COLUMBIA UNIVERSITY
Statistical Inference and Modeling
Graded Homework 1: solutions

Prof. Marco Avella
TA: Ian Kinsella
Fall 2019

---

**Exercise 1**

1. The quantile function is simply the inverse CDF

$$F_D(d) = 1 - \exp(-\lambda d) \implies Q_D(p) = F_D^{-1}(p) = \frac{-\log(1-p)}{\lambda}.$$

2. To derive the MLE $\hat{q}_p$ we use the MLE of the unknown parameter $\lambda$

$$\hat{\lambda} = \bar{D}_n^{-1} \implies \hat{q}_p = \frac{-\log(1-p)}{\hat{\lambda}} = -\log(1-p)\bar{D}_n.$$

3. To derive an approximate confidence interval for $Q_D)(p)$ we will want to derive an asymptotic result for $\hat{q}_p$. This can be done simply by noting that

$$\sqrt{n}(\hat{\lambda}^{-1} - \lambda^{-1}) \underset{n\to\infty}{\overset{D}{\to}} \mathcal{N}(0, \lambda^{-2})$$

by CLT. Noting that $\hat{q}_p$ is a linear function of $\hat{\lambda}^{-1}$this enables straightforward application of the delta method for $p \in (0,1)$

$$\sqrt{n}(\hat{q}_p - Q_D(p)) \underset{n\to\infty}{\overset{D}{\to}} \mathcal{N}\left(0, \left(\frac{-\log(1-p)}{\lambda}\right)^2\right).$$

Noting that the standard deviation is our MLE permits convenient standardization and application of the standard normal confidence interval

$$\left[\hat{q}_p - \frac{z_{1-\alpha/2}}{\sqrt{n}}\hat{q}_p,\ \hat{q}_p + \frac{z_{1-\alpha/2}}{\sqrt{n}}\hat{q}_p\right].$$

4. Noting that $\lambda D_i \overset{iid}{\sim} \text{Exp}(1)$ $i = 1, \cdots, n$, we may conclude that $\lambda n \bar{D}_n \sim \text{Gamma}(1, n)$. This makes it clear that $\lambda\bar{D}_n$ is a pivot as the previous expressions are free of the parameter $\lambda$. As such, we may choose to use critical values from the gamma distribution in order to derive an exact confidence interval. So letting $\gamma_{\alpha/2}, \gamma_{(1-\alpha/2)}$ be the relevant critical values, we have

$$\begin{aligned}
\alpha &= P(\gamma_{\alpha/2} \leq \lambda\bar{D}_n \leq \gamma_{(1-\alpha/2)}) \\
&= P(-\log(1-p)\gamma_{\alpha/2} \leq \lambda\hat{q}_p \leq -\log(1-p)\gamma_{(1-\alpha/2)}) \\
&= P\Big(\frac{\gamma_{\alpha/2}}{\hat{q}_p} \leq \frac{\lambda}{-\log(1-p)} \leq \frac{\gamma_{(1-\alpha/2)}}{\hat{q}_p}\Big) \\
&= P\Big(\frac{\hat{q}_p}{\gamma_{(1-\alpha/2)}} \leq Q_D(p) \leq \frac{\hat{q}_p}{\gamma_{\alpha/2}}\Big)
\end{aligned}$$

so we see our interval for the median is $\left[\frac{\hat{q}_{.5}}{\gamma_{(1-\alpha/2)}},\ \frac{\hat{q}_{.5}}{\gamma_{\alpha/2}}\right]$.

**Exercise 2**

1. Note that
$$\mu_2 = \mathbb{E}[x^2] = \lambda^2 + \lambda \iff \lambda^2 - \lambda - \mu_2 = 0.$$
Solving the latter quadratic equation we see that
$$\lambda = \frac{-1 + \sqrt{1 + 4\mu_2}}{2},$$
since $\lambda > 0$ excludes the other solution. Therefore,
$$\hat{\lambda}_M = \frac{-1 + \sqrt{1 + \frac{4}{n}\sum_{i=1}^n X_i^2}}{2}$$

2. Consider the function $g(x) = \frac{-1+\sqrt{1+4x}}{2}$ and note that $\hat{\lambda}_M = g\Big(\frac{1}{n}\sum_{i=1}^n X_i^2\Big)$. Furthermore
$$g'(x) = \frac{1}{\sqrt{1+4x}}\Big|_{x=\lambda^2+\lambda} = \frac{1}{1 + \sqrt{1 + 4(\lambda^2 + \lambda)}}.$$
By CLT,
$$\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^n X_i^2 - (\lambda^2 + \lambda)\Big) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, 4\lambda^3 + 6\lambda^2 + \lambda)$$
By Delta Method,
$$\sqrt{n}(\hat{\lambda}_M - \lambda) \xrightarrow[n\to\infty]{\mathcal{D}} N\Big(0, \frac{1}{(2\lambda+1)^2}(4\lambda^3 + 6\lambda^2 + \lambda)\Big)$$

3. The 1st moment estimator is $\tilde{\lambda} = \overline{X_n}$. By the CLT
$$\sqrt{n}(\overline{X_n} - \lambda) \xrightarrow[n\to\infty]{\mathcal{D}} N(0, \lambda)$$
Since moment estimators are consistent, their asymptotic relative efficiency is approximately
$$\frac{\mathrm{MSE}(\hat{\lambda}_M)}{\mathrm{MSE}(\tilde{\lambda})} \xrightarrow[n\to\infty]{} \frac{1}{(2\lambda+1)^2}(4\lambda^3 + 6\lambda^2 + \lambda) = \frac{\lambda(2\lambda+1)^2 + 2\lambda^2}{(2\lambda+1)^2} > 1$$
This shows that $\lambda_M$ is more asymptotically more efficient.

4. Confidence interval for $\hat{\lambda}_M$ is

$$\left[\hat{\lambda}_M - (2\hat{\lambda}_M + 1)^{-1}(4\lambda^3 + 6\lambda^2 + \lambda)^{1/2}\frac{z_{\alpha/2}}{\sqrt{n}}, \hat{\lambda}_M + (2\hat{\lambda}_M + 1)^{-1}(4\lambda^3 + 6\lambda^2 + \lambda)^{1/2}\frac{z_{\alpha/2}}{\sqrt{n}}\right]$$

And

$$\tilde{\lambda} = \frac{1}{n}\sum_{i=1} nx_i$$

Confidence interval is

$$[\tilde{\lambda} - \sqrt{\tilde{\lambda}}\frac{z_{\alpha/2}}{n}, \tilde{\lambda} + \sqrt{\tilde{\lambda}}\frac{z_{\alpha/2}}{n}]$$

where $z_{\alpha/2} = z$ such that $\mathbb{P}(N(0,1) > z) = \alpha/2$

Since $\tilde{\lambda}$ is more efficient, the confidence intervals computed using this estimator will tend to be shorter. In fact, the variance of $\tilde{\lambda}$ matches the Cramer-Rao lower bound and it is the MVUE.

**Exercise 3**

1. For notational convenience define $\tilde{R}_i := R_i - \mu$, so we have

$$\begin{aligned}\gamma = \mathbb{E}[R_1^3] &= \mathbb{E}[(\tilde{R}_1 + \mu)^3] \\ &= \mathbb{E}[\tilde{R}_1^3] + \mu^3 + 3\mu\mathbb{E}[\tilde{R}_1^2] + 3\mu^2\mathbb{E}[\tilde{R}_1] \\ &= \mu^3 + 3\mu\sigma^2.\end{aligned}$$

2. (a) Following the derivation above

$$\begin{aligned}\mathbb{E}[\hat{\gamma}_n] &= \mathbb{E}[\bar{R}_n^3] \\ &= \mathbb{E}[(\bar{R}_n - \mu)^3] + \mu^3 + 3\mu\mathbb{E}[(\bar{R}_n - \mu)^2] + 3\mu^2\mathbb{E}[(\bar{R}_n - \mu)] \\ &= \mu^3 + 3\mu\mathrm{Var}[\bar{R}_n] = \mu^3 + \frac{3\mu\sigma^2}{n}.\end{aligned}$$

From this result and part 1, it is immediate to see that

$$\mathrm{bias}(\gamma, n) = \mathbb{E}[\hat{\gamma}_n - \gamma] = \frac{3\mu\sigma^2}{n} - 3\mu\sigma^2 = 3\mu\sigma^2(n^{-1} - 1).$$

(b) Taking limits of the bias derived in part (a), we see that this estimator cannot be consistent for $\gamma$ since

$$\lim_{n\to\infty} \mathrm{bias}(\gamma, n) = -3\mu\sigma^2.$$

3. Following the results above, to obtain an unbiased estimator for $\gamma$ we must correct for the bias of $3\mu\sigma^2(n^{-1}-1)$. The bias depends on the unknown parameters $(\mu, \sigma^2)$ so we will use the estimators $\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^n (R_i - \bar{R}_n)^2$ and $\hat{\mu}_n = \bar{R}_n$ to define the estimator $U_n = \hat{\gamma}_n - 3\hat{\mu}_n\hat{\sigma}_n^2(n^{-1}-1)$. Students should use the fact that $\hat{\mu}, \hat{\sigma}^2$ are independent of one another and unbiased to show

$$\begin{aligned}\mathbb{E}\left[U_n - \gamma\right] &= \mathbb{E}\left[\hat{\gamma}_n - 3\hat{\mu}_n\hat{\sigma}_n^2(n^{-1}-1) - \gamma\right] \\ &= \mathbb{E}[\hat{\gamma}_n - \gamma] - 3\mathbb{E}[\hat{\mu}_n]\mathbb{E}[\hat{\sigma}_n^2](n^{-1}-1) \\ &= 3\mu\sigma^2(n^{-1}-1) - 3\mu\sigma^2(n^{-1}-1) = 0\end{aligned}$$

proving that $U_n$ is indeed unbiased.

4. (a) Based on our definition of $\gamma$, it is much easier to derive the expectation of this estimator and see that it is clearly unbiased

$$\mathbb{E}[\tilde{\gamma}_n] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[R_i^3] = \mathbb{E}[R_1^3] = \gamma.$$

   (b) Based on the above, to show consistency all that we now require is to show that the variance of our estimator goes to 0 as $n \to \infty$ which follows from

$$\lim_{n\to\infty} \mathrm{Var}[\tilde{\gamma}_n] = \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}[R_1^3] = 0$$

   since $\mathrm{Var}[R_1^3]$ is a finite constant.

5. We will use the bias-corrected result from part 4 and condition on the statistics $T_1 = \sum_{i=1}^n R_i$ and $T_2 = \sum_{i=1}^n R_i^2$ in order to derive the MVUE. Noting once again that $\hat{\mu} \perp \hat{\sigma^2}$ and that our estimators are functions of the statistics being conditioned on, the students should show that

$$\begin{aligned}\mathbb{E}\left[U_n \middle| T_1, T_2\right] &= \mathbb{E}\left[\hat{\gamma}_n - 3\hat{\mu}_n\hat{\sigma}_n^2(n^{-1}-1) \middle| T_1, T_2\right] \\ &= \mathbb{E}[\hat{\gamma}_n|T_1] - 3\mathbb{E}[\hat{\mu}|T_1]\mathbb{E}[\hat{\sigma}^2|T_1, T_2](n^{-1}-1) \\ &= \hat{\gamma}_n - 3\hat{\mu}\hat{\sigma}^2(n^{-1}-1) = U_n.\end{aligned}$$

This equivalence in conjunction with the Rao-Blackwell Theorem shows that our bias-corrected estimator was in fact the MVUE all along.

**Exercise 4**

1. From Figure there seems to be a negative correlation between the variables age and tot that one can expect to be captured by a straight line. There is probably no other obvious pattern and the variability of the observations seems to be more or less homogeneous across the range of values of age. One can certainly think of including more covariates as predictors for tot including body mass index, sex at birth, etc.

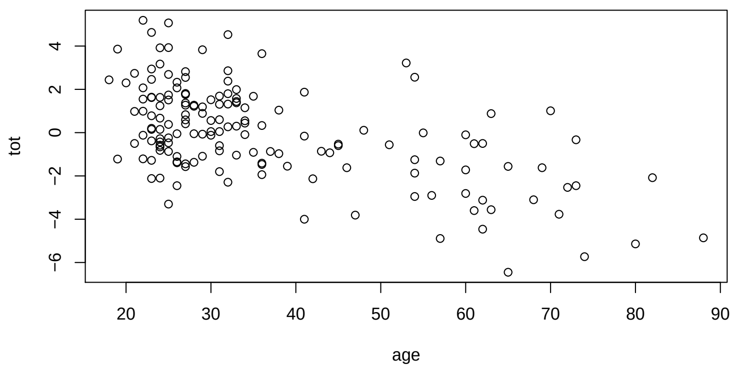


Figure 1: Scatter plot of Kidney data

2. The variable tot is the natural response variable here. One would like to understand how the kidney function varies as function of age which one use as predictor.

3. The sign of the intercept should be positive and the sign of the slope parameter negative. They would capture the intuition that the overall kidney function tot is positive at a young age and decreases as we grow older.

4. For model (a) one is tempted to extrapolate beyond the support of the data at hand (age $\in \{18, 19, \ldots, 88\}$) and interpret $\alpha$ as the tot at birth while $\beta$ would indicate the average yearly decay of tot.
   For model (b) $\alpha$ indicates the avarage level of tot for a patient with age $\bar{X}$ and $\beta$ still captures the decay rate of tot as age increases. This model is somehow awkward it changes as $n$ varies because of $\overline{X_n}$

5. A least squares fit gives the following output

```
Call:
lm(formula = tot ~ age, data = kidney)

Residuals:
    Min      1Q  Median      3Q     Max
-4.2018 -1.3451  0.0765  1.0719  4.5252

Coefficients:
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.860027   0.359565   7.954 3.53e-13 ***
age         -0.078588   0.009056  -8.678 5.18e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.801 on 155 degrees of freedom
Multiple R-squared:  0.327,Adjusted R-squared:  0.3227
F-statistic: 75.31 on 1 and 155 DF,  p-value: 5.182e-15
```

The sign of the estimated parameters match the intuition given in point 3. They are both statistically highly significant using the p-values computed using t-statistics. item

6. The least squares estimator can be seen as the best linear fit to the data at hand without any additional model assumptions. The estimated intercept and slopes simply reflect the observed decay in overall decay in kidney function as we age. However any statement about statistical significance has to rely on some model assumptions that are leveraged for the calculation of p-values.

7. The prediction of tot for a 100 years old person is

$$\hat{y} = \hat{\alpha} + \hat{\beta}100 \approx -4.99.$$

One has to do some extrapolation for this prediction because the oldest person in the data set was 88 and there are also far less older people than young people.

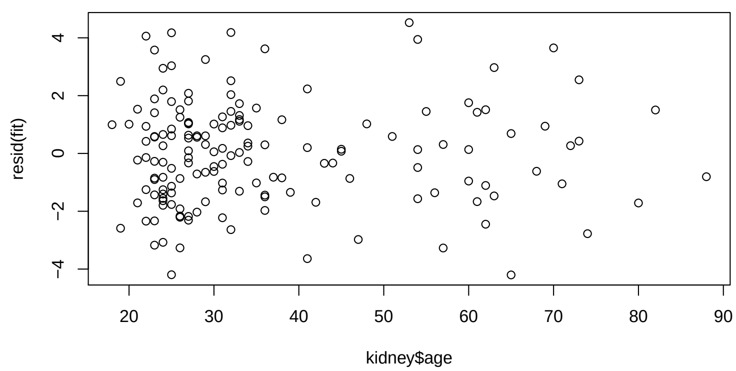


Figure 2: Residuals V. age

8. Figure 8 shows no obvious pattern in the residuals indicating that should not assume i.i.d. errors. As noted earlier, there is a larger number of young people in this study. In particular between 20 and 40.

9. The exact 95% confidence interval for $\beta$ can be obtained using the fact that the standardized least squares estimators of $\beta$ is distributed as a $t_{n-2}$. This leads to $[-0.0964, -0.0606]$.
We can also use the approximation $\sqrt{n}(\hat{\beta} - \beta) \approx N(0, \sigma^2/\sigma_X^2)$, where $\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X_n})^2$ and $\sigma^2 = \text{var}(\varepsilon_i)$. This leads to the confidence interval $[\hat{\beta} - z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}\sigma_X}, \hat{\beta} + z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}\sigma_X}] = [-0.09622; -0.0609]$, where $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2$ is the empirical variance of the residuals.

   Both intervals are very similar since the differences betwenn a $t_n$ and a standard normal disappear for large $n$. The normal confidence interval will always be more optimistic because it uses lighter tails!

10. A nonparametric bootstrap gave me the confidence interval $[-0.965; -0.785]$ using the empirical bootstrap qunatiles (see R code). It is an almost symmetric confidence interval around the least squares estimate $\hat{\beta} = -0.7858$, but is more conservative (wider) than the interval computed with the t-distribution.

11. In Figure 11 we can see that 3 individuals over 70 years old are the most influencial.

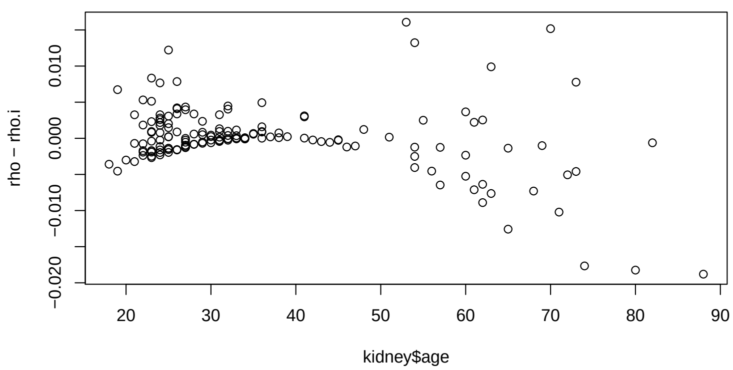


Figure 3: Individual influence on empirical correlation

**Exercise 5**

1. Anything that demonstrates that the students read the paper (or at least the indtroduction is sufficient). Some key things to note are that the author wanted to model the rate at which simulataneous discoverie are made in different fields and he wanted to use a poisson distribution to model this.

One challenge he faced was that his data only only consisted of observations valued $\geq 2$ making straightforward estimation of the mean parameter of the poisson difficult.

2. The key observation is that the way this estimator is defined does not make it ammenable to analysis. Therefore we cannot say much about it's statistical properties.

3. Using this model would allow us to derive estimators in a principled way and verify that they have favorable statistical properties.

4. 

$$
\begin{aligned}
\mathbb{E}[Y] &= \frac{\mu - \mu \exp(-\mu)}{1 - \exp(-\mu) - \mu \exp(-\mu)} \\
\mathbb{E}[Y^2] &= \frac{\mu^2 + \mu - \mu \exp(-\mu)}{1 - \exp(-\mu) - \mu \exp(-\mu)} \\
\mathrm{Var}[Y] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 &= \frac{\mu[(1 - \exp(-\mu))^2 - \mu^2 \exp(-\mu)]}{(1 - \exp(-\mu) - \mu \exp(-\mu))^2}
\end{aligned}
$$

5. Taking the log of the Likelihood function yields

$$
\begin{aligned}
l(\mu | Y_1, \cdots, Y_n) = -n\mu - n \log(1 - \exp(-\mu) - \mu \exp(-\mu)) \\
+ \log(\mu) \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \log(Y_i!).
\end{aligned}
$$

We can compute the MLE numerically by maximizing this function.

6. The MLE should be $\approx 1.398$, any function optimizer can be used. Alternatively, it is also valid if they interacted with the plot to find the maximizer in this simple case.

7. First, we derive the Fischer Information for the parameter $\mu$

$$
\begin{aligned}
\frac{dl(\mu|...)}{d\mu} &= \mu^{-1} \sum_{i=1}^{n} Y_i - \frac{n(1 - \exp(-\mu))}{1 - \exp(-\mu) - \mu \exp(-\mu)} \\
\implies \mathbb{E}\Big[\Big(\frac{dl(\mu|...)}{d\mu}\Big)^2\Big] &= \mathrm{Var}\Big(\frac{dl(\mu|...)}{d\mu}\Big) = \frac{n}{\mu^2}\mathrm{Var}\Big(Y\Big). \\
\implies I(\mu) &= \frac{(1 - \exp(-\mu))^2 - \mu^2 \exp(-\mu)}{\mu(1 - \exp(-\mu) - \mu \exp(-\mu))^2}.
\end{aligned}
$$

Next will use our results for MLEs to obtain the variance of the asymptotic distribution as

$$
\sqrt{n}(\hat{\mu} - \mu) \underset{n \to \infty}{\overset{D}{\to}} \mathcal{N}(0, I(\mu)^{-1}).
$$

8. Using the Fischer information derived above (and hence variance of the asymptotic distribution), we obtain the confidence interval

$$\left[\hat{\mu} - \frac{z_{1-\alpha/2}}{\sqrt{nI(\hat{\mu})}},\ \hat{\mu} + \frac{z_{1-\alpha/2}}{\sqrt{nI(\hat{\mu})}}\right].$$

9. The estimates are surprisingly similar! In the paper his method yielded an estimate of $\mu = 1.4$ whereas our MLE estimate is given in problem 6 above.