

Exemple de simulation sur la régression linéaire simple

Timothée Flutre (INRA)

12/02/2019

Question : comment évolue l'erreur standard de l'estimation de l'effet ($\hat{\beta}$) en fonction du nombre d'échantillons (n) ?

```
## create the vectors used to keep the results at the end
vec.n <- c(5, 10, 20, 50, 100, 200, 500)
vec.se.beta.hat <- rep(NA, length(vec.n))

calcGenoFreq <- function(maf){
  return(c((1 - maf)^2, 2 * (1 - maf) * maf, maf^2))
}

## set the inputs
set.seed(1234)
mu <- 50
f <- 0.3
beta <- 2.45
sigma <- 1

## "for" loop over the different sample sizes
for(i in 1:length(vec.n)){
  n <- vec.n[i]

  ## simulate data
  genotypes <- sample(x=c(0,1,2), size=n, replace=TRUE, prob=calcGenoFreq(f))
  errors <- rnorm(n=n, mean=0, sd=sigma)
  phenotypes <- mu + beta * genotypes + errors
  dat <- data.frame(x=genotypes, y=phenotypes)

  ## perform the inference
  fit <- lm(y ~ x, data=dat)

  ## extract the std error of beta-hat
  vec.se.beta.hat[i] <- summary(fit)$coefficients["x", "Std. Error"]
}

## plot the results
plot(vec.n, vec.se.beta.hat, type="b", las=1,
     xlab="nombre d'échantillons",
     ylab="erreur standard de l'estimation de l'effet",
     main=paste0("mu=", round(mu, 2),
                  ", f=", round(f, 2),
                  ", beta=", round(beta, 2),
                  ", sigma=", round(sigma, 2)))
```

$\mu=50, f=0.3, \beta=2.45, \sigma=1$

