

Initiation à l'analyse statistique des données expérimentales

(Les distributions statistiques sous R)

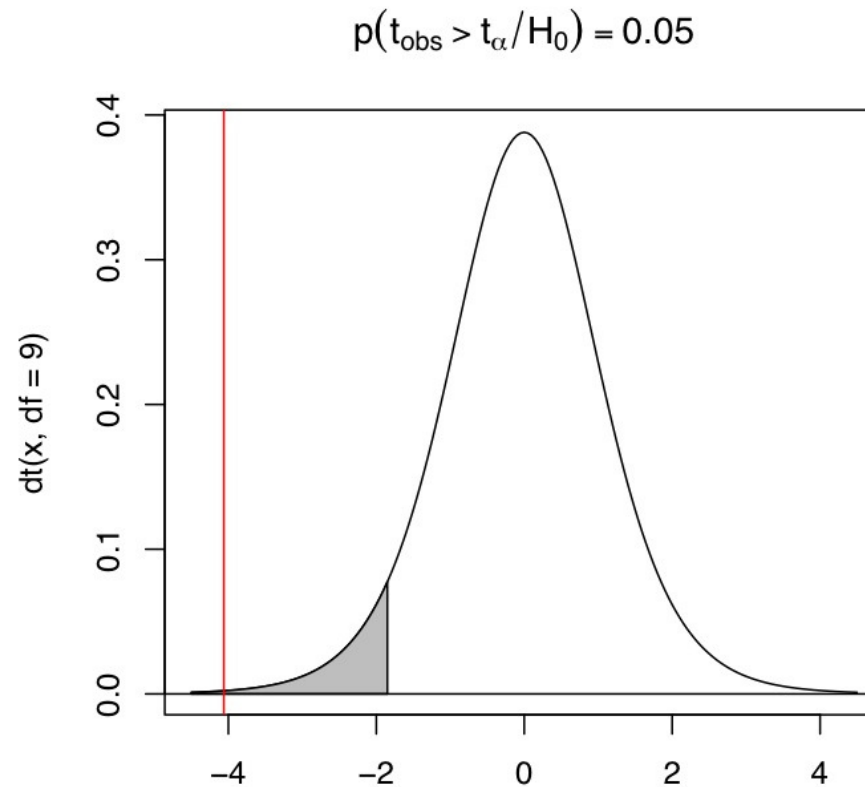
Christophe Pallier et Christophe Lalanne

`christophe.pallier@m4x.org / christophe.lalanne@gmx.net`

Mastère de Sciences Cognitives, EHESS – Paris 5 – ENS

Rappels

- ▶ probabilités vs. statistique
- ▶ lois de probabilités discrètes et continues
 - ▶ binomiale, poisson, ...
 - ▶ normale, χ^2 , Student, ...
- ▶ fonction de densité, fonction de répartition





Exemples de distributions

- ▶ sous R, il y a 4 fonctions principales pour travailler avec chaque fonction, avec comme préfixe: r-, d-, p-, q-
- ▶ e.g. `rnorm`, génère une séquence de nombres $\sim \mathcal{N}(\mu; \sigma)$; `pnorm` et `qnorm` donnent les probabilités et quantiles de F_X

- ▶ **exemples :**

- ▶ générateur de nombres aléatoires

```
round(runif(20), digits=2)
rnorm(10)+0.5*runif(10)
```

- ▶ jeu du pile ou face (1)

```
n <- 100
x <- sample(c(-1,1), n, replace=T)
plot(x, type='h', main="Bernoulli variables")
```

- ▶ jeu du pile ou face (2)

```
rbinom(100,1,0.5)
```



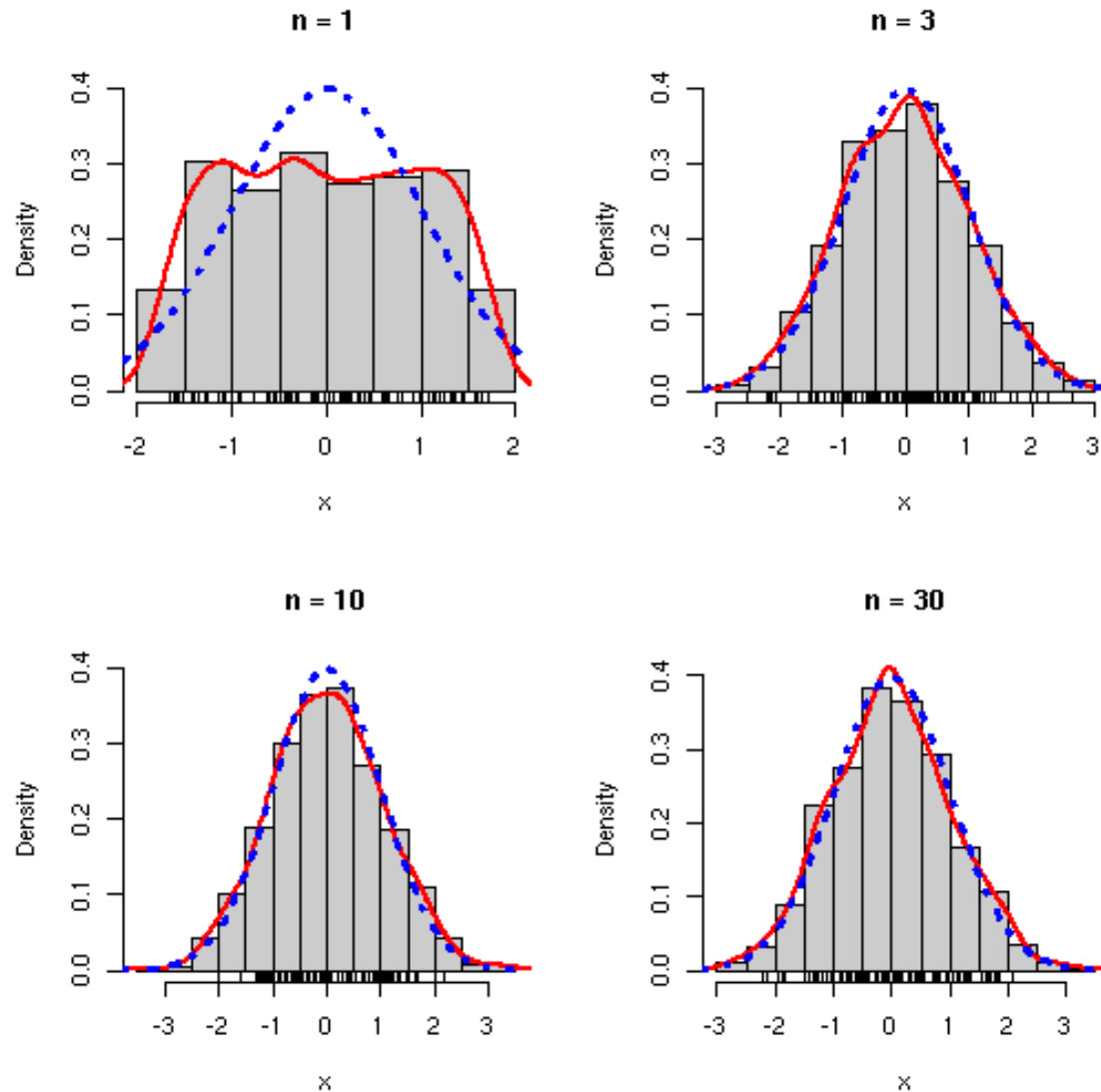
Paramètre et statistique

- ▶ diverses notions : paramètre, V.A., estimateur, statistique
- ▶ biais d'estimation et taille d'échantillon (LGN)
- ▶ **exemple :**

```
n <- seq(10,10000,by=10)
for (i in 1:length(n)) {
  x <- rbinom(n[i],1,0.5)
  f <- length(x[x==1])/length(x)
  plot(i,f,ylim=c(0,1),xlim=c(1,length(n)))
  par(new=T)
}
abline(h=0.5,col="red",lty=2)
```

La loi normale

- ▶ fondamentale en statistique (TCL, résidus)





Applications de la loi normale

- ▶ positionnement d'une statistique sur une loi connue
- ▶ e.g. situer un individu par rapport à une mesure de référence
- ▶ utilisation de la LNCR ($\frac{X-\mu}{\sigma} \sim \mathcal{N}(0; 1)$) avec les tables standards
- ▶ base du *test de typicalité* : notion d'extrêmérité d'une statistique (e.g. la moyenne d'un échantillon), par rapport à un seuil repère de 5 %

- ▶ **exemple :**

taille d'un individu

```
1-pnorm(180,mean=177,sd=7)
```

```
1-pnorm(188,mean=177,sd=7)
```

```
1-pnorm(198,mean=177,sd=7)
```

ou d'un groupe d'individus

```
x <- c(195,182,198,176,187,194)
```

```
1-pnorm(mean(x),mean=177,sd=7)
```



Résumés numérique et graphique

▶ graphiques :

```
plot(), boxplot(), hist(), stem(), stripchart(), qqnorm()
```

▶ synthèse numérique :

```
summary(), tapply()
```

▶ exemple :

```
x <- rnorm(100,mean=10,sd=2)
plot(x,pch=19)
a <- gl(2,50,100)
stripchart(x~a,method="jitter",vert=T)
boxplot(x~a)
summary(x)
tapply(x,a,mean)
y <- rnorm(100,0,(1+2*rbinom(100,1,0.05)))
hist(y)
plot(density(y),col="red",lty=3)
```



Exercices

- ▶ calculer les probabilités associées aux quantiles 1.64 et 1.96 de la LNCR
- ▶ calculer les bornes de l'intervalle centré sur la moyenne contenant 95 % des valeurs de la distribution de Student $T(9)$
- ▶ afficher sur le même graphique la loi de Student pour différents ddl
- ▶ un groupe d'individus tiré d'une population de référence de moyenne 30 et d'écart-type 10 ayant un score moyen de 49 peut-il être déclaré représentatif de cette population ?
- ▶ représenter des échantillons aléatoires $\sim \mathcal{N}(0; 1)$ de taille variable, tracer la fonction de densité estimée (`?density()`) et comparer celle-ci à la densité gaussienne correspondante (`?dnorm()`)
- ▶ générer deux échantillons aléatoires gaussiens ($\mu_1 = 5, \mu_2 = 7 ; \sigma = 1.5 * runif(1)$) de taille 30, créer une variable de classification à 2 modalités, et comparer les caractéristiques de ces échantillons