STA500 Introduction to Probability and Statistics 2, autumn 2018.

# Solution exercise set 5

### Exercise 1:

**a)**

$$
\begin{aligned}
L(\lambda) &= \prod_{i=1}^{n}(\lambda e^{-\lambda t_i}) = \lambda^n e^{-\lambda \sum_{i=1}^{n} t_i} \\
l(\lambda) &= \ln L(\lambda; t_1, ..t_n) = \ln(\lambda)^n + \ln(e^{-\lambda \sum_{i=1}^{n} t_i}) = n\ln(\lambda) - \lambda \sum_{i=1}^{n} t_i \\
\frac{dl(\lambda)}{d\lambda} &= \frac{n}{\lambda} - \sum_{i=1}^{n} t_i = 0 \qquad \Rightarrow \lambda = \frac{n}{\sum_{i=1}^{n} t_i}
\end{aligned}
$$

I.e. the MLE is given as: $\hat{\lambda} = \underline{\underline{\frac{n}{\sum_{i=1}^{n} T_i}}}$. Estimate: $\hat{\lambda} = \frac{12}{108.5} = \underline{\underline{0.11}}$.

**b)** We start by considering the distribution of $Y = 2\lambda T_i$. From the result on transformations (see collection of formulas) we have for the exponential distribution that $2X/\beta$ has a $\chi^2_2$-distribution if $X$ is exponentially distributed with parameter $\beta$. In the current setting we have that $T_i$ is exponentially distributed with expectation $1/\lambda$, i.e. $\beta = 1/\lambda$, and it thus follow that $Y = 2T_i/\beta = 2\lambda T_i$ is having a $\chi^2_2$-distribution.

Further we have from results of linear combinations (collection of formulas) that since a sum of independent $\chi^2$-distributed variables is $\chi^2$-distributed with parameter ("degrees of freedom") equal to the sum of the parameters in the distribution of each variable, we will have that $Z = 2\lambda \sum_{i=1}^{n} T_i = \sum_{i=1}^{n} 2\lambda T_i$ is having a $\underline{\underline{\chi^2_{2n}\text{-distribution}}}$ (by being a sum of $n$ indep. $\chi^2_2$-distributed variables).

**c)**

$$
\mathrm{E}(\hat{\lambda}) = n\mathrm{E}(\frac{1}{\sum_{i=1}^{n} T_i}) = n\mathrm{E}(\frac{2\lambda}{2\lambda\sum_{i=1}^{n} T_i}) = 2n\lambda\mathrm{E}(\frac{1}{Z}) = 2n\lambda \cdot (\frac{1}{2(n-1)}) = \lambda\frac{n}{n-1}
$$

I.e. $\hat{\lambda}$ is $\underline{\underline{\text{biased}}}$.

We try the slightly modified estimator $\lambda^* = \frac{n-1}{n}\hat{\lambda} = \frac{n-1}{\sum_{i=1}^{n} T_i}$.

$$
\mathrm{E}(\lambda^*) = \frac{n-1}{n}\mathrm{E}(\hat{\lambda}) = \frac{n-1}{n}\frac{n}{n-1}\lambda = \lambda
$$

I.e. $\lambda^* = \frac{n-1}{\sum_{i=1}^{n} T_i}$ is an unbiased estimator for $\lambda$.

Using this estimator gives the estimate: $\lambda^* = \frac{11}{108.5} = 0.10$.

Generally we know that maximum likelihood estimators can be biased, but asymptotically (as $n \to \infty$) they are unbiased. We have found the MLE $\hat{\lambda}$ to be biased, but we see that

$$\mathrm{E}(\hat{\lambda}) = \lambda \frac{n}{n-1} \to \lambda \text{ when } n \to \infty \quad (\text{since } \frac{n}{n-1} \to 1)$$

i.e. asymptotically the estimator is unbiased as it should be.

**d)**

Since

$$Z = 2\lambda \sum_{i=1}^{n} T_i = 2\lambda n \frac{\sum_{i=1}^{n} T_i}{n} = 2n\frac{\lambda}{\hat{\lambda}} \sim \chi^2_{2n}$$

we get

$$\begin{aligned} P(\chi^2_{1-\frac{\alpha}{2},2n} < Z < \chi^2_{\frac{\alpha}{2},2n}) &= 1-\alpha \\ P(\chi^2_{1-\frac{\alpha}{2},2n} < 2n\frac{\lambda}{\hat{\lambda}} < \chi^2_{\frac{\alpha}{2},2n}) &= 1-\alpha \\ P(\frac{\hat{\lambda}}{2n}\chi^2_{1-\frac{\alpha}{2},2n} < \lambda < \frac{\hat{\lambda}}{2n}\chi^2_{\frac{\alpha}{2},2n}) &= 1-\alpha \end{aligned}$$

I.e. a $(1-\alpha)100\%$ confidence interval for $\lambda$ becomes:

$$\left[\frac{\hat{\lambda}}{2n}\chi^2_{1-\frac{\alpha}{2},2n},\ \frac{\hat{\lambda}}{2n}\chi^2_{\frac{\alpha}{2},2n}\right]$$

By inserting $\alpha = 0.05$ which gives $\chi^2_{1-\frac{\alpha}{2},2n} = \chi^2_{0.975,24} = 12.401$, $\chi^2_{\frac{\alpha}{2},2n} = \chi^2_{0.025,24} = 39.364$ and $\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} T_i} = \frac{12}{108.5} = 0.111$ we get the 95% confidence interval:

$$[\frac{0.111}{2\cdot 12}\cdot 12.401,\ \frac{0.111}{2\cdot 12}\cdot 39.364] = [0.06,\ 0.18]$$

## Exercise 2:

**a)** If we first look at $Y_i = 2X_i/\beta$ we have from the overview of results on transformations in the collection of formulas that $Y$ has a $\chi^2_2$-distribution.

Further, among the result on linear combinations we have one result which says that a sum of independent $\chi^2$-distributed variables is $\chi^2$-distributed with parameter ("degrees of freedom") equal to the sum of the parameters in the distribution of each single variable. I.e. in our case will $Z = \sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} 2X_i/\beta$ be $\chi^2$-distributed with parameter $\sum_{i=1}^{n} 2 = 2n$, i.e. $Z$ is $\chi^2_{2n}$-distributed.

**b)** We have from the results on transformations that if $X$ is gamma$(\alpha, \beta)$-distributed then $Y = 2X/\beta$ is $\chi^2_{2\alpha}$-distributed. Turned the other way around this means that if $Y$ has a $\chi^2_{2\alpha}$-distribution then $X = Y\beta/2$ has a gamma$(\alpha, \beta)$-distribution.

Now, in our specific case we showed in a) that $Z = \sum_{i=1}^n 2X_i/\beta = 2(\sum_{i=1}^n X_i)/\beta = 2V/\beta$ has a $\chi^2_{2n}$-distribution. It then follows that $V = Z\beta/2$ has a gamma-distribution with $\alpha = n$ and $\beta = \beta$.

We have here shown the result that a sum of independent identically exponentially distributed variables is gamma distributed with $\alpha = n$ and $\beta = \beta$.

### Exercise 3:

**a)** The exact distribution of $X$ will be a hypergeometric distributions with parameters $N$, $k$ and $n$ where $N$ is the number of voters, $k$ is the number of voters voting on Ap and $n = 1000$ is the number of voters which is asked in the opinion poll. Since $n << N$ we have that $X$ will be approximately binomially distributed with parameters $n = 1000$ and $p = k/N$.

**b)** Since $np(1-p) > 5$ we can use the approximation to the normal distribution:

$$\begin{aligned} P(X > 300) &= 1 - P(X \leq 300) = 1 - P(Z \leq \frac{300 + 0.5 - 1000 \cdot 0.275}{\sqrt{1000 \cdot 0.275 \cdot (1 - 0.275)}}) \\ &= 1 - P(Z \leq 1.81) = 1 - 0.9649 = \underline{\underline{0.0351}} \end{aligned}$$

I.e. there is approximately a probability of 3.5% that the opinon poll will indicate that more than 30% would vote for Ap when the reality is that only 27.5% would vote for Ap.

**c)** In the binomial distribution we have $f(x;p) = \binom{n}{x} p^x (1-p)^{n-x}$, and since we have only made on binomial experiment (only one $X$, one poll) we get $L(p;x) = f(x;p)$, i.e.

$$\begin{aligned} L(p;x) &= \binom{n}{x} p^x (1-p)^{n-x} \\ l(p;x) &= \ln L(p;x) = \ln\binom{n}{x} + x \ln p + (n-x)\ln(1-p) \\ \frac{\partial l(p;x)}{\partial p} &= \frac{x}{p} - \frac{n-x}{1-p} = 0 \\ &\Rightarrow \quad (1-p)x \; - \; p(n-x) = 0 \\ & \quad\; x - px - pn + px = 0 \quad \Rightarrow \quad \underline{\underline{\hat{p} = \frac{X}{n}}} \end{aligned}$$

Check that we have found a maximum:

$$\frac{\partial^2 l(p; x_1, \ldots, x_n)}{\partial p^2} = -\frac{1}{p^2} x - \frac{1}{(1-p)^2}(n-x) < 0 \qquad \text{i.e. maximum!}$$

$$\begin{aligned} \mathrm{E}(\hat{p}) &= \mathrm{E}(\frac{X}{n}) = \frac{1}{n}\mathrm{E}(X) = \frac{1}{n} np = \underline{\underline{p}} \\ \mathrm{Var}(\hat{p}) &= \frac{1}{n^2}\mathrm{Var}(X) = \frac{1}{n^2} np(1-p) = \underline{\underline{\frac{p(1-p)}{n}}} \end{aligned}$$

**d)** $X \sim \text{bin}(n,p)$. Confidence interval for $p$:

$$\hat{p} = \frac{X}{n}$$

$$\frac{\hat{p} - \mathrm{E}(\hat{p})}{\sqrt{\mathrm{Var}(\hat{p})}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \overset{CLT}{\approx} \mathrm{N}(0,1)$$

To simplify the calculations of the confidence interval we further use the approximation that when $n$ is large $p(1-p) \approx \hat{p}(1-\hat{p})$, i.e. we use as starting point:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \approx \mathrm{N}(0,1)$$

$$\begin{aligned}
P(-z_{\alpha/2} \leq Z \leq z_{\alpha/2}) &\approx 1-\alpha \\
P(-z_{\alpha/2} \leq \frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq z_{\alpha/2}) &\approx 1-\alpha \\
&\vdots \\
P(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}) &\approx 1-\alpha
\end{aligned}$$

Observed: $n = 1000$ and $x = 297$ gives $\hat{p} = \frac{297}{1000} = 0.297$.

$\alpha = 0.05 \quad \Rightarrow \quad z_{\alpha/2} = z_{0.025} = 1.96$.

Inserted this give the approximate 95% conf. int. for $p$:

$$[0.297 - 1.96\sqrt{\frac{0.297 \cdot 0.703}{1000}}, 0.297 + 1.96\sqrt{\frac{0.297 \cdot 0.703}{1000}}] = \underline{\underline{[0.269, 0.325]}}$$