Laborator 2

Intervale de încredere

Obiectivul acestui laborator este de a ilustra noțiunea de interval de încredere și a face o serie de exemple.

1 Ilustrarea intervalelor de încredere pentru o populație normală

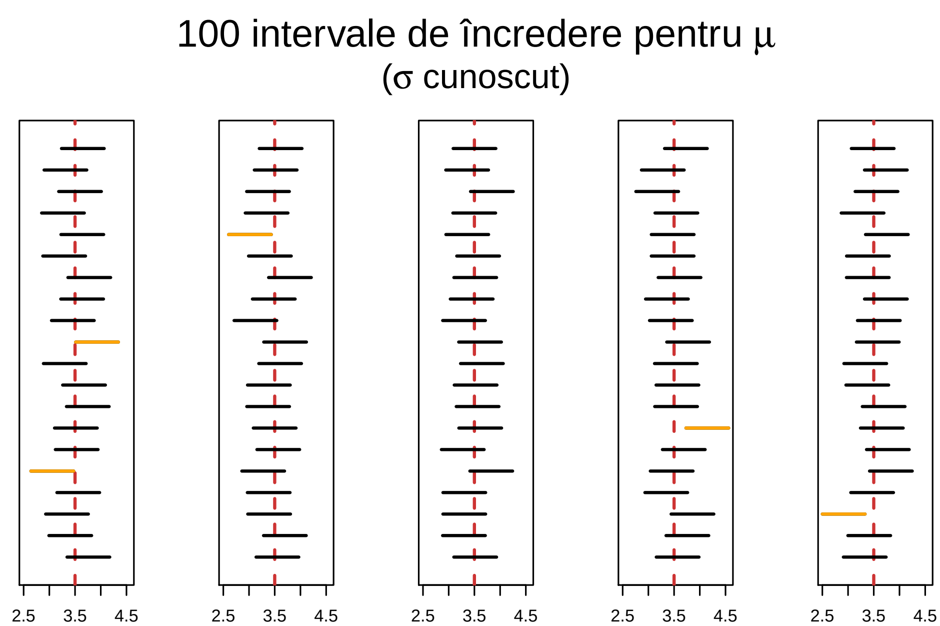
Generarea intervalelor de încredere:

```r
# cate panouri sa avem
p = 5

# nr de intervale de incredere per panou
n = 20

# talia esantionului
m = 50

# coeficient de incredere
alpha = 0.05

# media si sd populatia normala
mu = 3.5
sd = 1.5

lo3 <- hi3 <- lo2 <- hi2 <- lo <- hi <- vector("list", p)

for(i in 1:p) {
  dat = matrix(rnorm(n*m, mean = mu, sd = sd), ncol = m)

  # media si vaianta esantionului
  me = apply(dat,1,mean)
  se = apply(dat,1,sd)

  # calcul intervale de incredere
  lo[[i]] = me - qnorm(1-alpha/2)*sd/sqrt(m)
  hi[[i]] = me + qnorm(1-alpha/2)*sd/sqrt(m)

  lo2[[i]] = me - qnorm(1-alpha/2)*se/sqrt(m)
  hi2[[i]] = me + qnorm(1-alpha/2)*se/sqrt(m)

  lo3[[i]] = me - qt(1-alpha/2, m-1)*se/sqrt(m)
  hi3[[i]] = me + qt(1-alpha/2, m-1)*se/sqrt(m)
}
```

Intervale de încredere atunci când σ este cunoscut:

```r
r = range(unlist(c(lo,hi,lo2,hi2,lo3,hi3)))

par(mfrow=c(1,5), las=1, mar=c(5.1,2.1,6.1,2.1))
```

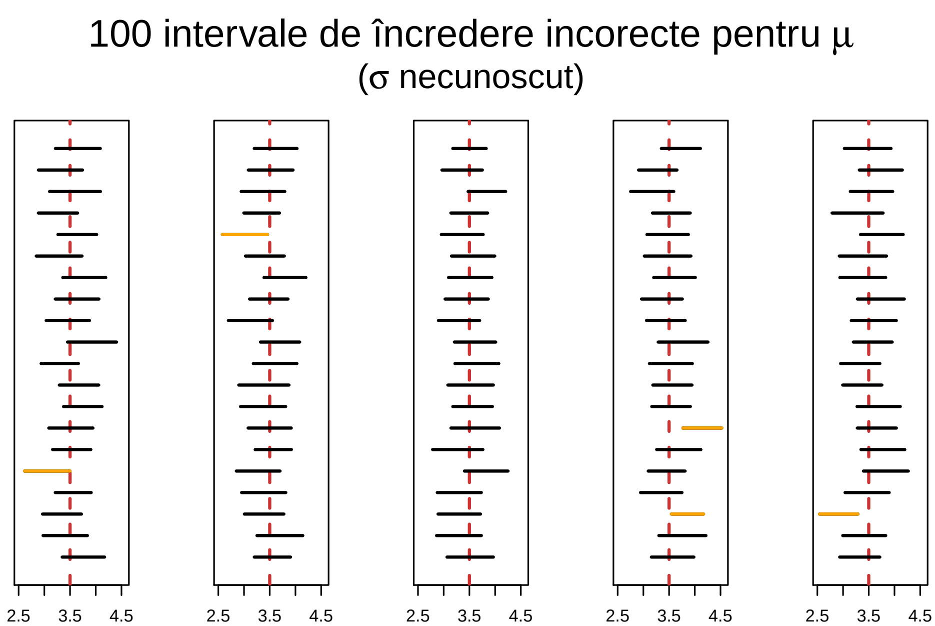
```
for(i in 1:p) {
  plot(0, 0, type="n",
       ylim = 0.5+c(0,n),
       xlim = r,
       ylab = "",
       xlab = "",
       yaxt = "n")

  abline(v = mu, lty=2, col="brown3", lwd=2)

  segments(lo[[i]], 1:n,
           hi[[i]], 1:n,
           lwd=2)

  o = (1:n)[lo[[i]] > 3.5 | hi[[i]] < 3.5]

  segments(lo[[i]][o], o,
           hi[[i]][o], o,
           lwd=2,col="orange")
}

par(mfrow=c(1,1))

mtext(expression(paste("100 intervale de încredere pentru ", mu)),
      side=3, cex=1.5, xpd=TRUE, line=4)
mtext(expression(paste("(",sigma," cunoscut)")), side=3, cex=1.3,
      xpd=TRUE,line=2.7)
```



Intervale de încredere **incorecte** atunci când σ nu este cunoscut:

```
par(mfrow=c(1,5), las=1, mar=c(5.1,2.1,6.1,2.1))
```

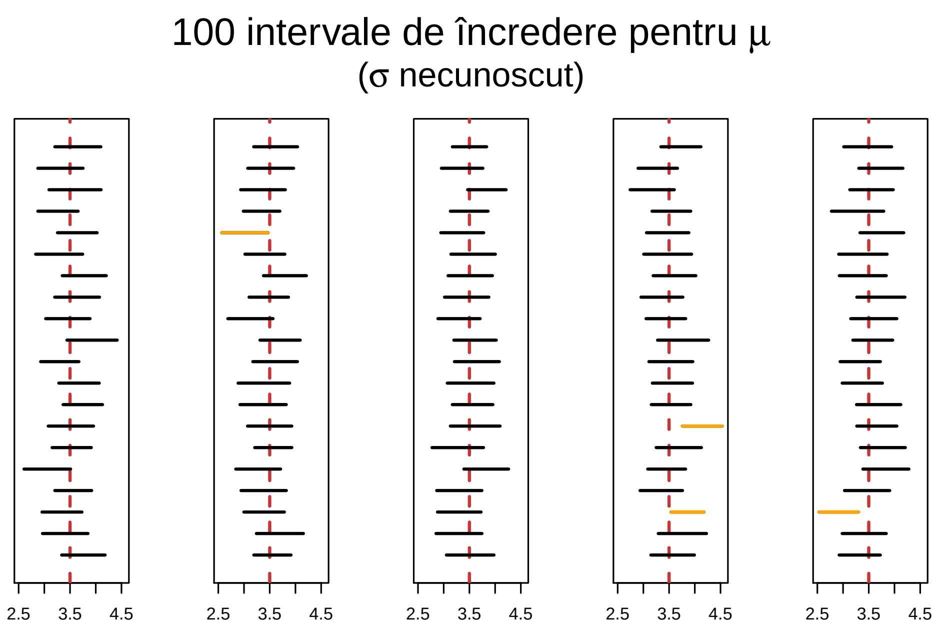
```
for(i in 1:p) {
  plot(0, 0,
       type="n",
       ylim=0.5+c(0,n),
       xlim=r,
       ylab="",
       xlab="",
       yaxt="n")

  abline(v = mu,lty = 2, col="brown3", lwd=2)

  segments(lo2[[i]], 1:n,
           hi2[[i]], 1:n,
           lwd=2)

  o = (1:n)[lo2[[i]] > 3.5 | hi2[[i]] < 3.5]

  segments(lo2[[i]][o],o,
           hi2[[i]][o],o,
           lwd=2, col="orange")
}

par(mfrow=c(1,1))
mtext(expression(paste("100 intervale de încredere incorecte pentru ", mu)),
      side=3, cex=1.5, xpd=TRUE, line=4)
mtext(expression(paste("(",sigma," necunoscut)")),
      side=3,cex=1.3,xpd=TRUE,line=2.7)
```



Intervale de încredere **corecte** atunci când σ nu este cunoscut:

```
par(mfrow=c(1,5), las=1, mar=c(5.1,2.1,6.1,2.1))
```

```r
for(i in 1:p) {
  plot(0,0,
       type="n",
       ylim=0.5+c(0,n),
       xlim=r,
       ylab="",
       xlab="",
       yaxt="n")

  abline(v = mu, lty=2, col="brown3", lwd=2)

  segments(lo3[[i]],1:n,
           hi3[[i]],1:n,
           lwd=2)

  o = (1:n)[lo3[[i]] > 3.5 | hi3[[i]] < 3.5]

  segments(lo3[[i]][o],o,
           hi3[[i]][o],o,
           lwd=2, col="orange")
}
par(mfrow=c(1,1))

mtext(expression(paste("100 intervale de încredere pentru ", mu)),
      side=3, cex=1.5, xpd=TRUE, line=4)

mtext(expression(paste("(",sigma," necunoscut)")),
      side=3, cex=1.3, xpd=TRUE, line=2.7)
```



2 Ilustrarea probabilității de acoperire

2.1 Intervale de încredere de tip Wald

Fie $X_1, X_2, \dots, X_n$ un eșantion de talie n dintr-o populație Bernoulli de medie θ. Determinați un interval de încredere asimptotic pentru θ cu un coeficient de încredere $1-\alpha$.

Ilustrați grafic *probabilitatea de acoperire* $\mathbb{P}_\theta\left(IC^{1-\alpha}(\theta) \ni \theta\right)$ ca funcție de θ pentru diferite valori ale lui $n \in \{50, 100\}$ și $\alpha = 0.05$. Ce observați?

Știm că $\hat{\theta}_n = \bar{X}_n$ este estimatorul de verosimilitate maximă pentru θ și folosind proprietatea asimptotică a estimatorilor de verosimilitate maximă găsim că un interval de încredere asimptotic pentru θ este (folosid o înlocuire de tip Wald)

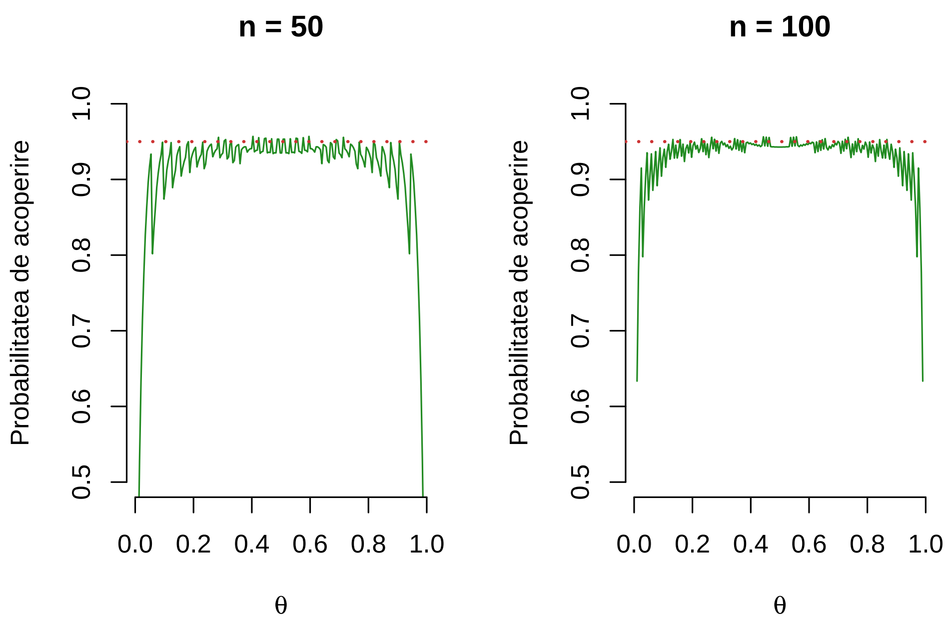
$$IC^{1-\alpha}(\theta) = \bar{X}_n \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{X}_n(1-\bar{X}_n)}{n}}.$$

Probabilitatea de acoperire este:

```r
binom.wald.cvg = function(theta, n, alpha) {
  z = qnorm(1 - alpha / 2)

  f = function(p) {
    t = 0:n

    s = sqrt(t * (n - t) / n)
    o = (t - z * s <= n * p & t + z * s >= n * p)

    return(sum(o * dbinom(t, size = n, prob = p)))
  }

  out = sapply(theta, f)
  return(out)
}
```

```r
# date intrare
par(mfrow = c(1,2))

n = 50
alpha = 0.05

theta = seq(0.01, 0.99, len=200)

plot(theta, binom.wald.cvg(theta, n, alpha),
     ylim=c(0.5, 1), type="l", lwd=1,
     bty = "n",
     col = "forestgreen",
     main = paste0("n = ", n),
     xlab = expression(theta),
     ylab = "Probabilitatea de acoperire")

abline(h = 1-alpha, lty=3, lwd=2,
       col = "brown3")
```

```
# al doilea grafic
n = 100

plot(theta, binom.wald.cvg(theta, n, alpha),
     ylim=c(0.5, 1), type="l", lwd=1,
     bty = "n",
     col = "forestgreen",
     main = paste0("n = ", n),
     xlab = expression(theta),
     ylab = "Probabilitatea de acoperire")

abline(h = 1-alpha, lty=3, lwd=2,
       col = "brown3")
```



Observăm că probabilitatea de acoperire tinde să fie mai scăzută decât pragul $1 - \alpha = 0.95$ ales pentru majoritatea valorilor lui θ.

Fie $X_1, X_2, \dots, X_n$ un eșantion de talie n dintr-o populație Exponențială de parametru θ. Determinați un interval de încredere asimptotic pentru θ cu un coeficient de încredere $1 - \alpha$.

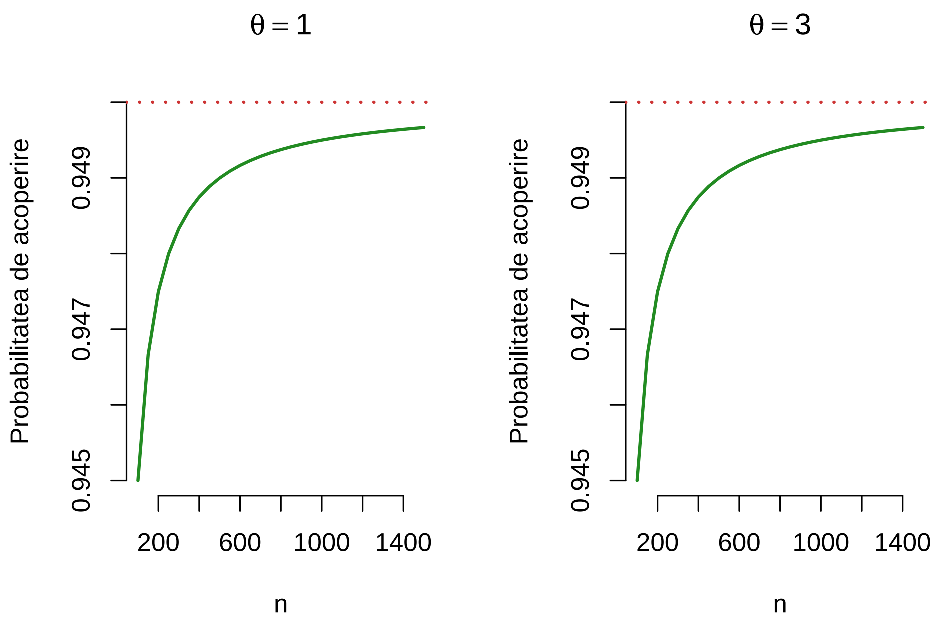
Ilustrați grafic *probabilitatea de acoperire* $\mathbb{P}_\theta\left(IC^{1-\alpha}(\theta) \ni \theta\right)$ ca funcție de n pentru diferite valori ale lui $\theta \in \{1, 3\}$ și $\alpha = 0.05$. Ce observați?

Știm că $\hat{\theta}_n = \bar{X}_n$ este estimatorul de verosimilitate maximă pentru θ și folosind proprietatea asimptotică a estimatorilor de verosimilitate maximă găsim că un interval de încredere asimptotic pentru θ este (folosid o înlocuire de tip Wald)

$$IC^{1-\alpha}(\theta) = \bar{X}_n \pm z_{1-\frac{\alpha}{2}} \frac{\bar{X}_n}{\sqrt{n}}.$$

```r
expo.wald.cvg = function(N, theta, alpha) {
    z = qnorm(1 - alpha / 2)

  f = function(n) {
    f1 = 1 - pgamma(n * theta / (1 - z / sqrt(n)),
                    shape=n, rate=1/theta)
    f2 = pgamma(n * theta / (1 + z / sqrt(n)),
                shape=n, rate=1/theta)
    return(1 - f1 - f2)
  }

  out = sapply(N, f)
  return(out)
}
```

```r
alpha = 0.05
n = seq(100, 1500, by=50)

par(mfrow = c(1,2))

plot(n, expo.wald.cvg(n, 1, alpha),
     ylim=c(0.945, 0.95), type="l", lwd=2,
     bty = "n", col = "forestgreen",
     main = TeX("$\\theta = 1$"),
     xlab="n", ylab="Probabilitatea de acoperire")

abline(h=1-alpha, lty=3, lwd=2,
       col = "brown3")

plot(n, expo.wald.cvg(n, 3, alpha),
     ylim=c(0.945, 0.95), type="l", lwd=2,
     bty = "n", col = "forestgreen",
     main = TeX("$\\theta = 3$"),
     xlab="n", ylab="Probabilitatea de acoperire")

abline(h=1-alpha, lty=3, lwd=2,
       col = "brown3")
```



2.2 Intervale de încredere folosid transformări stabilizatoare de varianță

Spune că o funcție g este stabilizatoare de varianță dacă verifică ecuația diferențială:

$$[g'(\theta)]^2 = c^2 I_1(\theta), \quad c > 0$$

unde $I_1(\theta)$ este informația lui Fisher.

Fie $X_1, X_2, \dots, X_n$ un eșantion de talie n dintr-o populație Bernoulli de medie θ. Determinați o funcție stabilizatoare de varianță și găsiâi un interval de încredere asimptotic pentru θ cu un coeficient de încredere $1-\alpha$.

Ilustrați grafic *probabilitatea de acoperire* $\mathbb{P}_\theta\left(IC^{1-\alpha}(\theta) \ni \theta\right)$ ca funcție de θ pentru diferite valori ale lui $n \in \{50, 100\}$ și $\alpha = 0.05$. Ce observați acum?

Observăm că pentru $g(\theta) = \arcsin\sqrt{\theta}$ avem

$$g'(\theta) = \frac{1}{2}\frac{1}{\sqrt{\theta(1-\theta)}}$$

deci

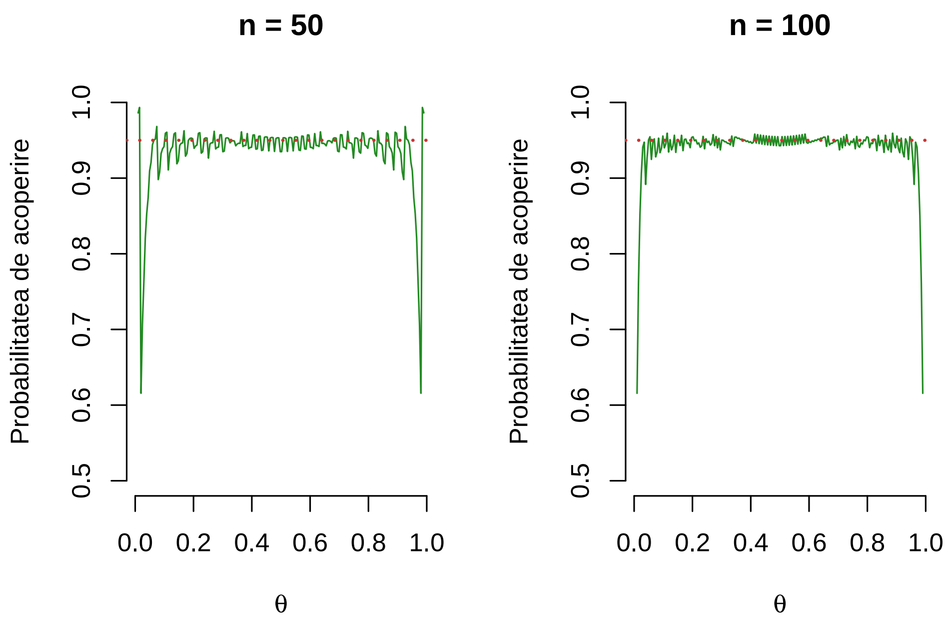
$$[g'(\theta)]^2 = \frac{1}{4}I_1(\theta)$$

și găsim un interval de încredere de tipul

$$IC^{1-\alpha}(\theta) = \sin^2\left(\arcsin\sqrt{\bar{X}_n} \pm z_{1-\frac{\alpha}{2}}\frac{1}{16n^2}\right)$$

```
binom.vst.cvg = function(theta, n, alpha) {
  z = qnorm(1 - alpha / 2)

  f = function(p) {
    t = 0:n
    a = asin(sqrt(t / n))
    s = z / 2 / sqrt(n)

    o = (a - s <= asin(sqrt(p)) & a + s >= asin(sqrt(p)))

    return(sum(o * dbinom(t, size=n, prob=p)))
  }

  out = sapply(theta, f)
  return(out)
}
```

```
# date intrare
par(mfrow = c(1,2))

n = 50
alpha = 0.05

theta = seq(0.01, 0.99, len=200)

plot(theta, binom.vst.cvg(theta, n, alpha),
     ylim=c(0.5, 1), type="l", lwd=1,
     bty = "n",
     col = "forestgreen",
     main = paste0("n = ", n),
     xlab = expression(theta),
     ylab = "Probabilitatea de acoperire")

abline(h = 1-alpha, lty=3, lwd=2,
       col = "brown3")

# al doilea grafic
n = 100

plot(theta, binom.vst.cvg(theta, n, alpha),
     ylim=c(0.5, 1), type="l", lwd=1,
     bty = "n",
     col = "forestgreen",
     main = paste0("n = ", n),
     xlab = expression(theta),
     ylab = "Probabilitatea de acoperire")

abline(h = 1-alpha, lty=3, lwd=2,
       col = "brown3")
```



Observăm că probabilitatea de acoperire în acest caz este mai aproape de ținta de $1-\alpha = 0.95$ comparativ cu exemplul anterior.