1 Laboratorinis darbas

5 grupė: Arnas Kazanzvičius, Arnas Usonis, Lukas Janušauskas, Simonas Lapinskas

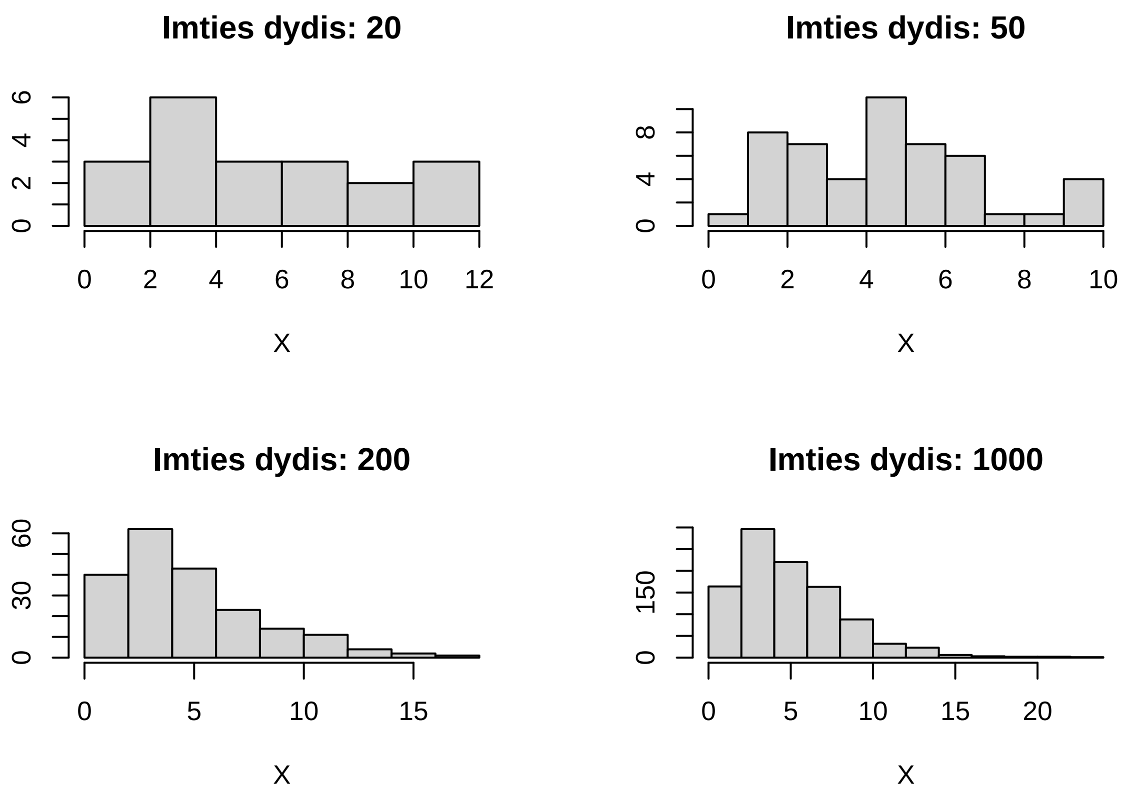
Skirstinys: χ^2

6. Fiksavome, pasirinktą a.d. parametrų rinkinį (`k=5`). Sugeneravome χ^2_5 duomenų rinkinius su 20, 50, 200, 1000 imčių dydžiais.

```
k <- 5
n <- c(20, 50, 200, 1000)

set.seed(42)

imtis1 <- rchisq(n[1], k)
imtis2 <- rchisq(n[2], k)
imtis3 <- rchisq(n[3], k)
imtis4 <- rchisq(n[4], k)
```

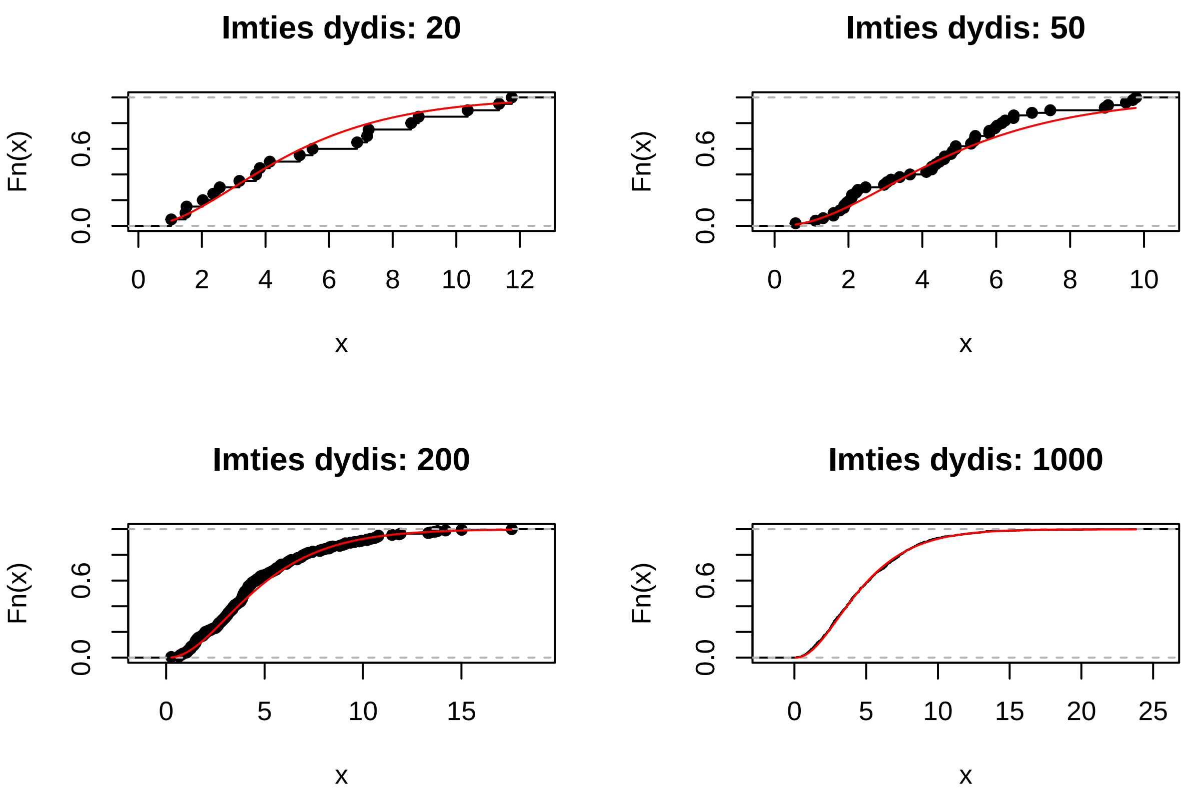
6a) Nubrėžėme histogramas

```
# Apibrėžiame pagalbinę funkciją, kadangi grafikai labai panašūs
plot_chisq_sample <- function(sample) {
  # sample - imtis, kurią įdedame į funkciją

  # Sudarome pavadinimą, į kurį įeis imties dydis
  imties_dydis <- length(sample)
  pavadinimas <- paste0("Imties dydis: ", imties_dydis)

  # Nubrėžiame histogramą
  hist(sample, main=pavadinimas,
       xlab = "X", ylab = "",
       freq = TRUE)
}

par(mfrow=c(2,2))
plot_chisq_sample(imtis1)
plot_chisq_sample(imtis2)
plot_chisq_sample(imtis3)
plot_chisq_sample(imtis4)
```



6b) Empirinės pasiskirstymo funkcijos mūsų darbe imčių pasiskistymo funkcijos pateikiamos su teorine pasiskirstymo funkcija.

```r
nubrezti_chisq_empirini <- function(sample, df=5) {
  # sample - imtis.
  # df - chi kvadratu parametras

  # Sudarome pavadinimą
  imties_dydis <- length(sample)
  pavadinimas <- paste0("Imties dydis: ", imties_dydis)

  # Nubrėžiame empirinę pasiskirstymo funkciją
  plot(ecdf(sample), main=pavadinimas)

  # Nubrėžiame teorinę pasiskirstymo funkciją
  x <- seq(min(sample), max(sample), by=0.01)
  lines(x, pchisq(x, df), col = 'red')
}

par(mfrow=c(2,2))

suppressWarnings({
  nubrezti_chisq_empirini(imtis1)
  nubrezti_chisq_empirini(imtis2)
  nubrezti_chisq_empirini(imtis3)
  nubrezti_chisq_empirini(imtis4)
})
```



7. Momentų metodu išvestas parametro k įvertinys:

Remiantis 3. punktu prisimename, kad $EX = k$. Pirmasis žingsnis, sudarant įverčius momentų metodu yra momentų prilyginimas empiriniams momentams. Taigi EX prilyginame $\overline{X}$. Gauname parametro k įvertinį $\widetilde{k}$:

$$\widetilde{k} = \overline{X} \tag{1}$$

```
imtys <- list(imtis1, imtis2, imtis3, imtis4)

sapply(imtys, function(x)
  paste(length(x), "dydžio imties parametro įvertinys", mean(x)))
```

```
## [1] "20 dydžio imties parametro įvertinys 5.42851954560543"
## [2] "50 dydžio imties parametro įvertinys 4.49953597150517"
## [3] "200 dydžio imties parametro įvertinys 4.82725161603656"
## [4] "1000 dydžio imties parametro įvertinys 4.99404810200757"
```

8. Didžiausio tikėtinumo metodu taip pat gavome įverčius. Tačiau, kadangi skaičiavimai per daug komplikuoti, naudojomės `optim` funkcija.

```
# install.packages('likelihoodExplore')
library(likelihoodExplore)

pakoreguota_tiketinumo <- function(x, par) {
  # Pakoreguojame tikėtinumo funkciją, kad tiktų optim funkcijai
  return( -1 * likchisq(x=x, df=par) )
}

mle_chisq_ivertis <- function(imtis) {
  res <- optim(par=c(1),                 # Pradedame nuo 1
```

```
                fn=pakoreguota_tiketinumo, # Pakoreguojame tikėtinumo funkciją
                method="L-BFGS-B",         # Naudojame L-BFGS optimizatorių
                x=imtis)
  return ( res$par )
}


sapply(imtys, function(x)
  paste(length(x), "dydžio imties parametro įvertinys", mle_chisq_ivertis(x)))
```

```
## [1] "20 dydžio imties parametro įvertinys 5.30031169143056"
## [2] "50 dydžio imties parametro įvertinys 4.76282529559125"
## [3] "200 dydžio imties parametro įvertinys 4.79013509203223"
## [4] "1000 dydžio imties parametro įvertinys 4.93233518543483"
```

9. Palyginkime 7 ir 8 punktuose gautus rezultatus

Nors, davus mažą imtį, gauti geresni rezultatai, panaudojus didžiausio tikėtinumo metodą, tačiau momentų metodo įverčiai, iš rezultatų atrodo, greičiau konverguoja link tikrojo parametro(5).

10. Parametrų patikimumo intervalai

Iš χ^2 apibrėžimo žinome, kad a.d. $X \sim \chi^2_k$, jei $X = \sum_{i=1}^{k} N_i$, kur $N_i \sim N(0,1)$. Todėl, n dydžio imties empirinio vidurkio skirstinį nustatome štai taip:

$$n\overline{X} = \sum_{j=1}^{n} X_j = \sum_{j=1}^{n} \sum_{i=1}^{k} N_{ji} \tag{2}$$

Kadangi kiekvienas N_{ji} yra pasiskirstęs pagal standartinį normalųjį skirstinį, tai vėl gauname χ^2 skirstinį:

$$n\overline{X} \sim \chi^2_{nk} \tag{3}$$

Taigi, pasikliautinio intervalo skaičiavimas (čia x_p - p-tasis χ^2_{nk} kvartilis):

$$P(x_{(1-\alpha)/2} < n\hat{k} < x_{(1+\alpha)/2}) \tag{4}$$

$$P(\frac{x_{(1-\alpha)/2}}{n} < \hat{k} < \frac{x_{(1+\alpha)/2}}{n}) \tag{5}$$

```
confidence_interval_chisq <- function(sample, alpha) {
  lower <- (1 - alpha) / 2
  upper <- (1 + alpha) / 2

  lower_b <- qchisq(lower, sum(sample)) / length(sample)
  upper_b <- qchisq(upper, sum(sample)) / length(sample)

  return( c(lower_b, upper_b) )
}

lapply(imtys, function(x)
  paste(length(x), "dydžio imties parametro pasikliautinio intervalo rėžiai",
        confidence_interval_chisq(x, 0.9)))
```

```
## [[1]]
## [1] "20 dydžio imties parametro pasikliautinio intervalo rėžiai 4.27607029946644"
## [2] "20 dydžio imties parametro pasikliautinio intervalo rėžiai 6.69455758007594"
##
## [[2]]
## [1] "50 dydžio imties parametro pasikliautinio intervalo rėžiai 3.8251888171799"
## [2] "50 dydžio imties parametro pasikliautinio intervalo rėžiai 5.21934234370739"
##
## [[3]]
## [1] "200 dydžio imties parametro pasikliautinio intervalo rėžiai 4.47163450665412"
## [2] "200 dydžio imties parametro pasikliautinio intervalo rėžiai 5.19423774046643"
##
## [[4]]
## [1] "1000 dydžio imties parametro pasikliautinio intervalo rėžiai 4.83080552588767"
## [2] "1000 dydžio imties parametro pasikliautinio intervalo rėžiai 5.15956468681696"
```