STA 610 Homework 5

Yuren Zhou

October 28, 2024

Question 1.

(a) (2 points)

- Macro explanatory variable: w_j.
- Micro explanatory variable: $x_{i,j}$.
- Fixed effect parameters: $\beta_0, \beta_1, \beta_2$.
- Random effects: $a_{0,j}, a_{1,j}$.
- Variance and covariance parameters: Ψ, σ^2.

(b) (2 points) The macro explanatory variable w_j does not possess within-group differences. Adding a random effect for it will almost duplicate the random intercept $a_{0,j}$.

(c) (2 points) The LRT statistic follows the distribution $\frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$, which is the mixture of χ_1^2 and χ_2^2 distributions. Letting $P_{\chi_p^2}(\cdot)$ denote the c.d.f. of χ_p^2 distribution, then the p-value of an LRT statistic λ is

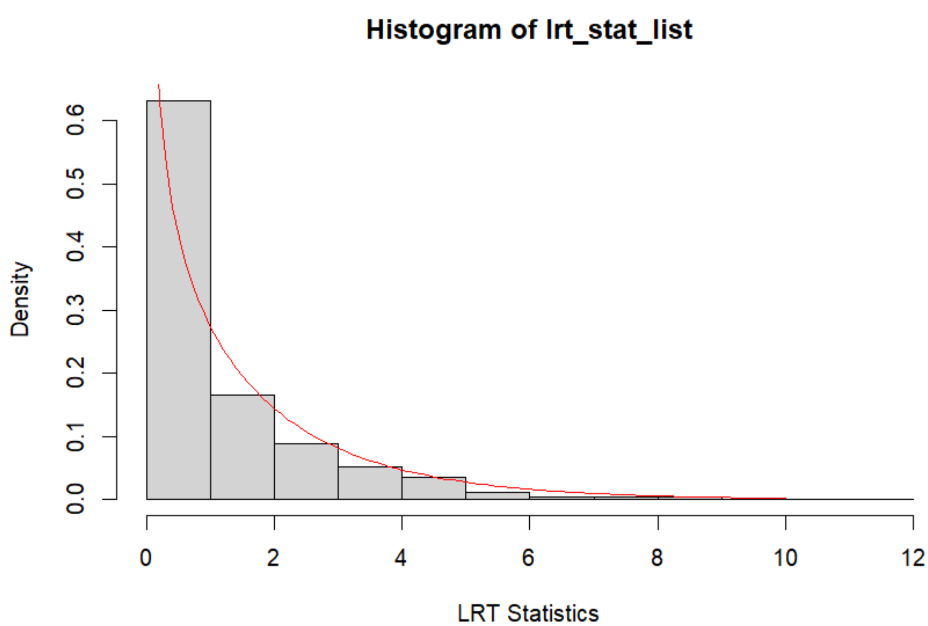
$$1 - \frac{1}{2}P_{\chi_1^2}^{-1}(\lambda) - \frac{1}{2}P_{\chi_2^2}^{-1}(\lambda).$$

(d) (2 points)

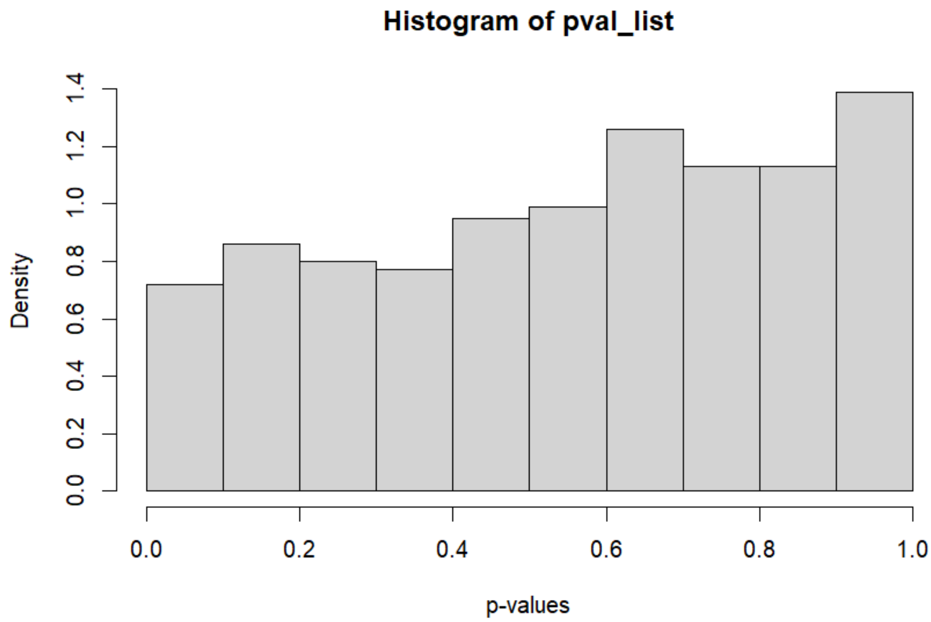
```
library(lme4)

lrt_pval <- function(
    seed,
    m = 20, n = 20,
    beta0 = 1, beta1 = 1, beta2 = 1,
    sigma = 1, psi0 = 1
){
  set.seed(seed)

  x <- rnorm(m * n)
  w <- rnorm(m)
  eps <- rnorm(m * n) * sigma
  a0 <- rnorm(m) * psi0
  y <- beta0 + beta1 * x + beta2 * rep(w, each = n) + rep(a0, each = n) + eps
  g <- as.factor(rep(1:m, each = n))

  model0 <- lmer(y ~ x + w[g] + (1 | g), REML = FALSE)
  model1 <- lmer(y ~ x + w[g] + (1 + x | g), REML = FALSE)
  lrt_stat <- 2 * (logLik(model1) - logLik(model0))
  pval <- 1 - (pchisq(lrt_stat, 1) + pchisq(lrt_stat, 2)) / 2
```

```
  return(c(lrt_stat, pval))
}


T = 1000
lrt_stat_list <- rep(0, T)
pval_list <- rep(0, T)
seed_list <- sample(1:1e8, T, replace = TRUE)
for (t in 1:T){
  result <- lrt_pval(seed_list[t])
  lrt_stat_list[t] <- result[1]
  pval_list[t] <- result[2]
}

hist(lrt_stat_list, probability = TRUE, xlab = "LRT Statistics")
curve((dchisq(x, 1) + dchisq(x, 2)) / 2,
      from = 0, to = 10, col = "red", add = TRUE)

hist(pval_list, probability = TRUE, xlab = "p-values")
```

Histogram of LRT statistics is roughly similar to its asymptotic distribution $\frac{1}{2}\chi^2_1 + \frac{1}{2}\chi^2_2$ under the null model:



Histogram of p-values is roughly similar to a uniform distribution $U(0, 1)$, although slightly skewed:



(e) (2 points) Repeat with the code in (d) with different values of parameters. Very open-ended.