T.P. Variables aléatoires et inférence statistique (Labo 2)

201-9F6-ST : Statistiques appliquées à l'informatique

Marc-André Désautels

2017-12-08

Instructions:

1. Le but de ce T. P. est de vous familiariser avec le langage `R`. Il vous faudra trouver et utiliser les commandes appropriées pour répondre aux questions. Vous devez vous aider de la documentation fournie dans le logiciel `RStudio` ou de la recherche `Google`.
2. Vous devez répondre aux questions directement dans ce document et vous assurez qu'il compile lorsque vous utilisez la commande `Knit`. Vous pouvez également compiler vos commandes au fur et à mesure dans ce document en appuyant sur la **flèche verte pointant vers la droite** en haut à droite de votre code `R`.

Installer `R` et `RStudio`

Vous pouvez télécharger `R` aux adresses suivantes:

- Pour Linux
- Pour (Mac) OS X
- Pour Windows

Une fois le logiciel `R` installé, vous pouvez télécharger et installer le logiciel `RStudio` à l'adresse suivante:

- Pour Linux, (Mac) OS X et Windows

Les lois de probabilités

Chaque distribution en `R` possède quatre fonctions qui lui sont associées. Premièrement, la fonction possède un *nom racine* (qui correspond au nom de la `loi`), par exemple le *nom racine* pour la distribution *binomiale* est `binom`. Cette racine est précédée par une de ces quatre lettre:

- `p` pour *probabilité*, qui représente la fonction de répartition
- `q` pour *quantile*, l'inverse de la fonction de répartition
- `d` pour *densité*, la fonction de densité de la distribution
- `r` pour *random* ou *simulation*, une variable aléatoire suivant la distribution spécifiée.

Pour la loi binomiale (*nom racine* `binom`) par exemple, ces fonctions sont `pbinom`, `qbinom`, `dbinom` et `rbinom`.

Nous avons donc:

Loi: `loi`	Densité	Fonction de répartition	Quantile	Simulation
Notations	$f(x)$ ou $P(X = x)$	$F(x)$	valeur liée à $F(x)$	$x_1, x_2, \ldots, x_n$
Commandes	`dloi`	`ploi`	`qloi`	`rloi`

Les noms de lois les plus célèbres sont : `norm` (pour la loi normale), `norm` (pour la loi binomiale), `unif` (pour la loi uniforme), `geom` (pour la loi géométrique), `pois` (pour la loi de Poisson), `t` (pour la loi de Student), `chisq` (pour la loi du Chi-deux), `exp` (pour la loi exponentielle), `f` (pour la loi de Fisher)...

Commandes

Si la loi de X dépend d'un ou de plusieurs paramètres, disons `par1` et `par2`, alors la densité de X en x est donnée par la commande : `dloi(x, par1, par2)`

Quelques exemples sont décrits ci-dessous:

Loi	Binomiale	Géométrique	Poisson
Paramètres	$n \in \mathbb{N}$, $p \in]0,1[$	$p \in]0,1[$	$\lambda > 0$
$X \sim$	$B(n;p)$	$G(p)$	$Po(\lambda)$
Ch(X)	$\{0,1,\dots,n\}$	$\mathbb{N}$	$\mathbb{N}$
$P(X=x)$	$C_x^n p^x q^{n-x}$	$p(1-p)^x$	$e^{-\lambda}\frac{\lambda^x}{x!}$
Commandes	`dbinom(x,n,p)`	`dgeom(x,p)`	`dpois(x,lambda)`

Loi	Uniforme	Exponentielle	Normale
Paramètres	$(a,b) \in \mathbb{R}^2$	$p \in]0,1[$	$\lambda > 0$
$X \sim$	$U([a,b])$	$E(\lambda)$	$N(\mu,\sigma^2)$
Ch(X)	$[a,b]$	$[0,\infty]$	$\mathbb{R}$
$P(X=x)$	$\frac{1}{b-a}$	$\lambda^{-\lambda x}$	$\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$
Commandes	`dunif(x,a,b)`	`dexp(x,lambda)`	`dnorm(x,mu,sigma)`

Exemples de calculs

Soit X une variable aléatoire telle que $X \sim B(8, 0.3)$.

1. Pour calculer $P(X = 4)$, nous devons utiliser la commande suivante:

```
dbinom(4,8,0.3)
```

```
## [1] 0.1361367
```

Ceci signifie que $P(X = 4) = 0.1361367$.

2. Pour calculer $P(X \leq 4)$, nous devons utiliser la commande suivante:

```
pbinom(4,8,0.3)
```

```
## [1] 0.9420324
```

Ceci signifie que $P(X \leq 4) = 0.9420324$.

3. Pour calculer $P(X > 4)$, nous pouvons utiliser une des commandes suivantes:

```
pbinom(4,8,0.3,lower.tail = FALSE)
```

```
## [1] 0.05796765
```

```
1-pbinom(4,8,0.3)
```

```
## [1] 0.05796765
```

Ceci signifie que $P(X > 4) = 0.0579676$.

4. Pour calculer $P(X \geq 4) = 1 - P(X \leq 3)$, nous pouvons utiliser la commande suivante:

```r
1-pbinom(3,8,0.3)
```

```
## [1] 0.1941043
```

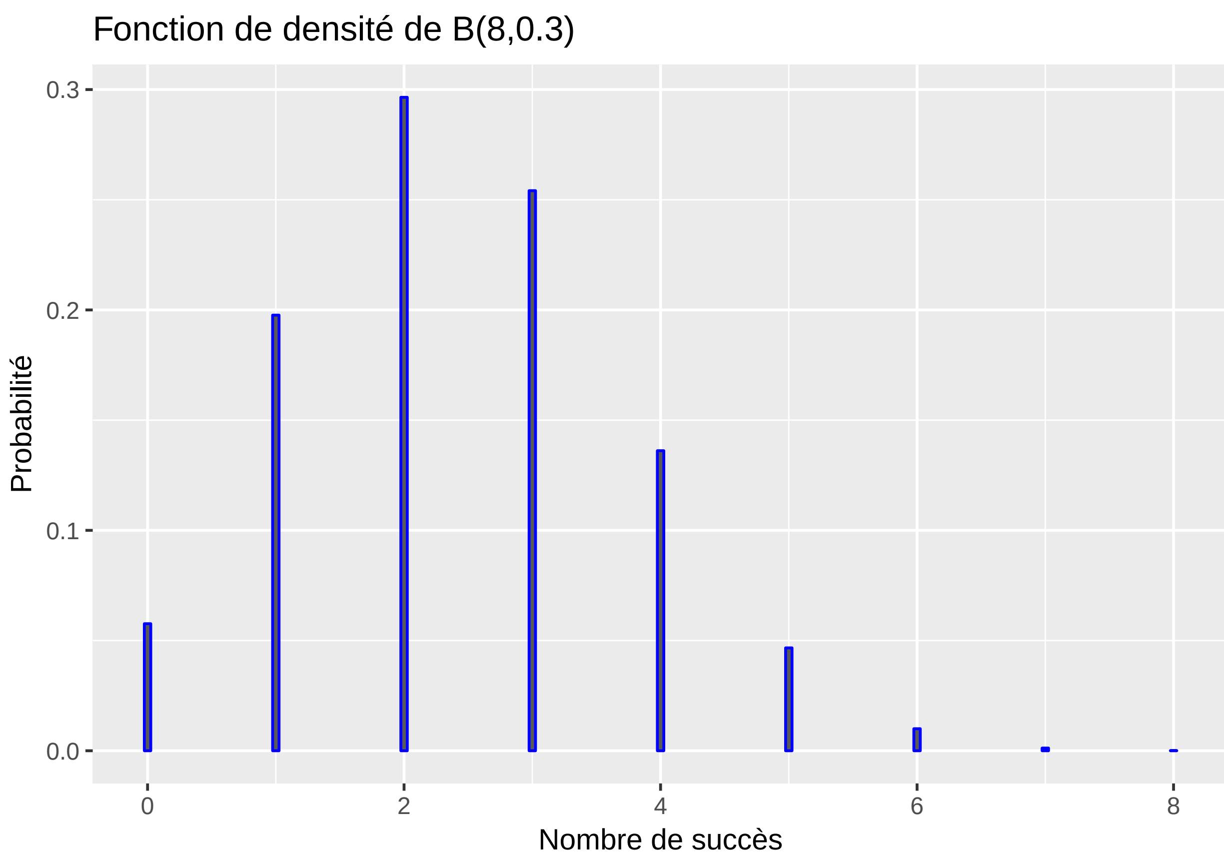
Ceci signifie que $P(X \geq 4) = 0.1941043$.

Représentation graphique

Les lois de probabilités discrètes

Nous pouvons représenter graphiquement la loi binomiale. Soit X $B(8, 0.3)$. Nous aurons:

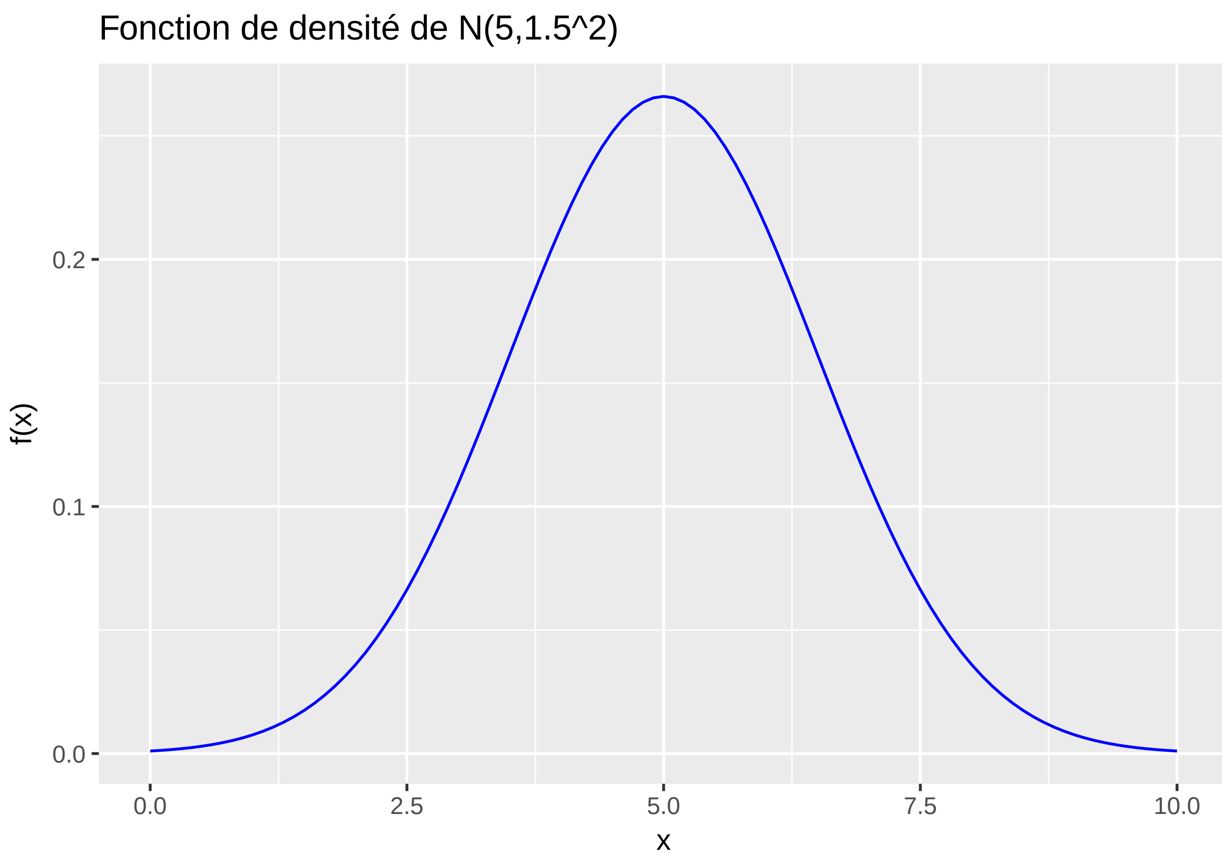
```r
n <- 8
p <- 0.3
fbinom <- data.frame(x = 0:n, y = dbinom(0:n, n, p))
ggplot(fbinom, aes(x = x, y = y)) +
  geom_bar(width = 0.05, stat = "identity", colour = "blue") +
  labs(
    x = "Nombre de succès",
    y = "Probabilité",
    title = "Fonction de densité de B(8,0.3)"
  )
```



Les lois de probabilités continues

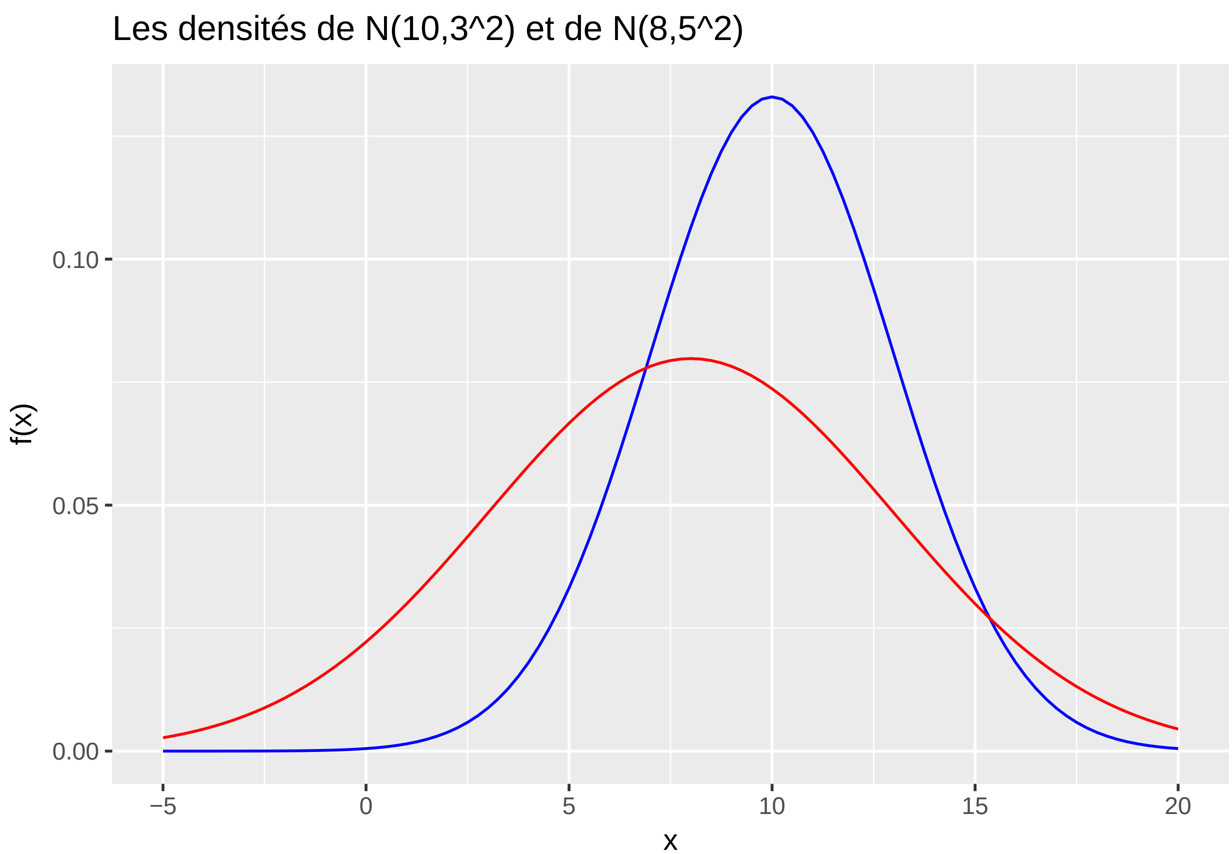
Nous pouvons représenter graphiquement la loi normale. Soit $X \sim N(5, 1.5^2)$. Nous aurons:

```
ggplot(data = data.frame(x = c(0, 10)), aes(x)) +
  stat_function(fun = dnorm, args = list(mean = 5, sd = 1.5), colour = "blue") +
  labs(
    x = "x",
    y = "f(x)",
    title = "Fonction de densité de N(5,1.5^2)"
  )
```



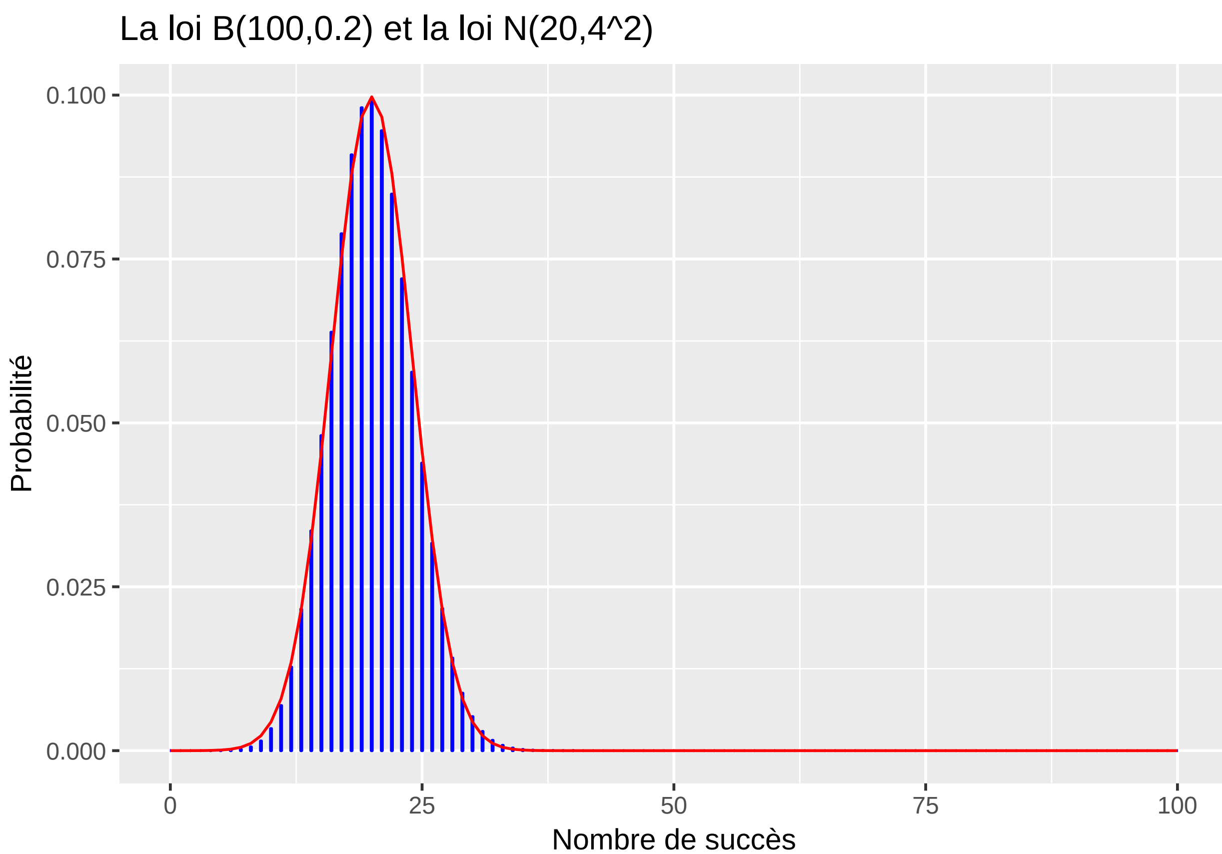
Nous pouvons également superposer plusieurs fonctions de densité. Par exemple, nous allons représenter la loi $N(10, 3^2)$ et la loi $N(8, 5^2)$ sur le même graphique.

```
ggplot(data = data.frame(x = c(-5, 20)), aes(x)) +
  stat_function(fun = dnorm, args = list(mean = 10, sd = 3), colour = "blue") +
  stat_function(fun = dnorm, args = list(mean = 8, sd = 5), colour = "red") +
  labs(
    x = "x",
    y = "f(x)",
    title = "Les densités de N(10,3^2) et de N(8,5^2)"
  )
```



Nous pouvons aussi superposer une variable aléatoire discrète et une variable aléatoire continue. Dans l'exemple suivant, nous avons la loi $B(100, 0.2)$ et son approximation par la loi normale $N(20, 4^2)$.

```
n <- 100
p <- 0.2
m <- n*p
s <- sqrt(n*p*(1-p))
fbinom <- data.frame(x = 0:n, y = dbinom(0:n, n, p))
ggplot(fbinom, aes(x = x, y = y)) +
  geom_bar(width = 0.1, stat = "identity", colour = "blue") +
  stat_function(fun = dnorm, args = list(mean = m, sd = s), colour = "red") +
  labs(
    x = "Nombre de succès",
    y = "Probabilité",
    title = "La loi B(100,0.2) et la loi N(20,4^2)"
  )
```



Exercices

Vous devez répondre aux questions suivantes dans les espaces prévus à cette fin.

1. Soit $X \sim B(15, 0.4)$.

   a) Calculez la probabilité $P(X = 4)$.

```
# Écrivez votre réponse ici.
dbinom(4,15,0.4)
```

```
## [1] 0.1267758
```

   b) Calculez la probabilité $P(X \leq 4)$.

```
# Écrivez votre réponse ici.
pbinom(4,15,0.4)
```

```
## [1] 0.2172777
```

   c) Calculez la probabilité $P(X > 8)$.

```
# Écrivez votre réponse ici.
1-pbinom(8,15,0.4)
```

```
## [1] 0.09504741
```

   d) Calculez la probabilité $P(X \geq 8)$.

```r
# Écrivez votre réponse ici.
1-pbinom(7,15,0.4)
```

```
## [1] 0.2131032
```

2. Soit $X \sim N(0, 1^2)$.

   a) Calculez la probabilité $P(X < -0.5)$.

```r
# Écrivez votre réponse ici.
pnorm(-0.5,0,1)
```

```
## [1] 0.3085375
```

   b) Calculez la probabilité $P(X > 1.5)$.

```r
# Écrivez votre réponse ici.
1-pnorm(1.5,0,1)
```

```
## [1] 0.0668072
```

3. Soit $X \sim N(15, 3^2)$.

   a) Calculez la probabilité $P(16 \leq X \leq 20)$.

```r
# Écrivez votre réponse ici.
pnorm(20,15,3)-pnorm(16,15,3)
```

```
## [1] 0.321651
```

   b) Calculez la probabilité $P(X > 18)$.

```r
# Écrivez votre réponse ici.
1-pnorm(18,15,3)
```

```
## [1] 0.1586553
```

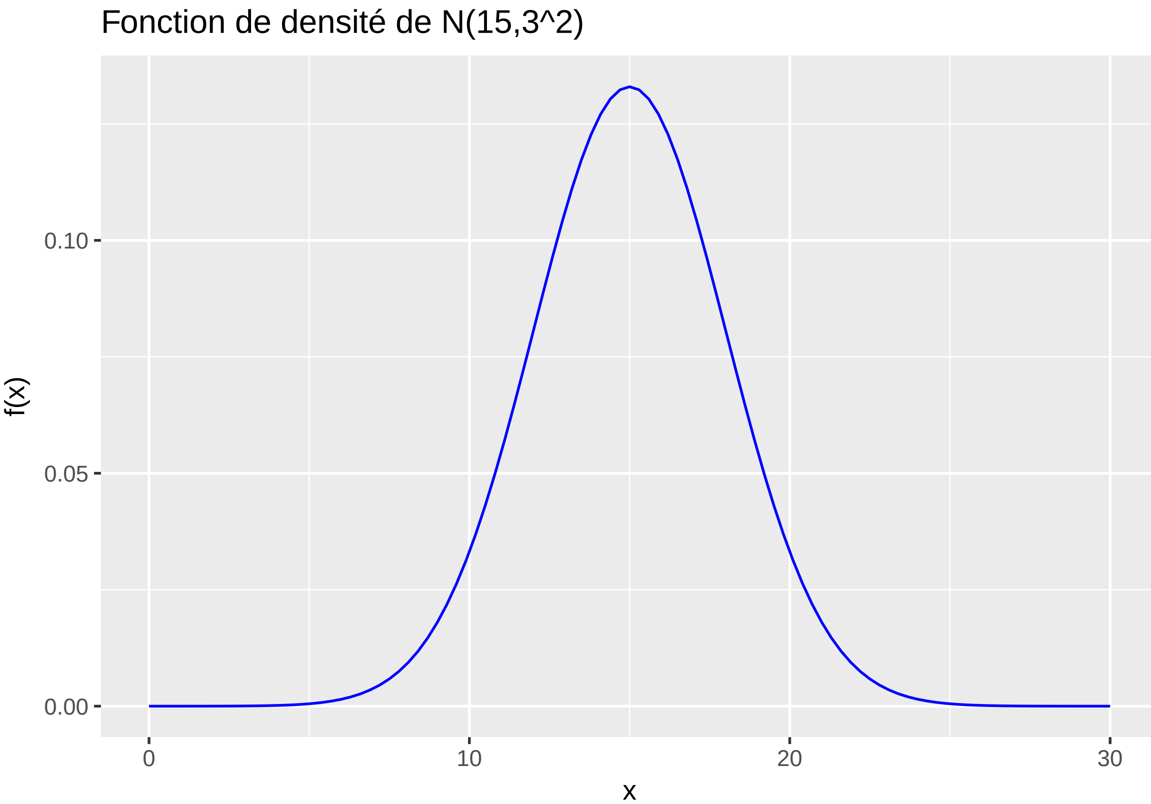
   c) Calculez la probabilité $P(X < 6)$.

```r
# Écrivez votre réponse ici.
pnorm(6,15,3)
```

```
## [1] 0.001349898
```

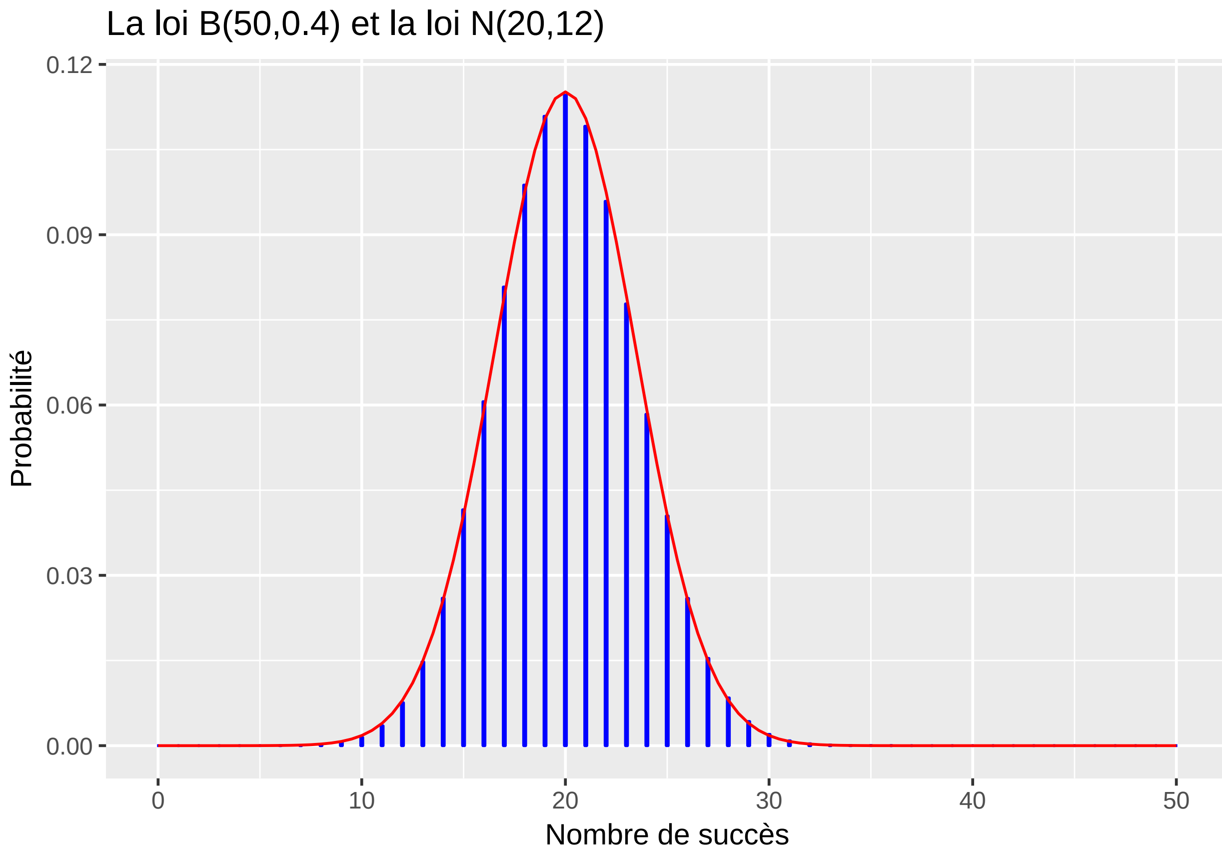
   d) Tracez la fonction de densité de la variable continue X.

```r
# Écrivez votre réponse ici.
ggplot(data = data.frame(x = c(0, 30)), aes(x)) +
stat_function(fun = dnorm, args = list(mean = 15, sd = 3), colour = "blue") +
labs(
  x = "x",
  y = "f(x)",
  title = "Fonction de densité de N(15,3^2)"
)
```



4. Représentez le graphe de la densité d'une variable $X \sim B(50, 0.4)$, puis ajoutez par dessus ce graphe celui de la densité d'une variable $Y \sim N(20, 12)$ (cela illustrera le fait que, lorsque $n > 30$, $np > 5$ et $n(1-p) > 5$, on peut approximer la loi binomiale $B(n,p)$ par la loi normale $N(np, np(1-p))$.

```r
# Écrivez votre réponse ici.
n <- 50
p <- 0.4
m <- n*p
s <- sqrt(n*p*(1-p))
fbinom <- data.frame(x = 0:n, y = dbinom(0:n, n, p))
ggplot(fbinom, aes(x = x, y = y)) +
  geom_bar(width = 0.1, stat = "identity", colour = "blue") +
  stat_function(fun = dnorm, args = list(mean = m, sd = s), colour = "red") +
  labs(
    x = "Nombre de succès",
    y = "Probabilité",
    title = "La loi B(50,0.4) et la loi N(20,12)"
  )
```


La loi B(50,0.4) et la loi N(20,12)
0.12
0.09
0.06
0.03
0.00
Probabilité
0
10
20
30
40
50
Nombre de succès

5.