ANSWER SHEET 7

Assignment 1. (a) We have $\overline{X} \sim N(\mu, 1/n)$.
(b) Under H_0, $\sqrt{n}\overline{X} \sim N(0,1)$. Letting Φ denote the Gaussian distribution function, we obtain the equation

$$1-\alpha = \mathbb{P}(-v_\alpha \leq \overline{X} \leq v_\alpha) = \mathbb{P}(-\sqrt{n}v_\alpha \leq \sqrt{n}\overline{X} \leq \sqrt{n}v_\alpha) = \Phi(\sqrt{n}v_\alpha) - \Phi(-\sqrt{n}v_\alpha).$$

By symmetry the right hand-side equals $2\Phi(\sqrt{n}v_\alpha)-1$. Thus $1-\alpha/2 = \Phi(\sqrt{n}v_\alpha)$ and $v_\alpha = n^{-1/2}\Phi^{-1}(1-\alpha/2) = n^{-1/2}z_{1-\alpha/2}$.
(c) The p-value is the infimum of the α's for which we reject,

$$p = p(X_1,\ldots,X_n) = \inf\{\alpha : |\overline{X}| > v_\alpha\} = \inf\{\alpha : |\overline{X}| > n^{-1/2}z_{1-\alpha/2}\}.$$

Since $z_{1-\alpha/2}$ is continuous and decreasing in α, the infimum is attained when $|\overline{X}|$ equals the threshold

$$|\overline{X}| = n^{-1/2}z_{1-p/2} = n^{-1/2}\Phi^{-1}(1-p/2) \quad\Longrightarrow\quad p(X_1,\ldots,X_n) = 2(1-\Phi(\sqrt{n}|\overline{X}|)).$$

(d) The code below carries out the simulation :

```
set.seed(25102017)
mu <- 0
n <- 11
REP <- 1000
p <- numeric(REP)
for(i in 1:REP)
{
  X <- rnorm(n, mean = mu, sd = 1)
  p[i] <- 2 - 2 * pnorm(sqrt(n) * abs(mean(X)))
}
hist(p)
```

The resulting histrogram suggests that p-value is uniformly distributed when H_0 holds ; in fact this can be shown to hold true in the continuous case by means of the probability transform. If μ is different than zero, than the histogram is concentrated around zero ; the concentration increases with n and with $|\mu|$.

Assignment 2. Observe that

$$\begin{aligned}
\Lambda(\theta_0) &= \frac{\sup_\theta \Pi_{j=1}^n \exp[-\frac{1}{2}(\theta - X_j)^2]}{\sup_{\theta=\theta_0} \Pi_{j=1}^n \exp[-\frac{1}{2}(\theta - X_j)^2]} = \frac{\Pi_{j=1}^n \exp[-\frac{1}{2}(\overline{X} - X_j)^2]}{\Pi_{j=1}^n \exp[-\frac{1}{2}(\theta_0 - X_j)^2]} \\
&\overset{a}{=} \exp\{-\frac{1}{2}\sum_{j=1}^n[(X_j^2 - \overline{X}^2) - (X_j^2 - \theta_0 X_j + \theta_0^2)]\} = \exp\{-\frac{1}{2}\sum_{j=1}^n[-\overline{X}^2 + 2\theta_0 X_j - \theta_0^2)]\} \\
&= \exp\{\frac{n}{2}(\overline{X} - \theta_0)^2\}
\end{aligned}$$

Where a is because $\sum_{j=1}^n(\overline{X} - X_j)^2 = \sum_{j=1}^n \overline{X}^2 - 2\overline{X}X_j + X_j^2 = n\overline{X}^2 - 2\overline{X}\sum_{j=1}^n X_j + \sum_{j=1}^n X_j^2 = \sum_{j=1}^n[X_j^2 - \overline{X}^2]$. Now,

$$\mathbb{P}_0\{\Lambda(\theta_0) > c_\alpha\} = \mathbb{P}_0\{\sqrt{n}|\overline{X} - \theta_0| > q_{1-\alpha/2}\} = \mathbb{P}_0\{n(\overline{X} - \theta_0)^2 > \chi^2_{1,1-\alpha}\} = \alpha$$

(a) (i) In this case $\mathcal{I}_n(\theta) = n$, which is free of the parameter θ. Hence, the Wald test is given by $\mathbf{1}(n(\overline{X} - \theta_0)^2 > \chi^2_{1,1-\alpha})$. So, the associated $100(1-\alpha)\%$ confidence interval for θ is given by

$$\{\theta : n(\overline{X} - \theta)^2 \leq \chi^2_{1,1-\alpha}\} = \left[\overline{X} - \sqrt{\chi^2_{1,1-\alpha}/n},\ \overline{X} + \sqrt{\chi^2_{1,1-\alpha}/n}\right].$$

(ii) The likelihood ratio test for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is given by $\mathbf{1}(\sqrt{n}|\overline{X} - \theta_0| > q_{1-\alpha/2})$. So, the associated $100(1-\alpha)\%$ confidence interval for θ is given by

$$\{\theta : \sqrt{n}|\overline{X} - \theta| \leq q_{1-\alpha/2}\} = \left[\overline{X} - q_{1-\alpha/2}/\sqrt{n},\ \overline{X} + q_{1-\alpha/2}/\sqrt{n}\right].$$

(iii) However, observe that $\sqrt{\chi^2_{1,1-\alpha}} = q_{1-\alpha/2}$ because of the following reason. Let $Z \sim N(0,1)$ and c be such that $P(|Z| > c) = \alpha \Leftrightarrow c = q_{1-\alpha/2} > 0$ since $\alpha < 1$. However, this is the same as saying $P(Z^2 > c^2) = \alpha \Leftrightarrow c^2 = \chi^2_{1,1-\alpha}$. So, $\sqrt{\chi^2_{1,1-\alpha}} = q_{1-\alpha/2}$. Thus, the two confidence intervals are the same.

(b) (i) For testing $H_0 : p = p_0$ versus $H_1 : p \neq p_0$, Wilks' theorem applied to the likelihood ratio statistic yields the test $\mathbf{1}(2n\overline{X}\log(\overline{X}/p_0) + 2(n - n\overline{X})\log\{(1-\overline{X})/(1-p_0)\} > \chi^2_{1,1-\alpha})$. So, the associated $100(1-\alpha)\%$ confidence interval for p is given by

$$\{p : 2n\overline{X}\log(\overline{X}/p) + 2(n - n\overline{X})\log\{(1-\overline{X})/(1-p)\} \leq \chi^2_{1,1-\alpha}\}.$$

Observe that the above inequation does not in general yield a closed form expression of the confidence interval.

(ii)The likelihood ratio test for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ is given by $\mathbf{1}(\sqrt{n}|\overline{X} - \theta_0| > q_{1-\alpha/2})$. So, the associated $100(1-\alpha)\%$ confidence interval for θ is given by

$$\{\theta : \sqrt{n}|\overline{X} - \theta| \leq q_{1-\alpha/2}\} = \left[\overline{X} - q_{1-\alpha/2}/\sqrt{n},\ \overline{X} + q_{1-\alpha/2}/\sqrt{n}\right].$$

(ii) In this case $\mathcal{I}_n(p) = n/(p(1-p))$. Hence, the Wald test is given by $\mathbf{1}(n(\overline{X} - p_0)^2/(\overline{X}(1-\overline{X})) > \chi^2_{1,1-\alpha})$. So, the associated $100(1-\alpha)\%$ confidence interval for p is given by

$$\{p : n(\overline{X} - p)^2/(\overline{X}(1-\overline{X})) \leq \chi^2_{1,1-\alpha}\}.$$

The above inequation can be solved explicitly (being a quadratic) to obtain a confidence interval for p.

(iii) The asymptotic test using the convergence in distribution as given is $\mathbf{1}(\sqrt{n}|\overline{X} - p_0| > q_{1-\alpha/2}\sqrt{p_0(1-p_0)})$. So, the associated $100(1-\alpha)\%$ confidence interval for p is given by

$$\{p : \sqrt{n}|\overline{X} - p| \leq q_{1-\alpha/2}\sqrt{p(1-p)}\} = \{p : n(\overline{X} - p)^2 \leq q^2_{1-\alpha/2}p(1-p)\}.$$

Once again, the above inequation can be solved explicitly (being a quadratic) to obtain a confidence interval for p.

(iv) No, these confidence intervals are not the same. Note that obtaining an explicit formula for the intervals is easy in the case of (ii), not so easy in the case of (iii) (this confidence interval is sometimes referred to as Wilson's confidence interval), and very hard in the case of (i).

Assignment 3. (a) Note that

$$\begin{aligned}\int_{-\infty}^{\infty} \text{Hist}_{X_1,X_2,\ldots,X_n}(x)dx &= \sum_{j\in\mathbb{Z}} \int_{I_j} \text{Hist}_{X_1,X_2,\ldots,X_n}(x)dx \\ &= \sum_{j\in\mathbb{Z}} \int_{I_j} \frac{1}{nh} \sum_{i=1}^{n} \mathbf{1}(X_i \in I_j)dx \\ &= \sum_{j\in\mathbb{Z}} \frac{1}{nh} \sum_{i=1}^{n} \mathbf{1}(X_i \in I_j) \int_{I_j} dx \\ &= \sum_{j\in\mathbb{Z}} \frac{1}{nh} \sum_{i=1}^{n} \mathbf{1}(X_i \in I_j) \times h \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j\in\mathbb{Z}} \mathbf{1}(X_i \in I_j) = 1\end{aligned}$$

since $\sum_{j\in\mathbb{Z}} \mathbf{1}(X_i \in I_j) = 1$ for each i as $\{I_j\}$ is a partition of $\mathbb{R}$.
(b) Fix any $x \in \mathbb{R}$. Then there exists a unique j such that $x \in I_j$. Then, $nh\text{Hist}_{X_1,X_2,\ldots,X_n}(x) = \sum_{i=1}^{n} \mathbf{1}(X_i \in I_j) \sim Bin(n, p_j)$, where $p_j = \mathbb{P}[X_1 \in I_j]$.
So, $\mathbb{E}[nh\text{Hist}_{X_1,X_2,\ldots,X_n}(x)] = np_j$ and $\text{Var}(nh\text{Hist}_{X_1,X_2,\ldots,X_n}(x)) = np_j(1-p_j)$.
(c) Note that

$$\mathbb{E}[\text{Hist}_{X_1,X_2,\ldots,X_n}(x)] = \frac{p_j}{h} = h^{-1}\mathbb{P}[X_1 \in I_j] = \frac{1}{h}\int_{I_j} f(y)dy \to f(x)$$

as $h \to 0$ by the continuity of f.
(d) Now,

$$\begin{aligned}\mathbb{E}\{[\text{Hist}_{X_1,X_2,\ldots,X_n}(x) - f(x)]^2\} &= (nh)^{-2}\,\mathbb{E}\{[nh\text{Hist}_{X_1,X_2,\ldots,X_n}(x) - nhf(x)]^2\} \\ &= (nh)^{-2}\left\{\text{Var}(nh\text{Hist}_{X_1,X_2,\ldots,X_n}(x)) + [np_j - nhf(x)]^2\right\} \\ &= (nh)^{-2}\left\{np_j(1-p_j) + (nh)^2[(p_j/h) - f(x)]^2\right\} \\ &= (nh)^{-1}(p_j/h)(1-p_j) + [(p_j/h) - f(x)]^2.\end{aligned}$$

(e) We have seen in part (c) that if $h \to 0$ then $p_j/h \to f(x)$. So, $p_j \to 0$ as $h \to \infty$. Thus, if $h \to 0$ and $nh \to \infty$, it follows from the above expression that $\mathbb{E}\{[\text{Hist}_{X_1,X_2,\ldots,X_n}(x) - f(x)]^2\} \to 0$.
(f) The limit $h \to 0$ implies that the we need to choose smaller and smaller values of the bin-width for the mean squared error to converge to zero.
The limit $nh \to \infty$ implies that the bin-width should not converge to zero arbitrarily fast – its rate of decay should not be slower than n^{-1}. Note that $\mathbb{E}[\sum_{i=1}^{n} \mathbf{1}(X_i \in I_j)] = np_j \approx nhf(x)$ for small enough h. So, the previous condition will also guarantee that even if we take a very small bin-width h for I_j, the average/expected number of observations in I_j grows to infinity (provided $f(x) > 0 \Leftrightarrow x$ is in the support of f), i.e., we still have enough sample points in that bin to be able to accurately estimate $f(x)$.
(g) By Chebyshev's inequality, for any $\epsilon > 0$, we have

$$\mathbb{P}[|\text{Hist}_{X_1,X_2,\ldots,X_n}(x) - f(x)| > \epsilon] \leq \frac{\mathbb{E}\{[\text{Hist}_{X_1,X_2,\ldots,X_n}(x) - f(x)]^2\}}{\epsilon^2} \to 0$$

as $h \to 0$ and $nh \to \infty$. Thus, under these conditions, it follows that $\text{Hist}_{X_1,X_2,\ldots,X_n}(x)$ is consistent for $f(x)$.

Assignment 4. (a) The $100(1-\alpha)\%$ confidence intervals for μ and σ^2 are

$$R_{1,\alpha}(\mathbf{X}) = \left[\overline{X} - \frac{q_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}, \overline{X} + \frac{q_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}}\right] \quad \text{and} \quad R_{2,\alpha}(\mathbf{X}) = \left[\frac{(n-1)S^2}{\chi^2_{(n-1),(1-\frac{\alpha}{2})}}, \frac{(n-1)S^2}{\chi^2_{(n-1),\frac{\alpha}{2}}}\right],$$

where q_γ is the γ quantile of the $N(0,1)$ distribution.
(b) No. This is because the independence of $\overline{X}$ and S^2 implies that $\mathbb{P}[R_{1,\alpha}(\mathbf{X}) \ni \mu, R_{2,\alpha}(\mathbf{X}) \ni \sigma^2] = \mathbb{P}[R_{1,\alpha}(\mathbf{X}) \ni \mu]\mathbb{P}[R_{2,\alpha}(\mathbf{X}) \ni \sigma^2] = (1-\alpha)^2 < (1-\alpha)$. The last inequality follows from the fact that $1-\alpha \in (0,1)$.
Using the Bonferroni method, we have

$$\mathbb{P}[R_{1,\beta}(\mathbf{X}) \ni \mu, R_{2,\beta}(\mathbf{X}) \ni \sigma^2] \geq \mathbb{P}[R_{1,\beta}(\mathbf{X}) \ni \mu] + \mathbb{P}[R_{2,\beta}(\mathbf{X}) \ni \sigma^2] - 1 = (2-2\beta) - 1.$$

Thus, we need β to satisfy $1-2\beta = 1-\alpha \Leftrightarrow \beta = \alpha/2$. So, the Bonferroni corrected $100(1-\alpha)\%$ confidence region for $(\mu,\sigma^2)^\top$ is $R_{1,\alpha/2}(\mathbf{X}) \times R_{2,\alpha/2}(\mathbf{X})$.
(c) Note that $\mathbb{P}[R_{1,\beta}(\mathbf{X}) \ni \mu, R_{2,\beta}(\mathbf{X}) \ni \sigma^2] = \mathbb{P}[R_{1,\beta}(\mathbf{X}) \ni \mu]\mathbb{P}[R_{2,\beta}(\mathbf{X}) \ni \sigma^2] = (1-\beta)^2$. So, we need $(1-\beta)^2 = 1-\alpha \Leftrightarrow \beta = 1-\sqrt{1-\alpha}$. Thus, a $100(1-\alpha)\%$ confidence region for $(\mu,\sigma^2)^\top$ is $R_{1,(1-\sqrt{1-\alpha})}(\mathbf{X}) \times R_{2,(1-\sqrt{1-\alpha})}(\mathbf{X})$.
(d) The confidence region in part (c) is preferable since it is exact, i.e., the coverage probability is equal to $(1-\alpha)$ for all values of n. Further, the Bonferroni corrected confidence interval is conservative.
(e) The likelihood ratio test statistic for testing $H_0 : \mu = \mu_0, \sigma^2 = \sigma_0^2$ vs $H_1 : \mu \neq \mu_0, \sigma^2 \neq \sigma_0^2$ is given by

$$l_n = \left(\frac{\widehat{\sigma}^2}{\sigma_0^2}\right)^{n/2} \exp\left[\frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^n (X_i - \overline{X})^2 - \frac{1}{2\sigma_0^2}\sum_{i=1}^n (X_i - \mu_0)^2\right],$$

where $\widehat{\sigma}^2 = (n-1)S^2/n$.
(f) Wilks' theorem says that under the null hypothesis, $-2\log l_n$ converges in distribution to the χ^2_2 distribution as $n \to \infty$. Now,

$$-2\log l_n = n\{\log\sigma_0^2 - \log(n-1) + \log n - \log S^2\} - n + \frac{(n-1)S^2}{\sigma_0^2} + \frac{n(\overline{X}-\mu_0)^2}{\sigma_0^2},$$

Thus, a $100(1-\alpha)\%$ confidence region for $(\mu,\sigma^2)^\top$ is given by

$$\left\{(\mu,\sigma^2) : n\{\log\sigma^2 - \log(n-1) + \log n - \log S^2\} - \frac{n}{2} + \frac{(n-1)S^2}{\sigma^2} + \frac{n(\overline{X}-\mu)^2}{\sigma^2} \leq \chi^2_{2,1-\alpha}\right\}.$$

(g) Using the continuous mapping theorem, it follows that the asymptotic distribution of U_n is χ^2_2 as $n \to \infty$.
(h) A $100(1-\alpha)\%$ confidence region for $(\mu,\sigma^2)^\top$ is given by

$$\left\{(\mu,\sigma^2) : \frac{n(\overline{X}-\mu)^2}{\sigma^2} + \frac{n(S^2-\sigma^2)^2}{2\sigma^4} \leq \chi^2_{2,1-\alpha}\right\}.$$

(i) Since S^2 converges in probability to σ^2 as $n \to \infty$, it follows from Slutsky's theorem and part (g) that V_n converges in distribution to the χ^2_2 distribution as $n \to \infty$.
(j) A $100(1-\alpha)\%$ confidence region for $(\mu, \sigma^2)^\top$ is given by

$$\left\{(\mu,\sigma^2) : \frac{n(\overline{X}-\mu)^2}{S^2} + \frac{n(S^2-\sigma^2)^2}{2S^4} \leq \chi^2_{2,1-\alpha}\right\}.$$

(k) It is easy to see that each of $R_B(\mathbf{X}), R_C(\mathbf{X})$ and $R_D(\mathbf{X})$ can be written in the form $\{(\mu,\sigma^2) : H(\mu,\sigma^2) \leq h\}$ for a real valued function H and a real number h. To write $R_A(\mathbf{X})$ in this form, note that

$$\begin{aligned} R_A(\mathbf{X}) &= \left\{(\mu,\sigma^2) : \overline{X} - \frac{q_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}} \leq \mu \leq \overline{X} + \frac{q_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}},\ \frac{(n-1)S^2}{\chi^2_{(n-1),(1-\frac{\alpha}{2})}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{(n-1),\frac{\alpha}{2}}}\right\} \\ &= \left\{(\mu,\sigma^2) : |\overline{X}-\mu| - \frac{q_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}} \leq 0,\ \left(\frac{(n-1)S^2}{\chi^2_{(n-1),(1-\frac{\alpha}{2})}} - \sigma^2\right)\left(\frac{(n-1)S^2}{\chi^2_{(n-1),\frac{\alpha}{2}}} - \sigma^2\right) \leq 0\right\} \\ &= \left\{(\mu,\sigma^2) : \max\left(|\overline{X}-\mu| - \frac{q_{1-\frac{\alpha}{2}}\sigma}{\sqrt{n}},\ \left[\frac{(n-1)S^2}{\chi^2_{(n-1),(1-\frac{\alpha}{2})}} - \sigma^2\right]\left[\frac{(n-1)S^2}{\chi^2_{(n-1),\frac{\alpha}{2}}} - \sigma^2\right]\right) \leq 0\right\}. \end{aligned}$$

Using the information given, we have the following simplified expressions :

$$\begin{aligned} R_A(\mathbf{X}) &= \left\{(\mu,\sigma^2) : \max\left(|\mu| - \frac{q_{0.975}\sigma}{\sqrt{10}},\ \left[\frac{9}{\chi^2_{9,0.975}} - \sigma^2\right]\left[\frac{9}{\chi^2_{9,0.025}} - \sigma^2\right]\right) \leq 0\right\} \\ R_B(\mathbf{X}) &= \left\{(\mu,\sigma^2) : 10(\log\sigma^2 - \log 0.9) - 10 + \frac{9}{\sigma^2} + \frac{10\mu^2}{\sigma^2} \leq \chi^2_{2,0.95}\right\} \\ R_C(\mathbf{X}) &= \left\{(\mu,\sigma^2) : \frac{10\mu^2}{\sigma^2} + \frac{5(1-\sigma^2)^2}{\sigma^4} \leq \chi^2_{2,0.95}\right\} \quad \text{and} \\ R_D(\mathbf{X}) &= \left\{(\mu,\sigma^2) : 10\mu^2 + 5(1-\sigma^2)^2 \leq \chi^2_{2,0.95}\right\}. \end{aligned}$$

The four confidence regions are displayed in the plots below (see Figures 1-3) with the x-axis for μ and the y-axis for σ^2.
It is observed that as the sample size grows, the confidence regions become more concentrated around the true value of μ and σ^2, namely, $\mu = 0$ and $\sigma^2 = 1$. The shapes of the three large sample confidence regions (in particular R_B and R_D) are quite similar when the sample size is large indicating that they will have similar properties (e.g., coverage probability, area etc.). The exact confidence region is always trapezoidal in shape.

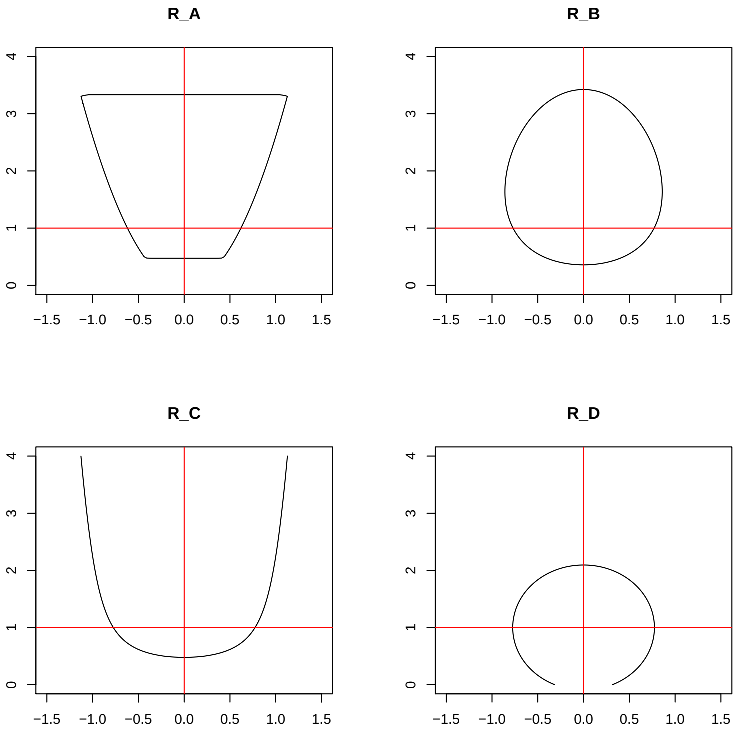

Figure 1 – Plots for $n = 10$

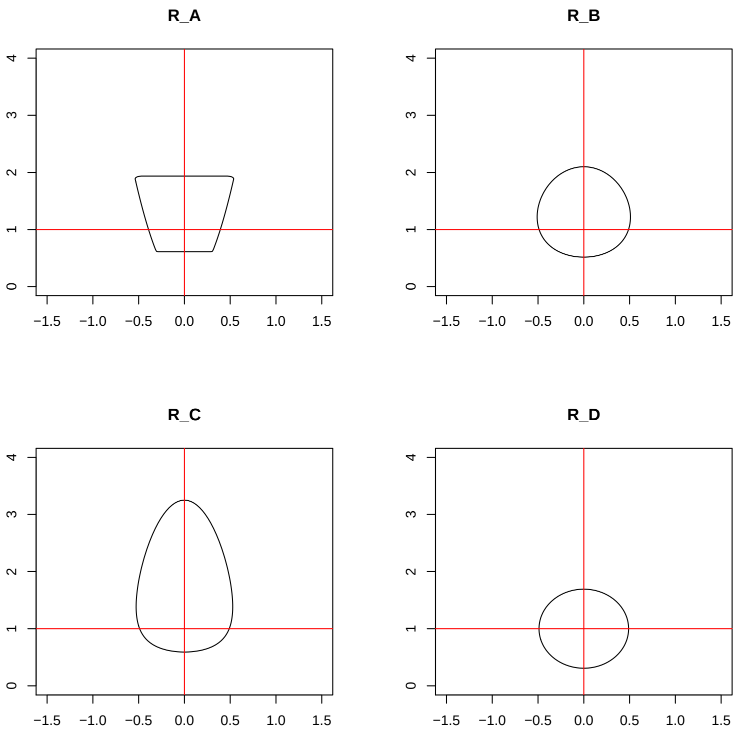

Figure 2 – Plots for $n = 25$

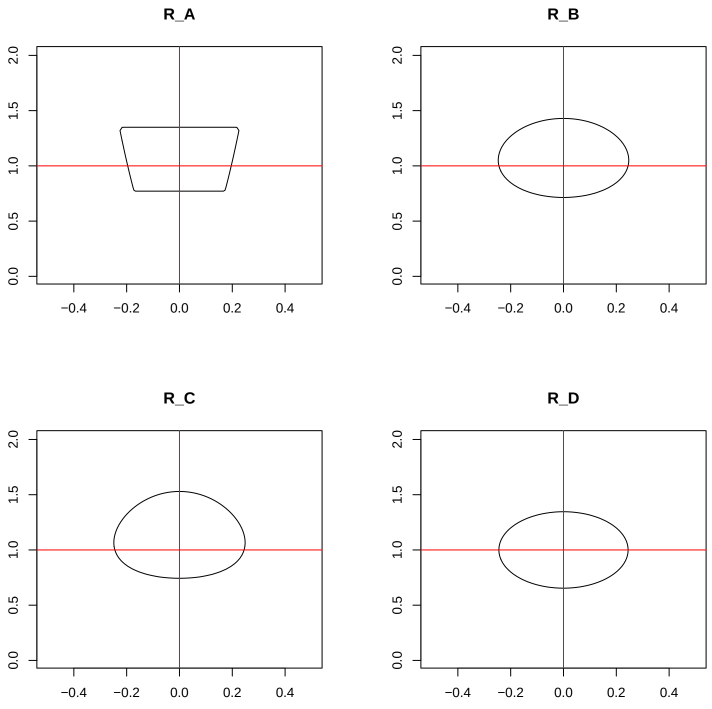

Figure 3 – Plots for $n = 100$ (note the change of scale of axes)

Assignment 5. (i).

```
# We are generating 100 binimials with n=trials and p=true.p,
 feed it to prop.test and get the p-values.
# The experiment is repeated nexp times.
p <- matrix(replicate(positions*nrep,prop.test(rbinom(1,trials,true.p),
              trials,true.p)$p.value),nrep)

# Every row contains the 100 p-values
dim(p)
[1] 1000  100

# Take the minimum of each p-value and test if it's significant
mean(apply(p,1,min)<alpha)
[1] 0.482
```

We have a significant result in nearly half of the cases, while under H_0 we expect to have water in only 5% of the sites.

(ii). When we increase α the probability of having a false positive is nearly 1.

```
alpha=0.05
mean(apply(p,1,min)<alpha)
[1] 0.973
```

(iii). For $\alpha = 0.01$ the adjusted p-values will be 0.015, while for $\alpha = 0.04$ they will be 0.04. Here the code

```
pa <- apply(p,1,p.adjust,method="bonferroni")
mean(apply(pa,2,min)<alpha)
```

```
pa <- apply(p,1,p.adjust,method="holm")
mean(apply(pa,2,min)<alpha)

 pa <- apply(p,1,p.adjust,method="hochberg")
mean(apply(pa,2,min)<alpha)
```

(iv). `prop.test` is using the normal approximation. It will give you warning when true.p$\times$trials is less than 5 (because of the Chi.square test). To overcome it, you could use `binom.test`

(v). To smile, look here https ://xkcd.com/882/