

c)

Because each X_i is from a Poisson distribution, $Var(X_i) = E(X_i)$, so large values of T would indicate that the sample variance is larger than the mean, which would let us reject the null.

d)

Large values of T would indicate a smaller $\bar{X}$ relative to the sample variance. If our alternative was that $H_A : X_i \sim Poisson(n\lambda)$, since this would become approximately normal for large n, it seems that we would not want to reject the null for large T because we could have either large or small variance and we would be mis-representing the parameter.

e)

0.05808896 as calculated below.

```
z = (29 * 46.02) / (957 /30)
1 -pchisq(z,29)
```

[1] 0.05808896

3.3)

a)

handwritten

b)

MC Approximation is below.

```
n = 30
R = 957
S_2 = 46.02
T_ = (n-1)*S_2 / (R/n)
data = c(T_ >= rchisq(1000000,29))
1 - mean(data)
```

[1] 0.058072

c)

By iterating over a range of values below, I attempt to show whether or not this is true for all values. Then, by taking the average of the differences, we can see if this is true in the case of every range of value tested. The value output is almost zero, so I believe this will hold generally.

```
delta = c()
for (i in seq(1,1000,10)){
  for (j in seq(300,900,20)){
    for (q in seq(10,70,3)){
    n = i
    R = j
    S_2 = 46
    T_ = (n-1)*S_2 / (R/n)
    data = c(T_ >= rchisq(10000,29))

    stat = 1 - pchisq(T_,29)
    mc_calc = 1 - mean(data)
    delta = c(delta, stat-mc_calc)
    }
  }
}

mean(delta)
```

[1] 1.63276e-06