2.1

V tomto příkladu chceme vzájemně porovnat průběh tří funkcí, a sice (a) $y = \frac{x}{1+x}$; (b) $y = \log(1+x)$ a (c) $y = x$. Řešení příkladu je velmi jednoduché. Vygenerujeme posloupnost bodů osy x a pro každý bod posloupnosti vypočítáme hodnotu podle předpisu funkce (a), resp. (b), či (c). Následně vykreslíme křivky těchto tří funkcí (jednak pro větší rozsah osy x (0–5) a osy y (0–5) a potom pro výřez, tj. pro rozsah osy x 0-0.05 a pro rozsah osy y 0-0.05.). Křivky vzájemně porovnáme.

Příprava podkladů

`x <- seq(...)` ... posloupnost od 0 do 5 o délce minimálně 512. (d = 512)

`y1 <- ...` ... vektor hodnot $\frac{x}{x+1}$ (d = 512)

`y2 <- ...` ... vektor hodnot $\log(x+1)$ (d = 512)

`y3 <- ...` ... vektor hodnot x (d = 512)

Graf - větší rozsah os x a y

`plot(x, y1, type='n', xlim=c(0,5), ylim=c(0,5), las=1, ylab=... )` ... příprava prázdného grafu

(type='n' … prázdný graf; xlim … rozsah osy x; ylim … rozsah osy y; las=1 … otočení měřítka osy y o 90°)

`lines(x, y1, col=..., lwd=..., lty=...)` ... křivka $y = \frac{x}{x+1}$ (např. červená)

`lines(x, y2, ...)` ... křivka $y = \log(x+1)$ (např. modrá)

`lines(...)` ... křivka $y = x$ (např. zelená)

`legend('topleft', lwd=..., lty=..., col=..., legend=..., bty='n')` ... legenda

('topleft' … umístění; lwd … vektor šířek čar; lty … vektor typů čar; col … vektor barev čar; legend … vektor popisků legendy; bty='n' … okolo legendy nebude rámeček)

Graf - užší rozsah os x a y

`plot(x, y3, xlim=c(0,0.05), ylim=c(0,0.05), ...)` ... příprava prázdného grafu

`lines(x, y1, ...)` ... křivka $y = \frac{x}{x+1}$

`lines(...)` ... křivka $y = \log(1+x)$

`lines(...)` ... křivka $y = x$

`legend(...)` ... legenda

Závěr: Z grafu vidíme, že $\frac{x}{1+x} \leq \log(1+x) \leq x$.

Poznámka: Uvedenou nerovnost $\frac{x}{1+x} \leq \log(1+x) \leq x$, jejíž platnost jsme v tomto příkladu graficky ověřili, využijeme ve 3. cvičení při porovnání testovacích statistik U_S, U_W a U_{LR} testu o μ, když σ^2 neznáme.

2.2

V tomto příkladu chceme graficky porovnat konvergenci Studentova rozdělení t_n ke standardizovanému normálnímu rozdělení N(0,1) pro $n \to \infty$. Konvergenci vizualizujeme pomocí animace. Nejprve vytvoříme funkci normalni_student(n), jejímž vstupem bude počet stupňů volnosti Studentova rozdělení t_n a výstupem bude dvojice grafů vykreslených vedle sebe. Na levém grafu bude zobrazena křivka hustoty rozdělení N(0,1) a rozdělení t_n, na pravém grafu bude zobrazena křivka distribuční funkce rozdělení N(0,1) a rozdělení t_n. Funkci normalni_student( ) následně použijeme k tvorbě animace.

povinný vstupní argument funkce

```
normalni_student <- function(n){
```

příprava podkladů

```
  mu <- ... parametr μ rozdělení N(0,1)
  sigma <- ... parametr σ rozdělení N(0,1)
  xfit <- seq(...)... posloupnost x od -3 do 3 o délce 512
  yfit <- dnorm(xfit, mean=..., sd=...)... hustota rozdělení N(0,1) nad posl. xfit (d=512)
  zfit <- dt(xfit, df=...)... hustota rozdělení t_n nad posl. xfit (d=512)
```

graf křivek hustoty

```
  par(mfrow=c(1,2), mar=c(...))... rozdělení obrázku na 2 okna vedle sebe + okraje 5,4,2,1
  plot(xfit, yfit, type='l', ylim=c(0,0.5), xlab='', ylab=...,
       col=..., lwd=..., las=...)... graf s křivkou hustoty N(0,1) (např. modrá)
  lines(xfit, zfit,...)... křivka hustoty t_n (např. červená)
  mtext('x', side=1, line=2.5)... popisek osy x
  mtext(paste('n =', n),...)... druhý popisek osy x (n=...); automatická aktualizace
                                hodnoty podle parametru n
  legend(..., lty=..., col=..., legend=..., bty=...)... legenda
```

příprava podkladů

```
  yfit <- pnorm(...)... distr. fce rozdělení N(0,1) nad posl. xfit (d=512)
  zfit <- pt(...)... distr. fce rozdělení t_n nad posloupností xfit (d=512)
```

graf křivek distr. fce

```
  plot(xfit, yfit,...)... graf s křivkou distr. fce N(0,1) (např. modrá)
  lines(...)... křivka distr. fce rozdělení t_n (např. červená)
  mtext(...)... popisek osy x
  mtext(...)... druhý popisek osy x (n=...); automatická aktualizace hodnoty podle parametru n
  legend(...)... legenda
}
```

animace

```
n <- seq(...)... posloupnost 5, 10, 50, 100, 500, 1000
oopts <- ani.options()
ani.record(reset=T)... vymazání předchozích grafů z paměti
saveLatex(for(i in 1:length(n)){... cyklus pro vytvoření obrázků animace
            normalni_student(n=n[i])
          },...)... animace
```

Závěr: Z průběhu animace vidíme, jak se pro $n \to \infty$ blíží tvar křivky hustoty (resp. distribuční fce) Studentova rozdělení t_n k tvaru křivky hustoty (resp. distribuční fce) rozdělení N(0,1).

V tomto příkladu si graficky ověříme tři vztahy, se kterými se setkáte na přednáškách ze S1-11.

(1) $\chi^2_1(1-\alpha) = (u_{1-\alpha/2})^2$

(2) $F_{1,n}(1-\alpha) = (t_n(1-\alpha/2))^2$

(3) $\chi^2_n(1-\alpha) = \Gamma_{\frac{n}{2},2}(1-\alpha)$

$n=10$

Řešení příkladu si demonstrujeme na situaci (1). Nejprve vygenerujeme posloupnost hodnot (alpha) z intervalu (0.01; 0.99). Dále pro každou hodnotu posl. alpha vypočítáme hodnotu $(1-\alpha)$-kvantilu χ^2 rozdělení o 1 stupni volnosti, tj. $\chi^2_1(1-\alpha)$ a hodnotu druhé mocniny $(1-\alpha/2)$-kvantilu rozdělení $N(0,1)$, tj. $(u_{1-\alpha/2})^2$. Následně vykreslíme graf, kde na ose x bude posloupnost alpha a na ose y posloupnost hodnot kvantilů $\chi^2_1(1-\alpha)$, resp. $(u_{1-\alpha/2})^2$. Obě křivky vzájemně porovnáme.

V situaci (2) a (3) postupujeme analogicky jako v situaci (1). Ve (3) situaci potřebujeme vypočítat hodnoty $(1-\alpha/2)$-kvantilu gamma rozdělení s parametry $\frac{n}{2}$ (shape) a 2 (scale). K tomu použijeme funkci qgamma().

---

(1)
```
alpha <- seq(...)
yfit <- qchisq(1-alpha,...)
zfit <- qnorm(1-alpha/2,...)^2
par(mar=...)
plot(alpha, yfit,...)
lines(alpha, zfit,...)
mtext(expression(alpha),...)
legend(...)
```
- alpha <- seq(...) ... posl. α od 0.01 do 0.99 o délce minimálně d=512
- yfit <- qchisq(1-alpha,...) ... hodnoty kvantilů $\chi^2_1(1-\alpha)$ nad posl. alpha (d=512)
- zfit <- qnorm(1-alpha/2,...)^2 ... druhé mocniny kvantilů $u_{1-\alpha/2}$ nad posl. alpha (d=512)
- par(mar=...) ... okraje grafu 4,4,2,1
- plot(alpha, yfit,...) ... graf s křivkou kvantilů $\chi^2_1(1-\alpha)$ (např. modrá plná křivka)
- lines(alpha, zfit,...) ... křivka druhých mocnin kvantilů $u_{1-\alpha/2}$ (např. červená přerušovaná křivka)
- mtext(expression(alpha),...) ... popisek osy x (α)
- legend(...) ... legenda

(2)
- n <- ... zadaná hodnota parametru n
- yfit <- qf(...) ... hodnoty kvantilů $F_{1,n}(1-\alpha)$ nad posl. alpha (d=512)
- zfit <- qt(...)^2 ... druhé mocniny kvantilů $t_n(1-\alpha/2)$ nad posl. alpha (d=512)
- plot(alpha, yfit,...) ... graf s křivkou kvantilů $F_{1,n}(1-\alpha)$ (např. modrá plná křivka)
- lines(alpha, zfit,...) ... křivka druhých mocnin kvantilů $t_n(1-\alpha/2)$ (např. červená přerušovaná křivka)
- mtext(...) ... popisek osy x (α)
- legend(...) ... legenda

(3)
- yfit <- qchisq( ) ... hodnoty kvantilů $\chi^2_n(1-\alpha)$ nad posl. alpha (d=512)
- zfit <- qgamma(1-alpha, shape = n/2, scale = 2) ... hodnoty kvantilů $\Gamma_{\frac{n}{2},2}(1-\alpha)$ nad posl. alpha (d=512)
- plot(...) ... graf s křivkou kvantilů $\chi^2_n(1-\alpha)$ (např. modrá plná křivka)
- ⋮
- legend(...) ... legenda

Závěr: Z vykreslených grafů vidíme že (1) $\chi^2_1(1-\alpha)$ a $(u_{1-\alpha/2})^2$ jsou identické; (2) $F_{1,n}(1-\alpha)$ a $(t_n(1-\alpha/2))^2$ jsou identické; (3) $\chi^2_n(1-\alpha)$ a $\Gamma_{\frac{n}{2},2}(1-\alpha)$ jsou identické, tedy, že všechny tři rovnosti skutečně platí.

2.4

V tomto příkladu si graficky ověříme, že pokud náhodná veličina $X \sim N(\mu, \sigma^2)$, potom pro výběrový rozptyl S_n^2, resp. pro výběrovou sm. odchylku S_n platí:

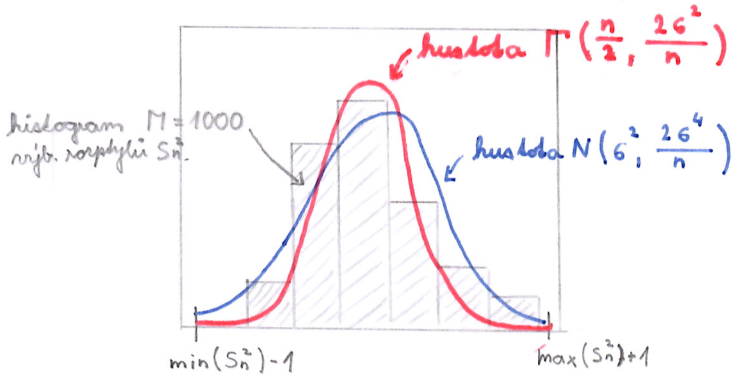
(1) $S_n^2 \sim \Gamma\left(\frac{n}{2}, \frac{2\sigma^2}{n}\right)$ exaktně

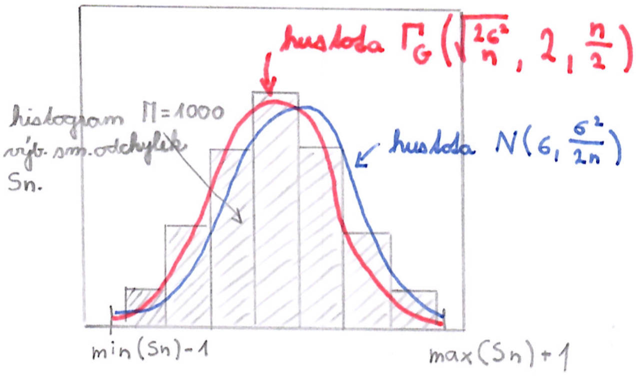
(2) $S_n^2 \sim N\left(\sigma^2, \frac{2\sigma^4}{n}\right)$ asymptoticky

(3) $S_n \sim \Gamma_G\left(\sqrt{\frac{2\sigma^2}{n}}, 2, \frac{n}{2}\right)$ exaktně

(4) $S_n \sim N\left(\sigma, \frac{\sigma^2}{2n}\right)$ asymptoticky.

$\Gamma\left(\frac{n}{2}, \frac{2\sigma^2}{n}\right)$ je gamma rozdělení s parametry shape $= \frac{n}{2}$ a scale $= \frac{2\sigma^4}{n}$. $\Gamma_G\left(\sqrt{\frac{2\sigma^2}{n}}, 2, \frac{n}{2}\right)$ je zobecněné gamma rozdělení s parametry scale $= \sqrt{\frac{2\sigma^2}{n}}$, $d = 2$ a $k = \frac{n}{2}$.





Nejprve naprogramujeme funkci **rozdeleni_Sn2_Sn()**, která pro zadané σ a n vrátí dvojici výše nakreslených histogramů. Pomocí funkce nakonec vytvoříme animaci. Hustotu rozdělení $\Gamma_{\frac{n}{2}, \frac{2\sigma^2}{n}}$ vypočítáme pomocí funkce **dgamma()**. Hustotu rozdělení $\Gamma_G\left(\sqrt{\frac{2\sigma^2}{n}}, 2, \frac{n}{2}\right)$ vypočítáme pomocí funkce **dgengamma.stacy()** z knihovny **VGAM**.

```
Var.n <- function(X) {      funkce, jejímž vstupem je náhodný výběr X1,...,Xn a výstupem výběrový rozptyl Sn^2
    n <- length(...)        ... délka vektoru x
    Sn2 <- ...              1/n sum_{i=1}^{n} (x_i - x̄)^2 ... 1/n(sum((x-mean(x))^2))
    return(...)             ... vrať hodnotu Sn2
}

rozdeleni_Sn2_Sn <- function(sigma, n, M=1000){      výstupem je vektor M=1000 výb. rozptylů Sn^2
  # příprava podkladů
    Sn2 <- replicate(M, Var.n(rnorm(n, 0, sigma)))   ... vytvoř vektor pseudonáhodných čísel o délce n z rozdělení N(0, sigma^2) a vypočítej z něj výb. rozptyl Sn^2. Toto proveď M=1000.
    Sn <- sqrt(...)         ... √Sn2
    x.Sn2 <- seq(...)       ... posl. od min(Sn2) do max(Sn2) o délce 512
    x.Sn <- seq(...)        ... posl. od min(Sn) do max(Sn) o délce 512
    y.e <- dgamma(...)      ... vektor hodnot hustoty rozdělení Γ(n/2, 2σ^2/n) nad posl. Sn2 (d=512)
    y.a <- dnorm(...)       ... vektor hodnot hustoty rozdělení N(σ^2, 2σ^4/n)  -||-  -||-
    z.e <- VGAM::dgengamma.stacy(xfit.Sn, scale = √(2σ^2/n), d= ..., k= ...)   vektor hodnot hustoty Γ_G(√(2σ^2/n), 2, n/2) nad posl. Sn (512)
    z.a <- dnorm(...)       ... vektor hodnot hustoty rozdělení N(σ, σ^2/2n) nad posl. Sn (d=512)

  # Sn^2
    par(mfrow= ... , mar= ... )   ... [][] , okraje 5,4,2,1
    hist(Sn2, prob=T, ylim=c(0, as.numeric(max(y.e, y.a))+0.1),
         density=30, col= ..., border= ..., ...)   ... histogram Sn^2     (hustota šrafování výplně histogramu)
    box(bty='o')            ... rámeček okolo histogramu
    lines(x.Sn2, y.e, ...)  ... křivka hustoty Γ(n/2, 2σ^2/n)
    lines(x.Sn2, y.a, ...)  ... křivka hustoty N(σ^2, 2σ^4/n)
```

histogram

```
mtext(expression(S[n]^2), ...)... popisek osy x (Sn^2)
mtext(bquote(paste(mu == 0, ', ', sigma^2 == .(sigma^2), ', ',
                   n == .(n))), ...)... druhý popisek osy x (μ=0, σ^2=..., n=...)
legend(...)... legenda
```

histogram Sn

```
hist(Sn, prob=..., ylim=c(0, as.numeric(max(zfit.e, zfit.a))+0.1),
     density=..., col=..., border=..., ...)... histogram Sn
box(...)... rámeček okolo grafu
lines(x.Sn, z.e, ...)... křivka hustoty Γ_G(√(2σ^2/n), 2, n/2)
lines(x.Sn, z.a, ...)... křivka hustoty N(σ, σ^2/2n)
mtext(expression(...), ...)... popisek osy x (Sn)
mtext(bquote(paste(...)), ...)... druhý popisek osy x (μ=0, σ^2=..., n=...)
legend(...)... legenda
}
```

```
n <- seq(...) ... posloupnost 5, 10, 50, 100, 500, 1000
oopts <- ani.options()
ani.record(reset=T)... vymazání obrázků z paměti
saveLatex(for (i in 1:length(n)){ cyklus generující obrázky animace
        rozdeleni_Sn2_Sn(sigma=..., n=n[i])
    },...)... animace
```

Závěr: Z uvedené animace pro S_n^2 je patrné, že exaktní rozdělení $\Gamma(\frac{n}{2}, \frac{2\sigma^2}{n})$ popisuje rozdělení S_n^2 velmi dobře i pro malé rozsahy náhodných výběrů n. Pro $n \to \infty$ se rozdělení S_n^2 blíží k normálnímu rozdělení a tedy $S_n^2 \sim N(\sigma^2; \frac{2\sigma^4}{n})$ asymptoticky. Z grafu dále vidíme, jak pro $n \to \infty$ se křivky obou hustot čím dál více jedna druhé podobají.

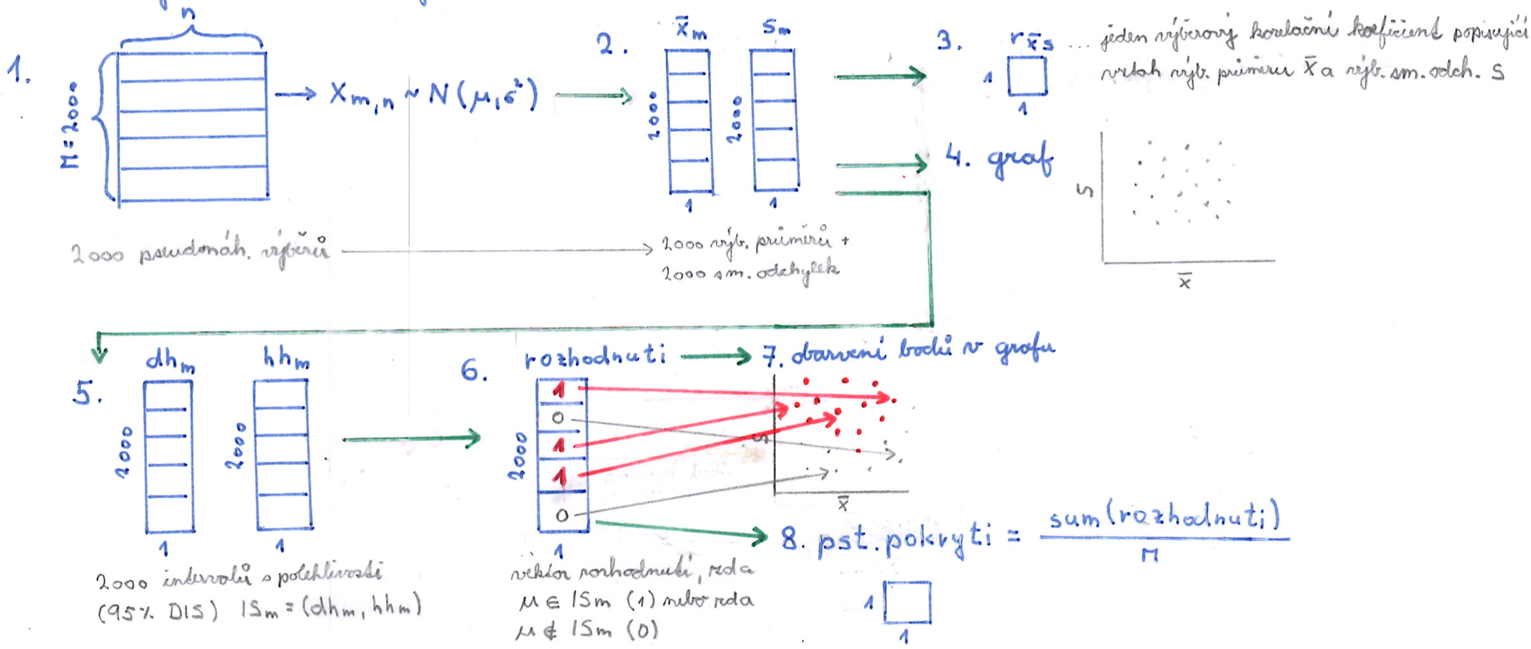
Z animace pro S_n je patrné že exaktní rozdělení $\Gamma_G(\sqrt{\frac{2\sigma^2}{n}}, 2, \frac{n}{2})$ popisuje rozdělení S_n velmi dobře i pro malé rozsahy náh. výběrů n. Pro $n \to \infty$ se rozdělení S_n tvarem blíží k normálnímu rozdělení $N(\sigma, \frac{\sigma^2}{n})$. S $n \to \infty$ se také křivky obou hustot vzájemně čím dál více podobají, a tedy pro $n \to \infty$ se $S_n \sim N(\sigma, \frac{\sigma^2}{n})$, tj. $S_n \overset{A}{\sim} N(\sigma, \frac{\sigma^2}{n})$.

Poznámka: Asymptotické rozdělení S_n^2 i S_n v praxi často používáme, protože normální rozdělení je jednodušší, než Γ, resp. Γ_G rozdělení a má skvělé vlastnosti. Znalost exaktních rozdělení S_n^2 a S_n je však také velmi užitečná, nicméně často opomíjená.

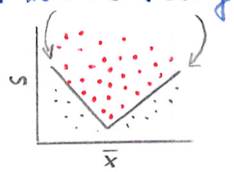
2.5

V tomto příkladu si graficky ověříme nezávislost parametrů μ a σ^2 normálního rozdělení $N(\mu,\sigma^2)$. Pomocí grafů si také zobrazíme pravděpodobnost pokrytí 95% Waldova empirického DIS pro μ když σ^2 neznáme. Nezávislost μ a σ^2 a pst pokrytí 95% DIS zhodnotíme nejprve za předpokladu, že $X \sim N(\mu,\sigma^2)$. Nakonec budeme zjišťovat, zda a případně jak se situace změní, pokud bude homogenní normální rozdělení $N(150,10^2)$ z 10% znečištěno rozdělením $N(150,20^2)$, tj. normálním rozdělením se stejnou střední hodnotou, ale větším rozptylem.

Příklad vyřešíme v následujících krocích (postup pro $n=5$; $X \sim N(\mu,\sigma^2)$):



9. hraniční čáry



m_i ... posl. od min(m) do max(m) o délce 2000 (osa x)

pro každé m_i vypočítáme hodnotu s_i (osa y) tak, aby platilo:

$dh_{m_i} = hh_{m_i} = \mu$, tj. $\left(m_i - t_{1-\alpha/2}\frac{s_i}{\sqrt{n}} = \mu = m_i - t_{\alpha/2}\frac{s_i}{\sqrt{n}}\right)$

I. $\quad$ II.

I. $m_i - t_{1-\alpha/2}\frac{s_i}{\sqrt{n}} = \mu$

$$s_i = \frac{\mu - m_i}{t_{1-\alpha/2}}\sqrt{n}$$

II. $m_i - t_{\alpha/2}\frac{s_i}{\sqrt{n}} = \mu$

$$s_i = \frac{\mu - m_i}{t_{\alpha/2}}\sqrt{n}$$

$$t_{1-\alpha/2} = -t_{\alpha/2} \Longrightarrow s_i = \left|\frac{\mu - m_i}{t_{1-\alpha/2}}\sqrt{n}\right| = \left|\frac{\mu - m_i}{t_{\alpha/2}}\sqrt{n}\right|$$

Protože v rámci příkladu máme vyhodnotit nezávislost μ a σ^2 a aktuální pst. pokrytí $1-\hat{\alpha}$ v šesti různých situacích ((1) $X \sim N(\mu,\sigma^2)$; (2) $X \sim pN(\mu,\sigma^2) + (1-p)N(\mu,\sigma_2^2)$) × ((a) $n=5$; (b) $n=50$; (c) $n=100$), bude nejlepší vytvořit funkci pokryti( ), která pro zadané μ, σ, σ_2, n, M, p a α vrátí graf s hodnotou výběrového korel. koef. $r_{\bar{x},s}$ a aktuální psti pokrytí $1-\hat{\alpha}$.

Na základě výběrového korel. koeficientu $r_{\bar{x},s}$ zhodnotíme pro každou situaci nezávislost mezi μ a σ^2. Nakonec pro každou situaci porovnáme hodnotu aktuální psti pokrytí $(1-\hat{\alpha})$ s nominální psti pokrytí $(1-\alpha)=0.95$ Waldova 95% empirického DIS pro μ když σ^2 neznáme.

defaultní nastavení $\sigma_2 = 6$

```
pokryti <- function(mu = ...(150), sigma = ...(10), sigma2 = sigma, n = ...(5), M = ...(2000), p = ...(0.9)) {
```

generování dat

X <- matrix(NA, ..., ...) ... příprava prázdné matice dimenze $M \times n$

for(i in 1:M) { cyklus: univerzální kód pro generování homogenních dat (pokud $\sigma = \sigma_2$) i heterogenních dat (směsi) (pokud $\sigma \neq \sigma_2$)

  bin <- rbinom(n, 1, p) ... posl. n hodnot $\sim Bin(1, p)$ ... $X = 1$ s pstí p, $X = 0$ s pstí $(1-p)$

  X[i,][bin == 1] <- rnorm(sum(bin == 1), mu, sigma) ... na i-tý řádek matice X na pozice, kde bin == 1 vlož hodnoty z $N(\mu, \sigma^2)$

  X[i,][bin == 0] <- rnorm(sum(bin == 0), mu, sigma2) ... na i-tý řádek matice X na pozice, kde bin == 0 vlož hodnoty z $N(\mu, \sigma_2^2)$

}

! pozor na syntaxi fce rnorm(n, mean = μ, sd = σ) !

příprava podkladů

m <- apply(X, 1, mean) ... vektor výb. průměrů $\bar{x}_m$ (d = 2000)

s <- apply(...) ... vektor výb. sm. odchylek s_m (d = 2000)

r <- round(cor(...), digits = ...) ... výb. korel. koef cor(m, s) zaokr. na 4 des. místa (d = 1)

dh <- $\bar{x}_m - t_{n-1}(1-\alpha/2)\frac{s}{\sqrt{n}}$ ... vektor dolních hranic 95% DIS pro μ když σ^2 neznáme (d = 2000)

hh <- $\bar{x}_m - t_{n-1}(\alpha/2)\frac{s}{\sqrt{n}}$ ... vektor horních hranic – || – – || – (d = 2000)

rozhodnuti <- (dh < mu) & (mu < hh) ... vektor rozhodnutí, zda $\mu = 150 \in IS_m$, $m = 1, \dots, 2000$ (d = 2000)

pst.pokryti <- $\frac{\Sigma(\text{rozhodnuti})}{M}$ ... aktuální pst pokrytí $1-\hat{\alpha}$

vykreslení grafu

par(...) ... okraje 5, 4, 2, 1

plot(m, s, asp = T, xlim = c(130, 170), ylim = c(0, 30), cex = 0.4, ...) ... graf; šedé body $(\bar{x}_m, s_m)$

zachová se poměr stran osy x a y; osa x a y bude pevná pro všechny grafy; velikost bodů bude 0.4 (defaultně cex = 1)

mtext(expression(bar(x)), ...) ... popisek osy x, tj. $\bar{x}$

mtext(bquote(paste(n == .(n), ', ', r == .(r), ', ', 1 - alpha == .(1-alpha), ', ', 1 - widehat(alpha) == .(pst.pokryti), sep = '')), ...) ... druhý popisek osy x

points(m[rozhodnuti == 1], s[rozhodnuti == 1], col = 'coral', cex = ..., ...) ... červené obarvení bodů $(\bar{x}_m, s_m)$, pro které platí, že $\mu \in IS_m$.

mi <- seq(...) ... posl. od min(m) do max(m) o délce 2000

si <- $\left|\frac{m_i - \mu}{t_{n-1}(1-\alpha/2)}\sqrt{n}\right|$ ... vektor hodnot s_i odpovídající hodnotám m_i tak, aby platilo $dh_{m_i} = hh_{m_i} = \mu$

lines(mi, si, ...) ... hraniční čáry, kde $dh_{m_i} = hh_{m_i} = \mu$.

}

$X \sim N(\mu, \sigma^2)$, $n = 5$: **pokryti(n = 5)** ... použití fce pokryti() k vykreslení grafu pro homogenní data (n = 5)

$X \sim pN(\mu, \sigma^2) + (1-p)N(\mu, \sigma_2^2)$, $n = 5$: **pokryti(sigma2 = 20, n = 5)** ... vykreslení grafu pro směs (n = 5)

Z obou obrázků vidíme, že pro homogenní data i pro data s 10% příměsí hodnot s vyšším rozptylem jsou parametry μ a σ^2 nezávislé. Vykreslené body jsou nahodile rozmístěny po ploše grafu a hodnota výb. korel. koeficientu je velmi blízká nule.

Z grafů dále vidíme, že pro homogenní data i pro zadanou směs se hodnota aktuální psti pokrytí $(1-\hat{\alpha})$ 95% Waldova empirického DIS pro μ když σ^2 neznáme pohybuje okolo nominální psti pokrytí $(1-\alpha) = 0.95$.

Jediný rozdíl, který u obou grafů pozorujeme, je, že v případě směsi jsou body z hlediska osy y více rozházené (mají větší rozptyl). To způsobuje 10% příměs dat s vyšším rozptylem.

Analogicky bychom vykreslili a vyhodnotili dvojici grafů pro (b) n = 50 a (c) n = 100.

V tomto příkladu si pomocí simulační studie vizualizujeme aktuální pst pokrytí 95% Waldova empirického DIS pro μ při známém σ^2. Aktuální pst pokrytí si ukážeme přímo na grafu s vynesenými hranicemi $M = 100$ 95% DIS. Hodnotu aktuální psti pokrytí prostudujeme ve třech různých situacích:

(1) u homogenních dat $X \sim N(\mu, \sigma^2)$;

(2) u dat $X \sim N(\mu, \sigma^2)$ s 10% příměsí hodnot se stejnou střední hodnotou a vyšším rozptylem, tj. + 10% hodnot z $N(\mu, \sigma_2^2)$;

(3) u dat $X \sim N(\mu, \sigma^2)$ s 10% příměsí hodnot se stejným rozptylem a nižší střední hodnotou, tj. + 10% hodnot z $N(\mu_2, \sigma^2)$.

Poznámka: Analogickou simulační studii bychom mohli vytvořit pro libovolný typ IS (oboustranný, levostranný, pravostranný, 90%, 95%, 99%, …, parametru $\mu, \sigma^2, \lambda, \rho, \mu_1 - \mu_2, \sigma_1^2/\sigma_2^2, \dots$)

Řešení příkladu: Nejprve vygenerujeme $M = 100$ náh. výběrů. Z každého výběru vypočítáme výb. průměr $\bar{x}_m$, výběrovou sm. odchylku s_m a nakonec dolní hranici dh_m a horní hranici hh_m 95% Waldova emp. DIS pro μ, když σ^2 známe. $IS = (dh, hh) = \left(m - u_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} ; m - u_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right)$. Dále vytvoříme vektor rozhodnuti, pro který platí, že $\text{rozhodnuti}_m == 1$, pokud $\mu = 5 \in IS_m$ a $\text{rozhodnuti} == 0$, pokud $\mu = 5 \notin IS_m$. Dále vypočítáme aktuální pst pokrytí $1 - \hat{\alpha}$ jako $(1 - \hat{\alpha}) = \frac{\Sigma(\mu \in IS)}{M} = \frac{\Sigma(\text{rozhodnuti})}{M}$ a aktuální hl. významnosti $\hat{\alpha} = 1 - (1 - \hat{\alpha})$. Následně vykreslíme graf, kde každý IS bude reprezentován svislou úsečkou, přičemž IS_m, pro které $\mu = 5 \notin IS_m$, barevně odlišíme. Nakonec stanovíme hranice 95% DIS pro aktuální pst. pokrytí $(1 - \hat{\alpha})$ podle vzorce:

$$(dh_{1-\hat{\alpha}} ; hh_{1-\hat{\alpha}}) = \left((1-\hat{\alpha}) - u_{1-\alpha/2} \widehat{SE} ; \quad (1-\hat{\alpha}) - u_{\alpha/2} \widehat{SE}\right),$$

kde $\widehat{SE} = \sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{M}}$. Hodnoty $(1-\alpha)$, $(1-\hat{\alpha})$, $dh_{1-\alpha}$, $hh_{1-\alpha}$ vložíme do souhrnné tabulky.

Protože v rámci příkladu máme aktuální pst pokrytí prozkoumat v šesti situacích, vytvoříme funkci DIS.mu( ), která pro zadané $n, \mu, \mu_2, \sigma, \sigma_2, M, p$ a α vykreslí graf a vrátí příslušnou souhrnnou tabulku výsledků. Šestici grafů a tabulek potom vygenerujeme pomocí této funkce.

```
DIS.mu <- function(n, mu=...(5), mu2=mu, sigma=...(√15), sigma2=sigma, M=...(100), p=...(0.9),
                   alpha=...(0.05)) {

příprava dat
  X <- matrix(...) ... příprava NA matice dimenze M×n
  for(i in 1:M) { cyklus; analogický jako v příkladu 2.5
     bin <- rbinom(...)
     X[i, ][...] <- rnorm(...) ... X~N(μ,σ²)
     X[i, ][...] <- rnorm(...) ... X~N(μ2,σ2²)
  }

podklady
  m <- apply(...) ... vektor výb. průměrů x̄m (d=100)
  s <- apply(...) ... vektor výb. sm. odchylek sm (d=100)
  dh <- m - u_{1-α/2} σ/√n ... vektor dolních hranic 95% DIS pro μ když σ² známe (d=100)
  hh <- m - u_{α/2} σ/√n ... vektor horních hranic -||- (d=100)
  rozhodnuti <- dh < μ < hh ... viz příklad 2.5
```

příprav

a.pst.pokryti <- $\frac{\Sigma(\text{rozhodnuti})}{M}$ ... aktuální pst pokrytí $(1-\hat{\alpha})$

a.alpha <- $1-(1-\hat{\alpha})$ ... aktuální hl. významnosti $\hat{\alpha}$

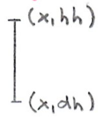
x ← 1:M ... diskrétní posloupnost čísel 1, 2, 3, ..., 99, 100

vykreslení grafu

par(...) ... okraje $\overset{1}{4\square 1}_{5}$

plot(x, dh, type=..., ylim=c(min(dh), max(hh)), xlab='',

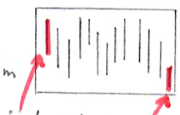
ylab=expression($\overset{\mu}{...}$), ...) ... příprava prázdného grafu



mtext(bquote(...), ...) ... popisek osy x ... $1-\hat{\alpha}$ = ...

mtext(paste(...), ...) ... druhý popisek osy x ... n = ...

segments(x, dh, x, hh, col='grey40') ... dokreslení šedých úseček IS_m



segments(x[rozhodnuti==0], dh[rozhodnuti==0], ...) ... obarvení úseček IS_m, pro které $\mu = 5 \notin IS$.

abline(h=mu) ... vodorovná čára v hodnotě $\mu = 5$

souhrnná tabulka

SE <- $\sqrt{\frac{\hat{\alpha}(1-\hat{\alpha})}{M}}$ ... standardní chyba $\widehat{SE}$ pst pokrytí $(1-\hat{\alpha})$

dh.alpha <- $1-\hat{\alpha}-u_{1-\alpha/2}\widehat{SE}$ ... dolní hranice 95% DIS pro $(1-\hat{\alpha})$, tj. $dh_{1-\hat{\alpha}}$

hh.alpha <- $1-\hat{\alpha}-u_{\alpha/2}\widehat{SE}$ ... horní hranice -||- -||- , tj. $hh_{1-\hat{\alpha}}$

tab <- data.frame($1-\alpha$, $1-\hat{\alpha}$, $dh_{1-\hat{\alpha}}$, $hh_{1-\hat{\alpha}}$) ... souhrnná tabulka výsledků

return(...) ... navrací souhrnnou tabulku tab.

}

(1)

$X \sim N(5, 15)$, n=5: DIS_mu(n=5) ... použití fce DIS_mu( ) pro vykreslení grafu a výpočtu tabulky pro homogenní data (n=5)

$X \sim N(5, 15)$, n=1000: DIS_mu(n=1000) ... -||- -||- (n=1000)

Analogicky bychom vykreslili dvojice grafů (n=5; n=1000) pro každou směs.

Poznámka: Komentáře jednotlivých grafů viz zadání samostatné práce.

Nezapomeňte u každého grafu rozhodnout, zda je DIS konzervativní nebo liberální. Závěr uveďte v řešení jako komentář v R-Skriptu (varianta .R) nebo slovně v .pdf souboru (varianta Rnw).