

Probabilidad y Distribuciones

Pablo de la Cuesta García

2024-01-08

Contents

1	Conceptos básicos de probabilidad	2
2	Distribuciones de probabilidad	2
2.1	Distribución Uniforme Discreta	2
2.1.1	Ejemplo distribución uniforme discreta	2
2.2	Distribución de Bernoulli	4
2.2.1	Propiedades	4
2.2.2	Función de Probabilidad Puntual	5
2.2.3	Función de Probabilidad Acumulada	5
2.2.4	Ejemplo distribución de Bernoulli	5
2.3	Distribución Binomial	6
2.3.1	Conceptos clave	6
2.3.2	Propiedades	7
2.3.3	Función de probabilidad Puntual	7
2.4	Distribución de Poisson	7
2.4.1	Función de Probabilidad	8
2.4.2	Función de Distribución Acumulada	8
2.4.3	Usos de la distribución de Poisson	8
2.4.4	Ejemplo	9
2.5	Distribución Normal	10
2.5.1	Función de Probabilidad	10
2.5.2	Función de Distribución Acumulada	12
2.5.3	Tipificación de una Variable Aleatoria Normal	13
2.5.4	Cálculo de probabilidades y percentiles	14
2.5.5	Relación entre distribuciones	19
2.5.5.1	Relación entre la distribución Binomial y Poisson	20
2.5.5.2	Relación entre la distribución Binomial y Normal	21

2.5.5.3	Relación entre la distribución de Poisson y Normal	22
2.6	Distribución t de Student	24
2.6.1	Conceptos	24
2.6.2	Propiedades	24
2.6.3	Representación gráfica y comparación con la distribución normal	24
2.6.4	Ejemplos	25
2.6.4.1	Ejemplo 1: Prueba de Hipótesis para la media de una muestra.	25
3	Teoremas y propiedades	25

1 Conceptos básicos de probabilidad

2 Distribuciones de probabilidad

2.1 Distribución Uniforme Discreta

Una distribución uniforme discreta es aquella en la que todos los valores posibles de la variable aleatoria tienen la misma probabilidad de ocurrir. Por ejemplo, si lanzamos un dado, la probabilidad de obtener cualquier número entre 1 y 6 es la misma.

Hay dos tipos de distribuciones uniformes: discretas y continuas.

- **La distribución uniforme discreta** surge en espacios de probabilidad equiprobables, es decir, en situaciones donde de n resultados diferentes, todos tienen la misma probabilidad de ocurrir.
- **La distribución uniforme continua** describe un experimento en el que los resultados de la variable aleatoria tienen las mismas probabilidades de ocurrir dentro de un intervalo $[a, b]$.

Tipo	Función de probabilidad	Esperanza	Varianza
Discreta	$p_x(x) = \frac{1}{n}$	$\mu = \sum_i \frac{x_i}{n}$	$\sigma^2 = \sum_i \frac{(x_i^2)}{n^2} - \mu^2$
Continua	$f(x) = \frac{1}{b-a}$	$\mu = \frac{a+b}{2}$	$\sigma^2 = \frac{(b-a)^2}{12}$

2.1.1 Ejemplo distribución uniforme discreta

Ejemplo del lanzamiento de un dado:

- Función de probabilidad: $p_x(x) = \frac{1}{6}$
- Función de distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < 1 \\ \frac{1}{6} & \text{si } 1 \leq x < 2 \\ \frac{2}{6} & \text{si } 2 \leq x < 3 \\ \frac{3}{6} & \text{si } 3 \leq x < 4 \\ \frac{4}{6} & \text{si } 4 \leq x < 5 \\ \frac{5}{6} & \text{si } 5 \leq x < 6 \\ 1 & \text{si } x \geq 6 \end{cases}$$

$$\mu = E(X) = 3.5\sigma^2 = 35/12 = 2.92$$

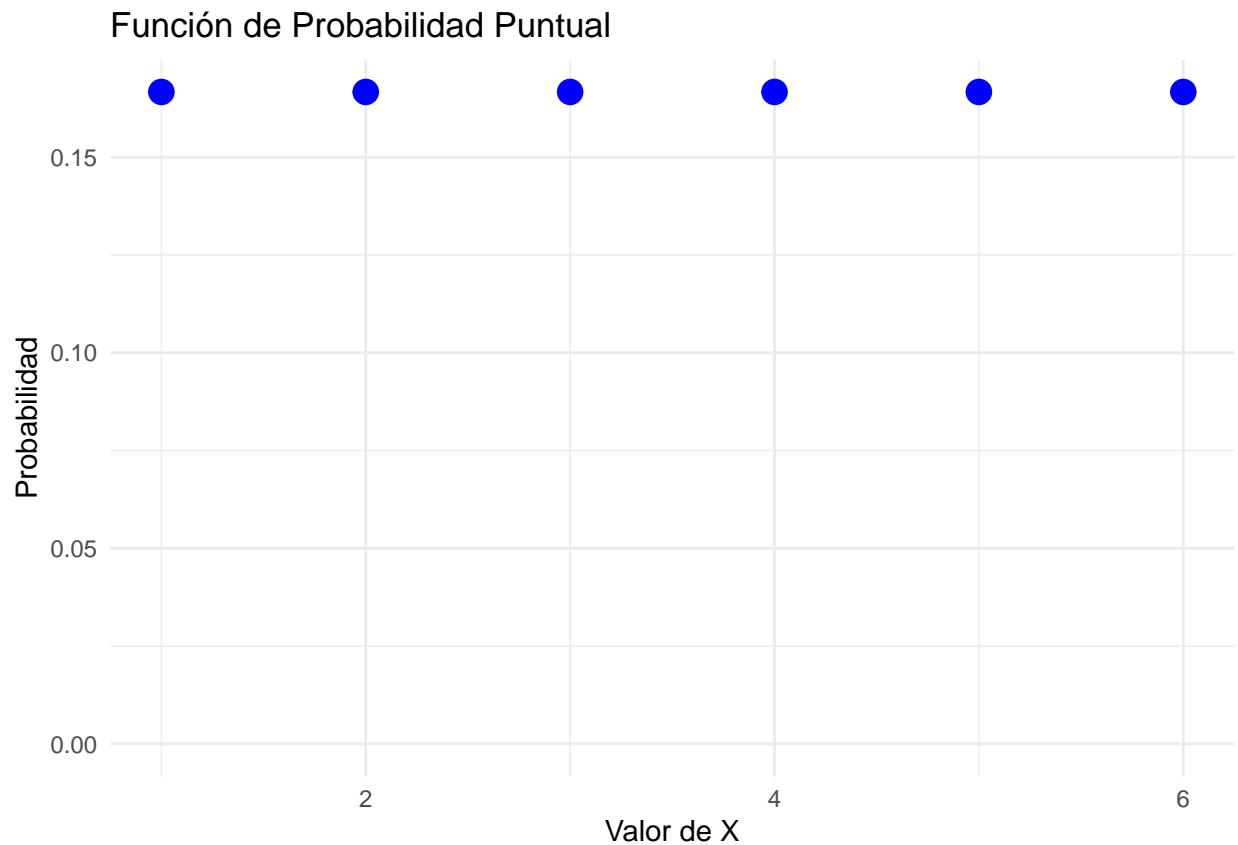
Función de probabilidad puntual:

```
library(ggplot2)

# Tus datos
x <- 1:6
p_x <- rep(1/6, times = 6)

# Crear un dataframe para ggplot
df_puntual <- data.frame(x, p_x)

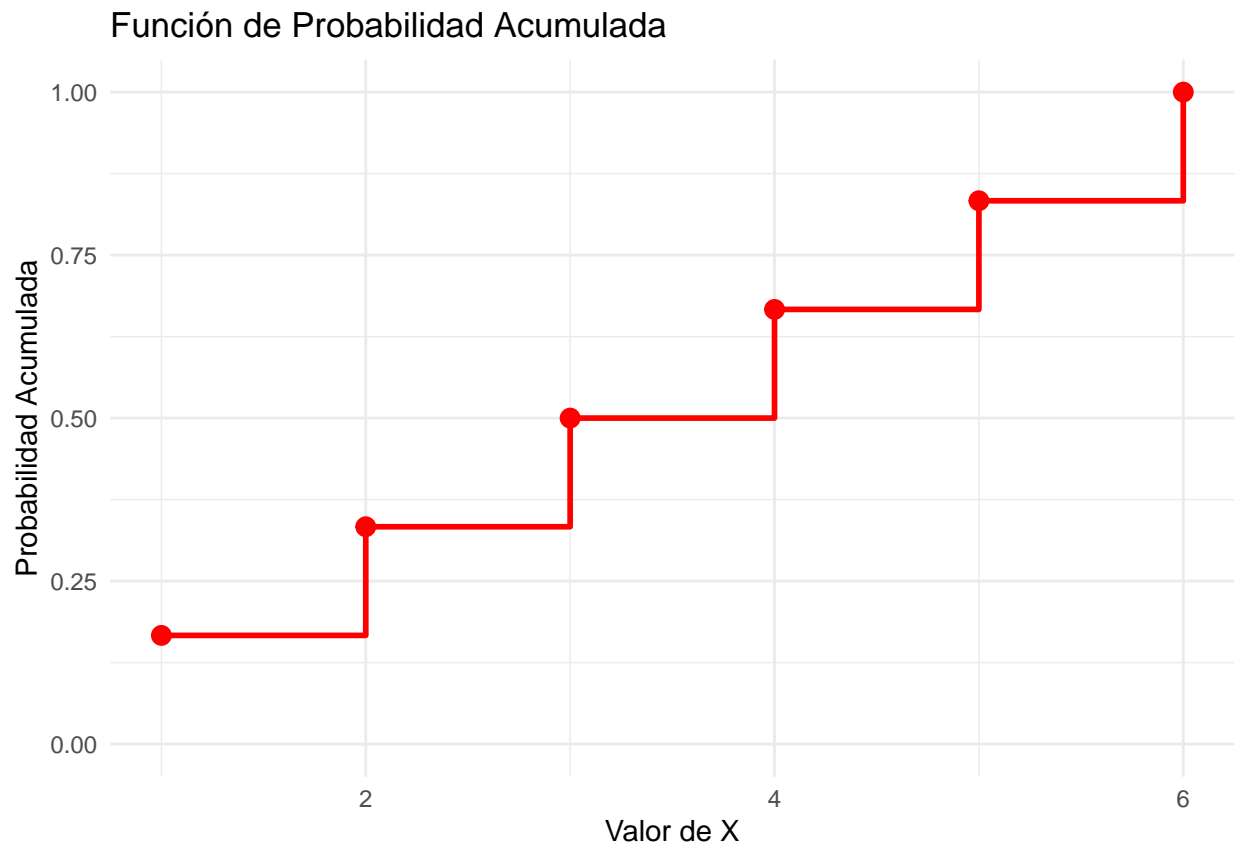
# Gráfica de la función de probabilidad puntual
ggplot(df_puntual, aes(x = x, y = p_x)) +
  geom_point(size = 4, color = "blue") +
  theme_minimal() +
  labs(title = "Función de Probabilidad Puntual",
       x = "Valor de X",
       y = "Probabilidad") +
  ylim(0, 1/6)
```



```
# Calcular la función de probabilidad acumulada
p_acumulada <- cumsum(p_x)
```

```
# Crear un dataframe para ggplot
df_acumulada <- data.frame(x, p_acumulada)

# Gráfica de la función de probabilidad acumulada
ggplot(df_acumulada, aes(x = x, y = p_acumulada)) +
  geom_step(direction = "hv", linewidth = 1, color = "red") +
  geom_point(size = 3, color = "red") +
  theme_minimal() +
  labs(title = "Función de Probabilidad Acumulada",
       x = "Valor de X",
       y = "Probabilidad Acumulada") +
  ylim(0, 1)
```



2.2 Distribución de Bernoulli

Supongamos que realizamos el experimento de tirar una moneda equilibrada y observamos que salió cara. Este tipo de experimentos donde hay dos posibles resultados se denominan experimentos de Bernoulli.

2.2.1 Propiedades

- **Media:** $\mu = p$
- **Varianza:** $\sigma^2 = p(1 - p)$

2.2.2 Función de Probabilidad Puntual

La función de probabilidad puntual o f.p.p. se puede denotar como $p_x(x)$, se define para todo x de la siguiente manera:

$$p_x(x) = \begin{cases} p & \text{si } x = 1 \\ 1 - p & \text{si } x = 0 \end{cases}$$

Esta fórmula puede resumirse en una expresión más compacta utilizando la notación de potencias:

$$p_x(x) = p^x(1 - p)^{1-x}$$

Siendo:

- $P(X = 1) = p^1 (1 - p)^0 = p$
- $P(X = 0) = p^0 (1 - p)^1 = 1 - p$

2.2.3 Función de Probabilidad Acumulada

La función de probabilidad acumulada o f.p.a. se puede denotar como $F(x)$, se define para todo x de la siguiente manera:

Para obtener la función de distribución acumulada (FDA) a partir de la función de probabilidad puntual de una distribución de Bernoulli, sumamos las probabilidades de los posibles resultados hasta el valor de interés.

La FPP de la distribución de Bernoulli es $p_x(x) = p^x(1 - p)^{1-x}$, entonces la FDA denotada como $F_X(x)$, se define como la suma de las probabilidades de todos los resultados menores o iguales a x . Para la distribución de Bernoulli, esto significa que:

- Cuando $x < 0$, no hay resultados posibles, ya que los únicos resultados posibles son 0 y 1, así que $F_X(x) = 0$.
- Cuando $0 \leq x < 1$, solo hay un resultado posible, $F_X(x) = p^0(1 - p)^{1-0} = 1 - p$.
- Cuando $x \geq 1$, consideremos ambos resultados posibles (0 y 1), sumando sus probabilidades da: $F_X(x) = p^1(1 - p)^{1-1} + p^0(1 - p)^{1-0} = p + 1 - p = 1$.

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1 - p & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x \geq 1 \end{cases}$$

2.2.4 Ejemplo distribución de Bernoulli

```
# Instalar y cargar el paquete ggplot2 si aún no está instalado
if(!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)

# Parámetro p para la distribución de Bernoulli
p = 0.5 # Probabilidad de éxito. Modifica este valor para ver diferentes distribuciones.

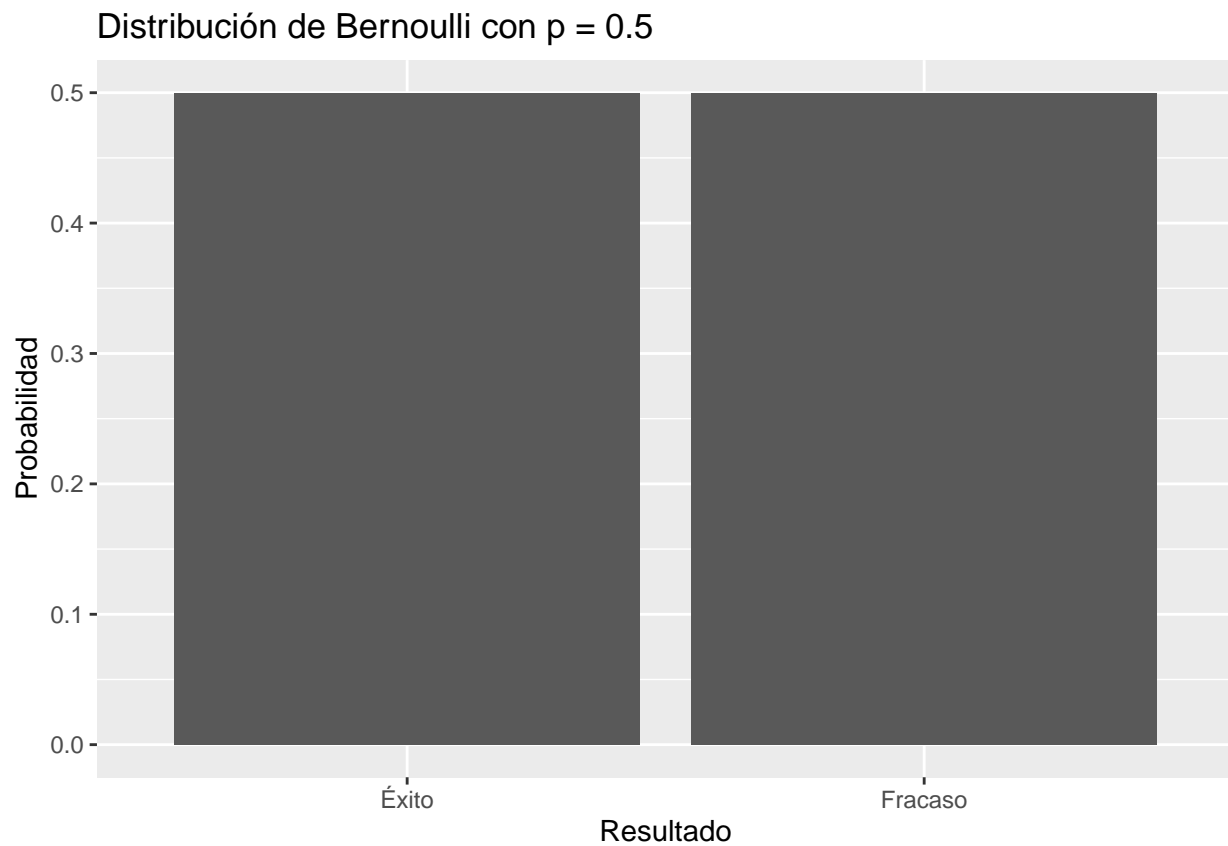
# Crear un dataframe con los resultados y probabilidades
bernoulli_df <- data.frame(
```

```

resultado = c("Fracaso", "Éxito"),
probabilidad = c(1-p, p)
)

# Crear el gráfico de barras
ggplot(bernoulli_df, aes(x = resultado, y = probabilidad)) +
  geom_bar(stat = "identity") +
  ylab("Probabilidad") +
  xlab("Resultado") +
  ggtitle(paste("Distribución de Bernoulli con p =", p))

```



2.3 Distribución Binomial

La distribución binomial es un concepto fundamental en la estadística, utilizado para modelar el número de éxitos en una secuencia de ensayos independientes de Bernoulli con la misma probabilidad de éxito.

2.3.1 Conceptos clave

- **Ensayos de Bernoulli:** Un experimento de Bernoulli es un experimento aleatorio con dos posibles resultados, éxito o fracaso. Por ejemplo, lanzar una moneda equilibrada es un experimento de Bernoulli, ya que solo hay dos posibles resultados: cara o cruz.
- **Ensayos independientes:** Los ensayos de Bernoulli son independientes si el resultado de un ensayo no afecta el resultado de los otros ensayos. Por ejemplo, lanzar una moneda dos veces es un experimento

de Bernoulli con dos ensayos independientes, ya que el resultado de la primera tirada no afecta el resultado de la segunda tirada.

- **Probabilidad de éxito constante:** La probabilidad de éxito debe ser la misma para todos los ensayos.
- **Número de ensayos:** Se denota como n y representa el número total de ensayos.

Es decir, una variable aleatoria tiene una distribución de Bernoulli si solo puede tomar dos valores. Las diferentes variables de Bernoulli tienen que ser independientes. Al sumar estas variables de Bernoulli independientes, estamos contando el número total de éxitos que ocurrieron en todas ellas. La suma de m variables, cada una con la misma probabilidad de éxito p , resulta en una nueva variable aleatoria con distribución binomial con parámetros m y p .

Ejemplo ilustrativo: Supongamos que lanzamos una moneda 5 veces y estamos interesados en contar cuántas veces obtenemos cara (éxito). Cada lanzamiento es un ensayo de Bernoulli independiente donde $p = 0.5$, entonces la suma de estas 5 variables de Bernoulli será una variable aleatoria X con una distribución binomial $Bin(5, 0.5)$

2.3.2 Propiedades

- **Media:** $\mu = np$
- **Varianza:** $\sigma^2 = np(1 - p)$

2.3.3 Función de probabilidad Puntual

También conocida como función de masa de probabilidad, nos da la probabilidad de que una variable aleatoria binomial tome exactamente un valor específico.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Función de Distribución Acumulada

Por otro lado, la función de distribución acumulada es la probabilidad de que una variable aleatoria tome un valor menor o igual a un valor específico.

$$F(k; n, p) = P(X \leq k) = \sum_{i=0}^k \binom{n}{i} p^i (1 - p)^{n-i}$$

2.4 Distribución de Poisson

La distribución de Poisson es una distribución de probabilidad discreta que expresa la probabilidad de que un número dado de eventos ocurra en un intervalo fijo de tiempo o espacio si estos eventos ocurren con una tasa media conocida e independientemente del tiempo desde el último evento.

- **Eventos independientes:** la ocurrencia de un evento no afecta a la probabilidad de otro.
- **Tasa media conocida:** la tasa de ocurrencia es constante.
- **Intervalo fijo:** el intervalo de tiempo o espacio es fijo.
- **Eventos raros:** la distribución de Poisson es adecuada para eventos que ocurren con poca frecuencia.

2.4.1 Función de Probabilidad

La función de probabilidad puntual de la distribución de Poisson se define como:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

donde:

- $P(X = k)$ es la probabilidad de que ocurran k eventos en el intervalo dado.
- λ es el número esperado de ocurrencias en el intervalo dado.
- k es el número de ocurrencias que se desea conocer.

2.4.2 Función de Distribución Acumulada

La función de probabilidad acumulada da la probabilidad de que la variable aleatoria tome un valor menor o igual a un cierto valor. Para la distribución de Poisson, la función de probabilidad acumulada $P(X \leq k)$ se define como:

$$P(X \leq k) = \sum_{i=0}^k \frac{\lambda^i e^{-\lambda}}{i!}$$

```
# Definir el parámetro lambda
lambda <- 2

# Calcular la probabilidad puntual de tener exactamente 3 errores
probabilidad_puntual <- dpois(3, lambda)

# Calcular la probabilidad acumulada de tener 3 errores o menos
probabilidad_acumulada <- ppois(3, lambda)

# Imprimir los resultados
print(paste("Probabilidad puntual:", probabilidad_puntual))
```

```
## [1] "Probabilidad puntual: 0.180447044315484"
```

```
print(paste("Probabilidad acumulada:", probabilidad_acumulada))
```

```
## [1] "Probabilidad acumulada: 0.857123460498547"
```

2.4.3 Usos de la distribución de Poisson

- Número de clientes que llegan a un supermercado en una hora.
- Número de llamadas que recibe un centro de atención telefónica en una hora.
- Accesos a un servidor web de una empresa en un día.
- Número de defectos en un rollo de tela.

2.4.4 Ejemplo

Supongamos que un barco recibe en promedio 5 clientes por hora. ¿Cuál es la probabilidad de que lleguen 10 clientes en una hora?

Aquí, $k = 10$ y $\lambda = 5$. Entonces, la probabilidad de que lleguen 10 clientes en una hora es:

$$P(X = 10) = \frac{5^{10}e^{-5}}{10!} = 0.018$$

Esto significa que hay aproximadamente un 2% de probabilidad de que lleguen 10 clientes en una hora.

```
# Definir el número promedio de eventos (lambda)
lambda <- 5

# Calcular la probabilidad de recibir exactamente 3 clientes
probabilidad <- dpois(10, lambda)

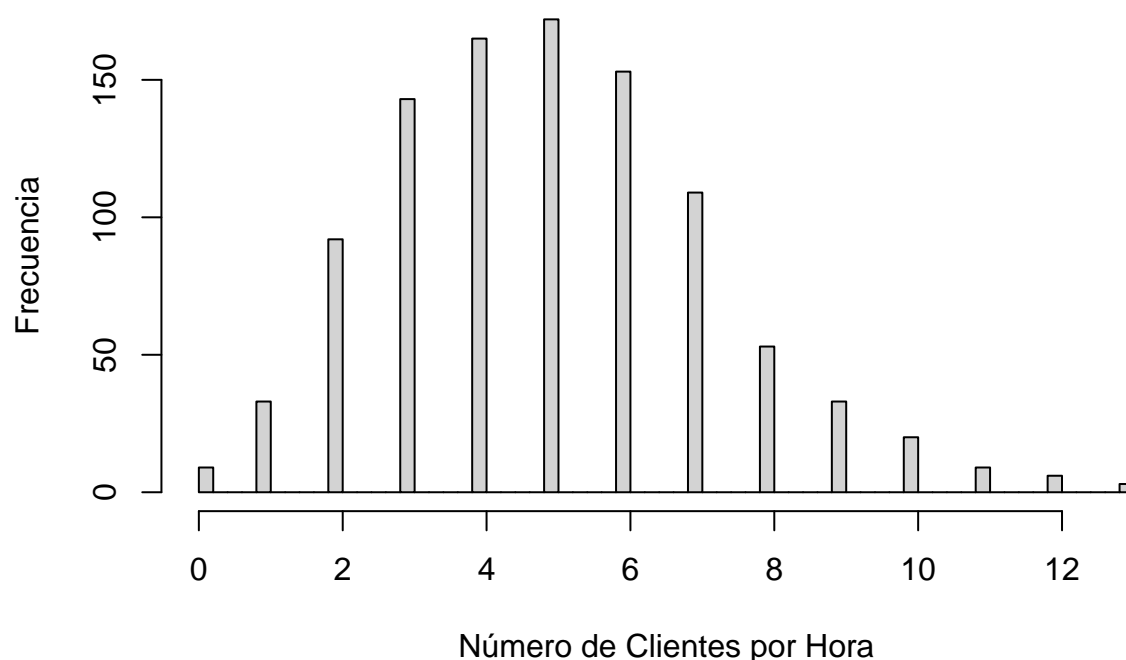
# Mostrar la probabilidad
print(probabilidad)
```

```
## [1] 0.01813279
```

```
# Generar una muestra de 1000 eventos de Poisson
muestra <- rpois(1000, lambda)

# Visualizar la distribución
hist(muestra, breaks = 50, main = "Distribución de Poisson de Eventos en un Banco", xlab = "Número de C
```

Distribución de Poisson de Eventos en un Banco



2.5 Distribución Normal

La distribución normal es una de las distribuciones de probabilidad más importantes en estadística. Se utiliza para modelar muchos fenómenos naturales, como la altura o el peso de una población, la temperatura en un día determinado, los errores de medición, etc. Es decir, variables aleatorias continuas. Se caracteriza por su forma de campana simétrica y se describe completamente por dos parámetros: la media (μ) y la desviación estándar (σ). Estos conceptos clave tienen las siguientes implicaciones:

- **Media (μ):** La media es el centro de la distribución. La media de una distribución normal es igual a su mediana y su moda. La media determina la ubicación horizontal de la distribución.
- **Desviación estándar (σ):** La desviación estándar es una medida de dispersión. La desviación estándar determina la altura de la distribución.
- **Simetría:** La distribución normal es simétrica alrededor de su media. Esto significa que la probabilidad de obtener un valor por encima de la media es la misma que la probabilidad de obtener un valor por debajo de la media.
- **Área bajo la curva:** El área total bajo la curva de una distribución normal es igual a 1. Esto significa que la probabilidad de obtener cualquier valor de la distribución es igual a 1. El 68% de los valores se encuentran dentro de una desviación estándar de la media, el 95% dentro de dos desviaciones estándar y el 99.7% dentro de tres desviaciones estándar. Esto es conocido como la regla 68-95-99.7.

2.5.1 Función de Probabilidad

La función de probabilidad puntual de la distribución normal se define como:

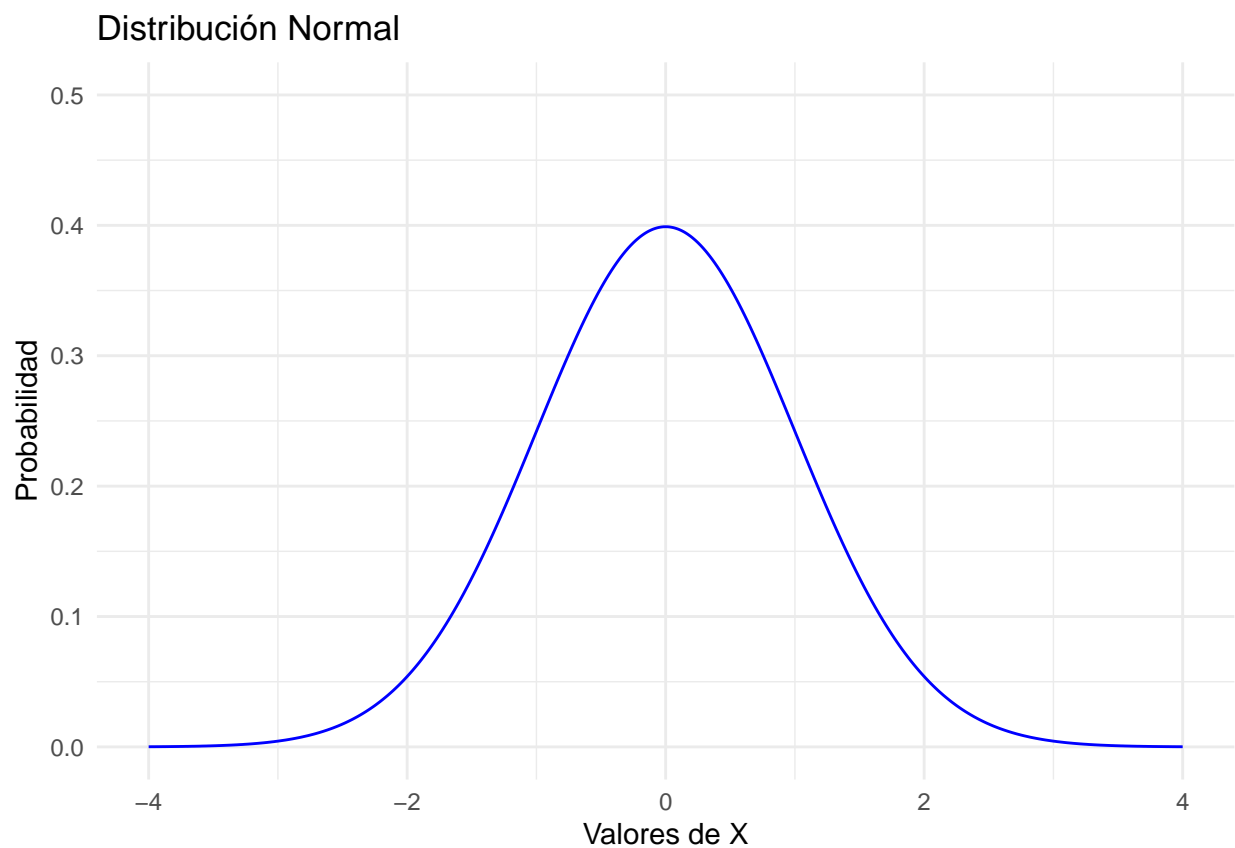
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

```
# Instalar y cargar el paquete ggplot2 si aún no está instalado
if(!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)

# Definir la media y la desviación estándar
mu <- 0
sigma <- 1

# Crear un dataframe con los valores de x y y
x <- seq(-4, 4, 0.01)
y <- dnorm(x, mu, sigma)
df <- data.frame(x, y)

# Crear el gráfico de la distribución normal
ggplot(df, aes(x = x, y = y)) +
  geom_line(color = "blue") +
  theme_minimal() +
  labs(title = "Distribución Normal",
       x = "Valores de X",
       y = "Probabilidad") +
  ylim(0, 0.5)
```



2.5.2 Función de Distribución Acumulada

La función de probabilidad acumulada de la distribución normal se define como:

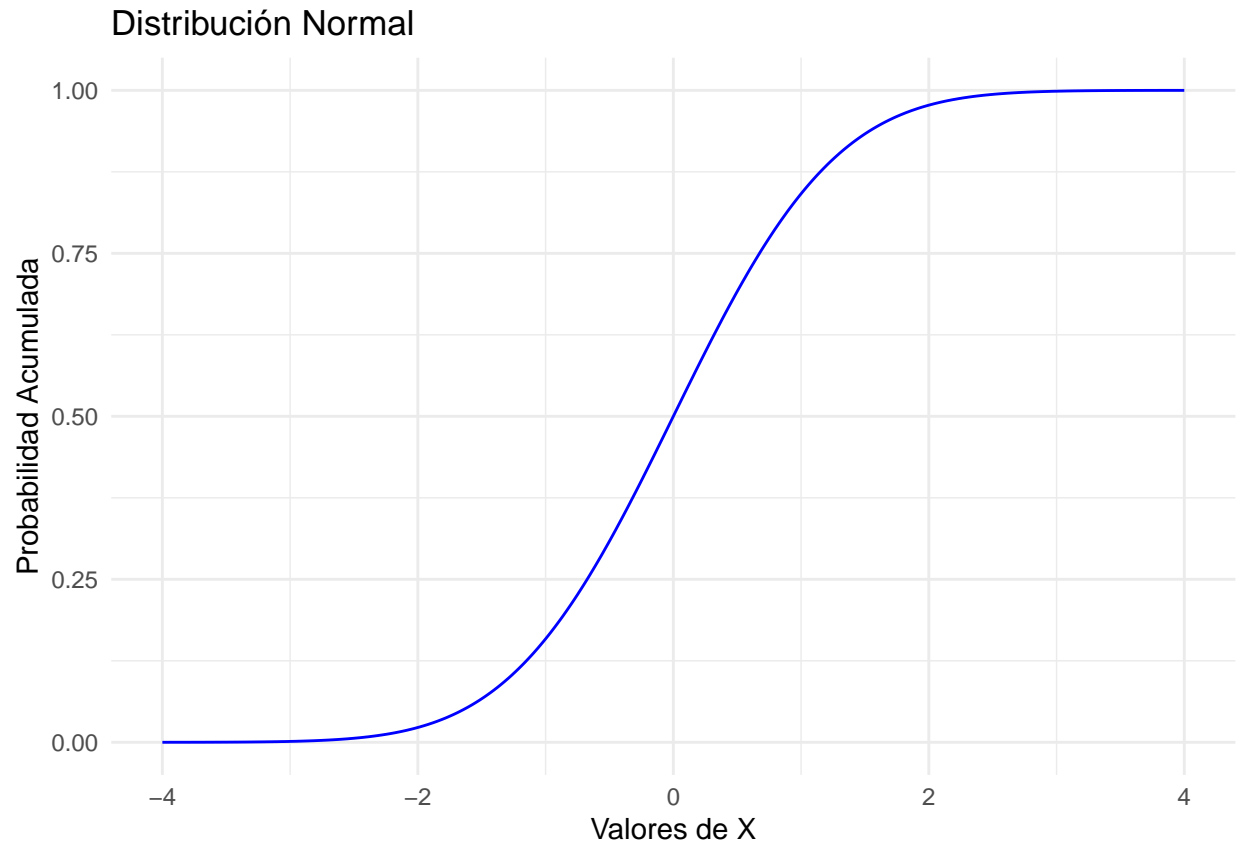
$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

```
# Instalar y cargar el paquete ggplot2 si aún no está instalado
if(!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)

# Definir la media y la desviación estándar
mu <- 0
sigma <- 1

# Crear un dataframe con los valores de x y y
x <- seq(-4, 4, 0.01)
y <- pnorm(x, mu, sigma)
df <- data.frame(x, y)

# Crear el gráfico de la distribución normal
ggplot(df, aes(x = x, y = y)) +
  geom_line(color = "blue") +
  theme_minimal() +
  labs(title = "Distribución Normal",
       x = "Valores de X",
       y = "Probabilidad Acumulada") +
  ylim(0, 1)
```



2.5.3 Tipificación de una Variable Aleatoria Normal

La tipificación de una variable aleatoria normal es un proceso que transforma una variable aleatoria normal X con media μ y desviación estándar σ en una nueva variable aleatoria Z con media 0 y desviación estándar 1. Esto se logra mediante el siguiente cambio de variable:

$$Z = \frac{X - \mu}{\sigma}$$

```
# Instalar y cargar el paquete ggplot2 si aún no está instalado
if(!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)

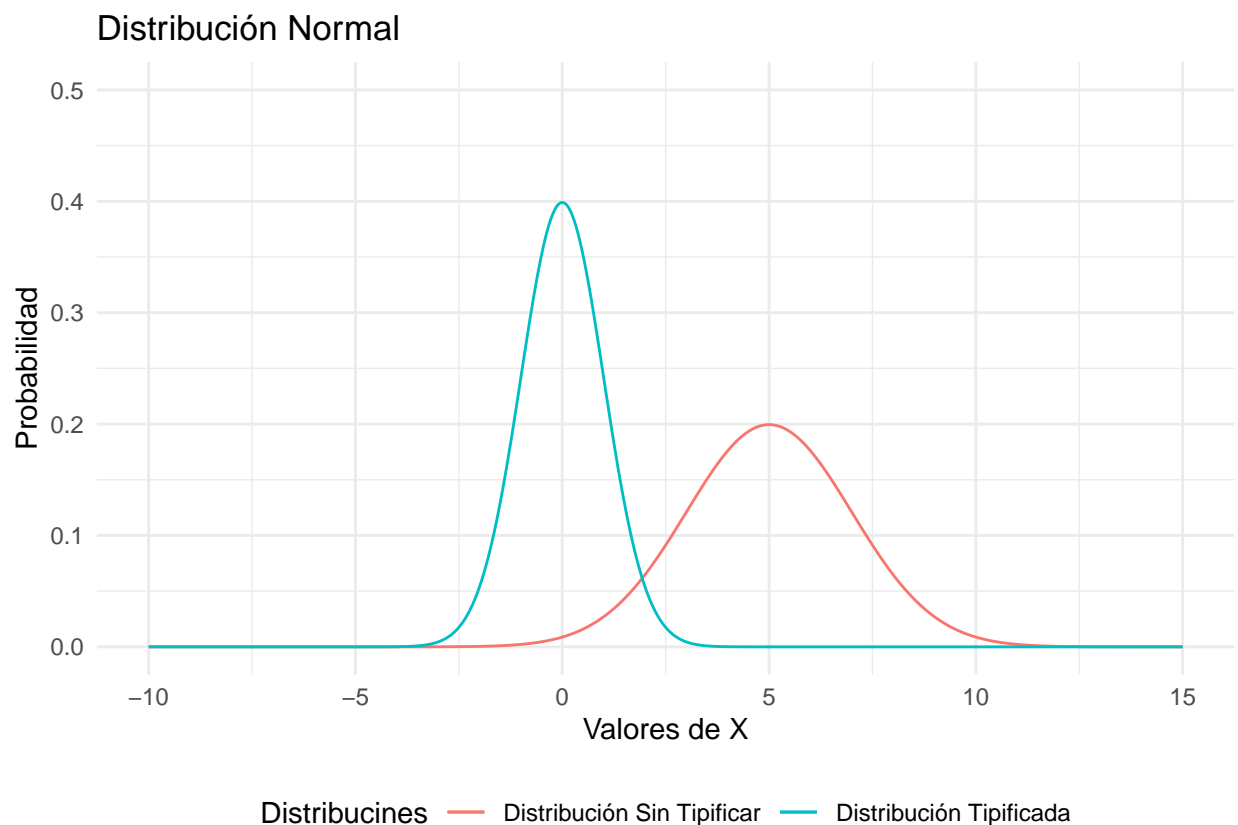
# Definir la media y la desviación estándar sin tipificar
mu_1 <- 5
sigma_1 <- 2

# Tipificamos la variable aleatoria
mu_2 <- 0
sigma_2 <- 1

# Crear un dataframe con los valores de x y y
x <- seq(-10, 15, 0.01)
y_1 <- dnorm(x, mu_1, sigma_1)
y_2 <- dnorm(x, mu_2, sigma_2)
```

```
df <- data.frame(x, y_1, y_2)

# Crear el gráfico de la distribución normal
ggplot(df, aes(x = x)) +
  geom_line(aes(y = y_1, colour = "Distribución Sin Tipificar")) +
  geom_line(aes(y = y_2, colour = "Distribución Tipificada")) +
  theme_minimal() +
  labs(title = "Distribución Normal",
       x = "Valores de X",
       y = "Probabilidad",
       colour = "Distribuciones") +
  ylim(0, 0.5) +
  theme(legend.position = "bottom")
```



2.5.4 Cálculo de probabilidades y percentiles

Supongamos que los resultados de un examen de matemáticas siguen una distribución normal con media 75 y desviación estándar 10. ¿Cuál es la probabilidad de que un estudiante obtenga la nota mínima de 50 puntos?

1. Proporción de exámenes con puntajes por encima de 90
2. Proporción de exámenes con puntajes entre 85 y 98
3. Supongamos que el profesor determina que el 20% de los estudiantes con menor nota deben ver a un tutor para ayuda extra. ¿Cuál es la nota mínima que un estudiante debe obtener para no tener que ver a un tutor?

```

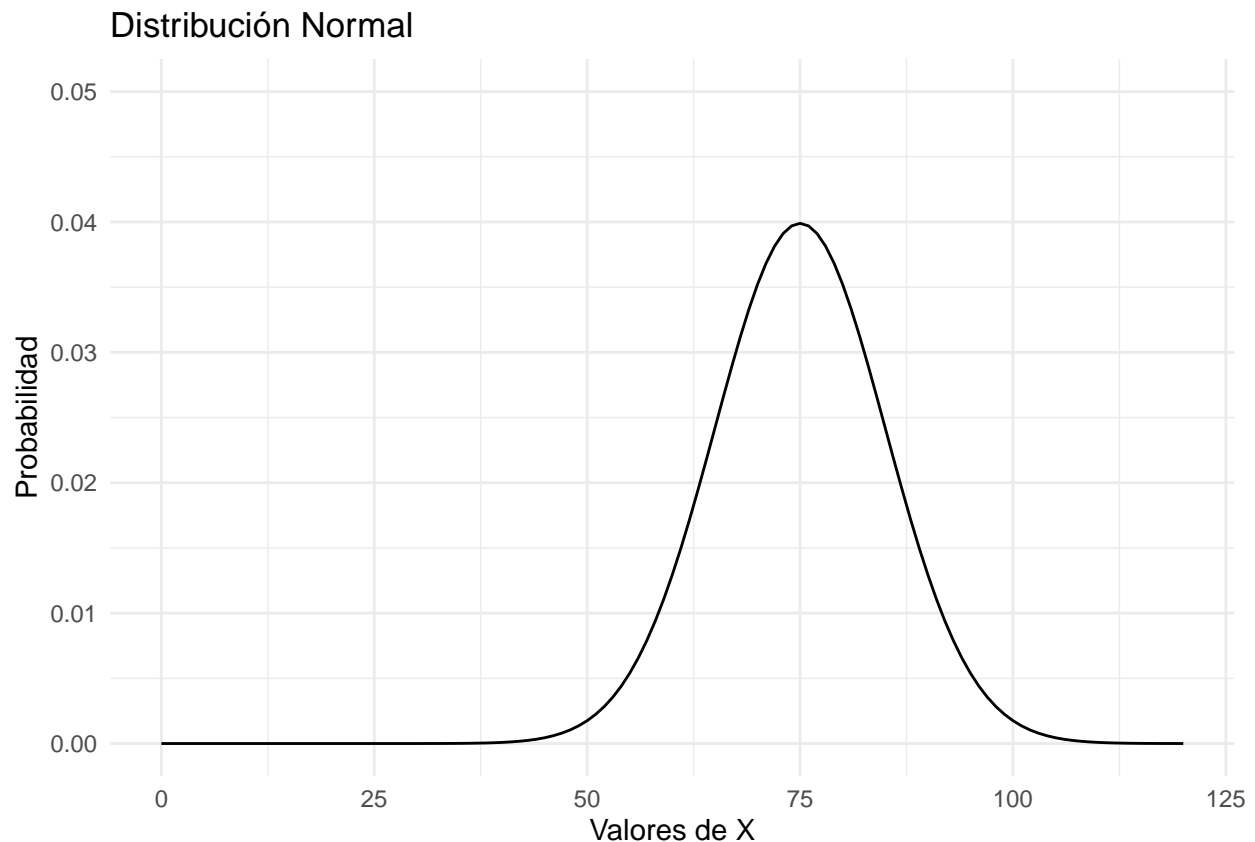
# Instalar y cargar el paquete ggplot2 si aún no está instalado
if(!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)

# Definir la media y la desviación estándar
mu <- 75
sigma <- 10

# Crear un dataframe con los valores de x e y
x <- seq(0, 120, 1)
y <- dnorm(x, mu, sigma)
df <- data.frame(x, y)

# Crear el gráfico de la distribución normal
ggplot(df, aes(x = x, y = y)) +
  geom_line() +
  theme_minimal() +
  labs(title = "Distribución Normal",
       x = "Valores de X",
       y = "Probabilidad") +
  ylim(0, 0.05)

```



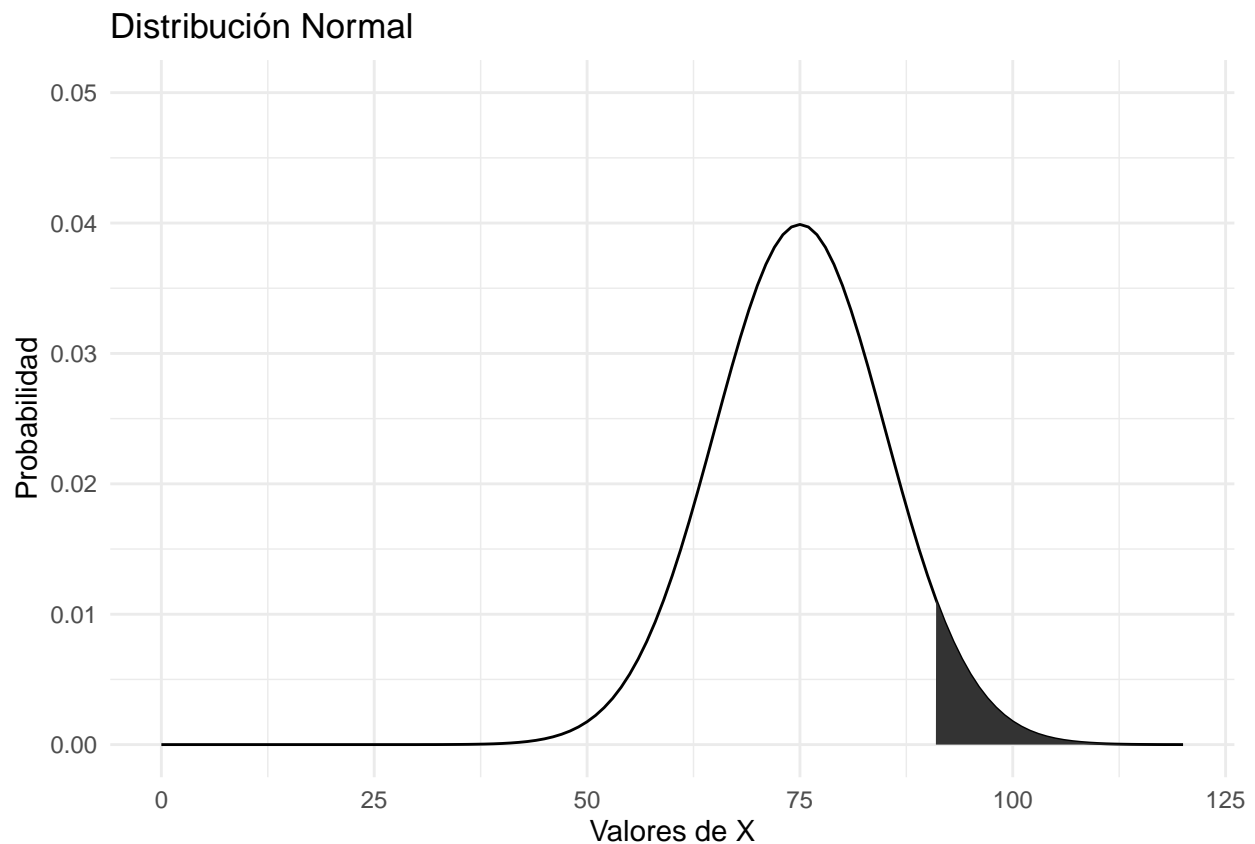
a) Proporción de exámenes con puntajes por encima de 90

Debemos calcular $P(X > 90)$, que es la probabilidad de que un estudiante obtenga una nota mayor a 90. Esto es equivalente a calcular $P(X \leq 90)$ y restar el resultado de 1.

```
# Calcular la probabilidad acumulada de obtener una nota menor o igual a 90
prob_90 <- pnorm(90, mu, sigma)
print(paste("Probabilidad de obtener una nota menor o igual a 90:", prob_90))
```

```
## [1] "Probabilidad de obtener una nota menor o igual a 90: 0.933192798731142"
```

```
# visualización del área bajo la curva para  $P(X > 90)$ 
ggplot(df, aes(x = x, y = y)) +
  geom_line() +
  geom_area(data = subset(df, x > 90)) +
  theme_minimal() +
  labs(title = "Distribución Normal",
       x = "Valores de X",
       y = "Probabilidad") +
  ylim(0, 0.05)
```



b) Proporción de exámenes con puntajes entre 85 y 98

Debemos calcular $P(85 \leq X \leq 98)$, que es la probabilidad de que un estudiante obtenga una nota entre 85 y 98. Esto es equivalente a calcular $P(X \leq 98)$ y restar el resultado de $P(X \leq 85)$.

```
# Calcular la probabilidad acumulada de obtener una nota menor o igual a 98
prob_98 <- pnorm(98, mu, sigma)
```

```
# Calcular la probabilidad acumulada de obtener una nota menor o igual a 85
prob_85 <- pnorm(85, mu, sigma)
```

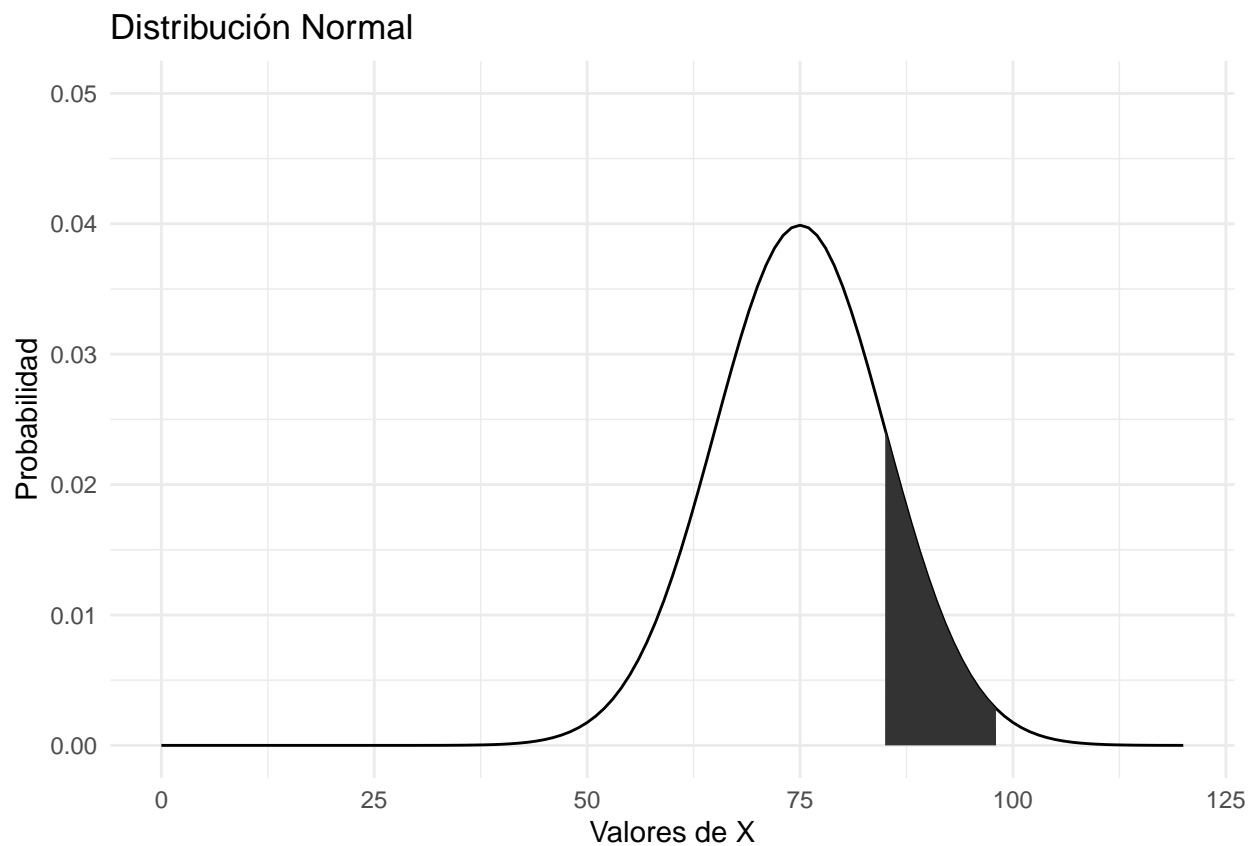


```
# Calcular la probabilidad de obtener una nota entre 85 y 98
prob_85_98 <- prob_98 - prob_85

print(paste("Probabilidad de obtener una nota entre 85 y 98:", prob_85_98))

## [1] "Probabilidad de obtener una nota entre 85 y 98: 0.147931143909781"

# visualización del área bajo la curva para  $P(85 \leq X \leq 98)$ 
ggplot(df, aes(x = x, y = y)) +
  geom_line() +
  geom_area(data = subset(df, x >= 85 & x <= 98)) +
  theme_minimal() +
  labs(title = "Distribución Normal",
       x = "Valores de X",
       y = "Probabilidad") +
  ylim(0, 0.05)
```



- c) 3. Supongamos que el profesor determina que el 20% de los estudiantes con menor nota deben ver a un tutor para ayuda extra. ¿Cuál es la nota mínima que un estudiante debe obtener para no tener que ver a un tutor?

Debemos calcular $P(X \leq x) = 0.8$, que es la probabilidad de que un estudiante obtenga una nota menor o igual a x y no tenga que ver a un tutor. Esto es equivalente a calcular $P(X > x) = 0.2$ y restar el resultado de 1.

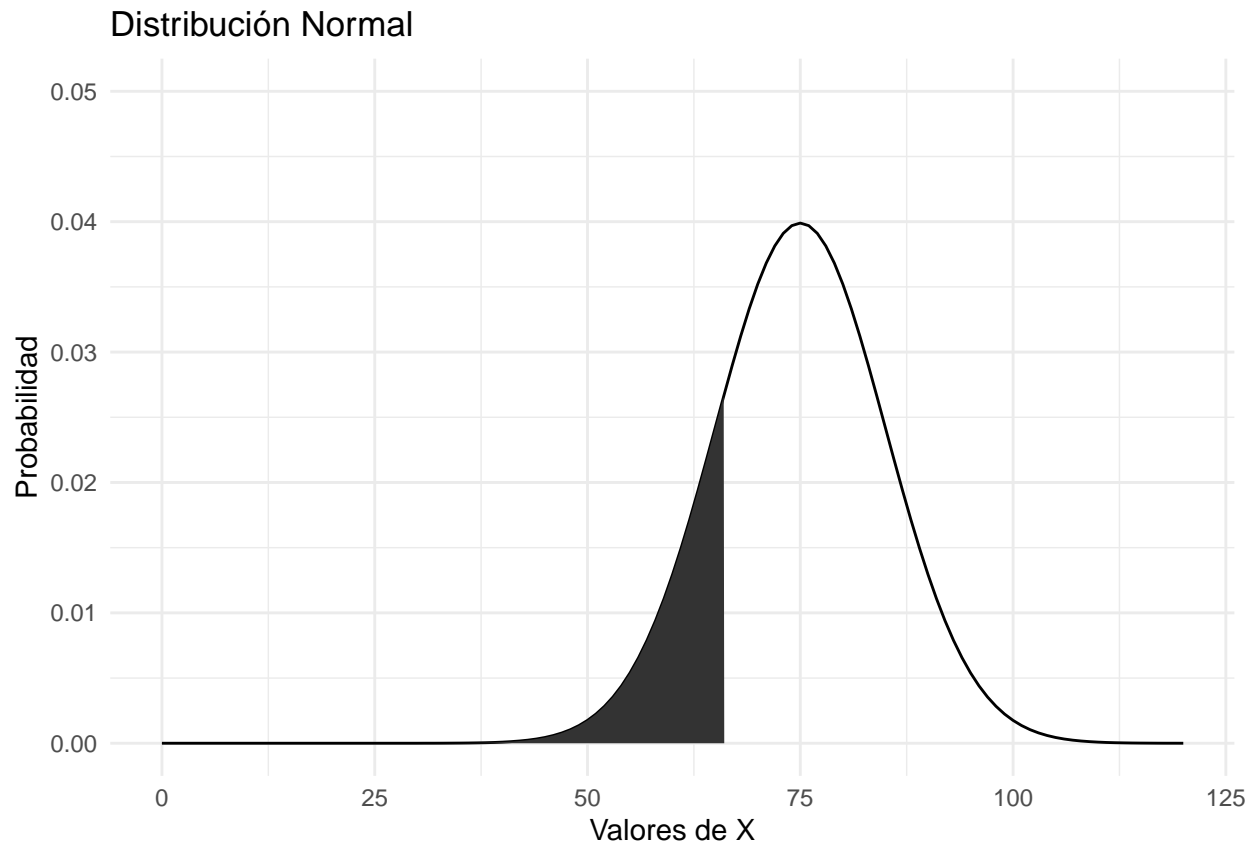
```
# Calcular la probabilidad acumulada de obtener una nota menor o igual a x
prob_x <- qnorm(0.8, mu, sigma, lower.tail = FALSE)
```

```
print(paste("Nota mínima para no tener que ver a un tutor:", prob_x))
```

```
## [1] "Nota mínima para no tener que ver a un tutor: 66.5837876642709"
```

```
# visualización del área bajo la curva para  $P(X \leq x) = 0.8$ 
```

```
ggplot(df, aes(x = x, y = y)) +
  geom_line() +
  geom_area(data = subset(df, x <= prob_x)) +
  theme_minimal() +
  labs(title = "Distribución Normal",
       x = "Valores de X",
       y = "Probabilidad") +
  ylim(0, 0.05)
```



```
# Proporción de exámenes con puntajes por encima de 90
```

```
a_1 <- pnorm(90, mu, sigma, lower.tail = FALSE)
```

```
# Proporción de exámenes con puntajes entre 85 y 98
```

```
a_2 <- pnorm(98, mu, sigma) - pnorm(85, mu, sigma)
```

```
# Nota mínima para no tener que ver a un tutor
```

```

a_3 <- qnorm(0.8, mu, sigma, lower.tail = FALSE)

print(paste("Proporción de exámenes con puntajes por encima de 90:", a_1))

## [1] "Proporción de exámenes con puntajes por encima de 90: 0.0668072012688581"

print(paste("Proporción de exámenes con puntajes entre 85 y 98: ", a_2))

## [1] "Proporción de exámenes con puntajes entre 85 y 98:      0.147931143909781"

print(paste("Nota mínima para no tener que ver a un tutor:      ", a_3))

## [1] "Nota mínima para no tener que ver a un tutor:      66.5837876642709"

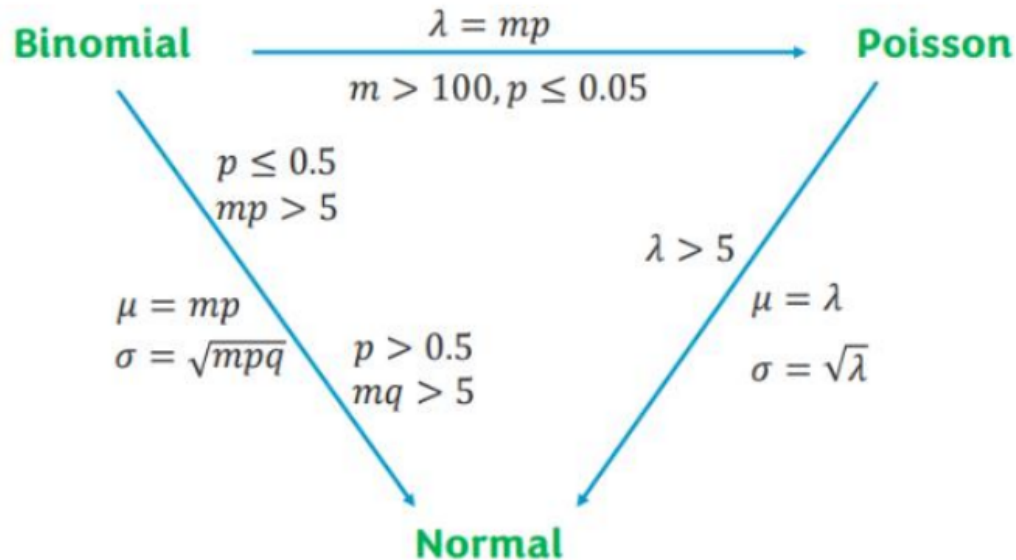
```

2.5.5 Relación entre distribuciones

- **Binomial a Normal:** cuando el número de ensayos n en una distribución binomial es grande, y p no está cerca de 0 ni de 1, la distribución binomial se puede aproximar a una distribución normal con $\mu = np$ y $\sigma = \sqrt{np(1-p)}$. Esto se conoce como el teorema del límite central. Para ello tiene que cumplirse:
 - $np \geq 5$ si $p \leq 0.5$
 - $n(1-p) \geq 5$ si $p \geq 0.5$
- **Poisson a Normal:** cuando la tasa de ocurrencia λ en una distribución de Poisson es grande, la distribución de Poisson se puede aproximar a una distribución normal con $\mu = \lambda$ y $\sigma = \sqrt{\lambda}$. Esto se conoce como la aproximación de Poisson a Normal.
- **Binomial a Poisson:** cuando el número de ensayos n en una distribución binomial es grande y la probabilidad de éxito p es pequeña, la distribución binomial se puede aproximar a una distribución de Poisson con $\lambda = np$. Esto se conoce como la aproximación de Binomial a Poisson.

Estas aproximaciones son útiles porque, para valores grandes, trabajar con distribuciones normales es matemáticamente más sencillo que trabajar con distribuciones binomiales o de Poisson.

Relación entre Binomial, Poisson y Normal



Relación entre Binomial, Poisson y Normal

Veamos lo descrito anteriormente en una gráfica.

2.5.5.1 Relación entre la distribución Binomial y Poisson Distribución binomial de parámetros $n = 20$ y $p = 0.05$

```
# Instalar y cargar el paquete ggplot2 si aún no está instalado
if(!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)

# Definir los parámetros de la distribución binomial
n <- 20
p <- 0.05

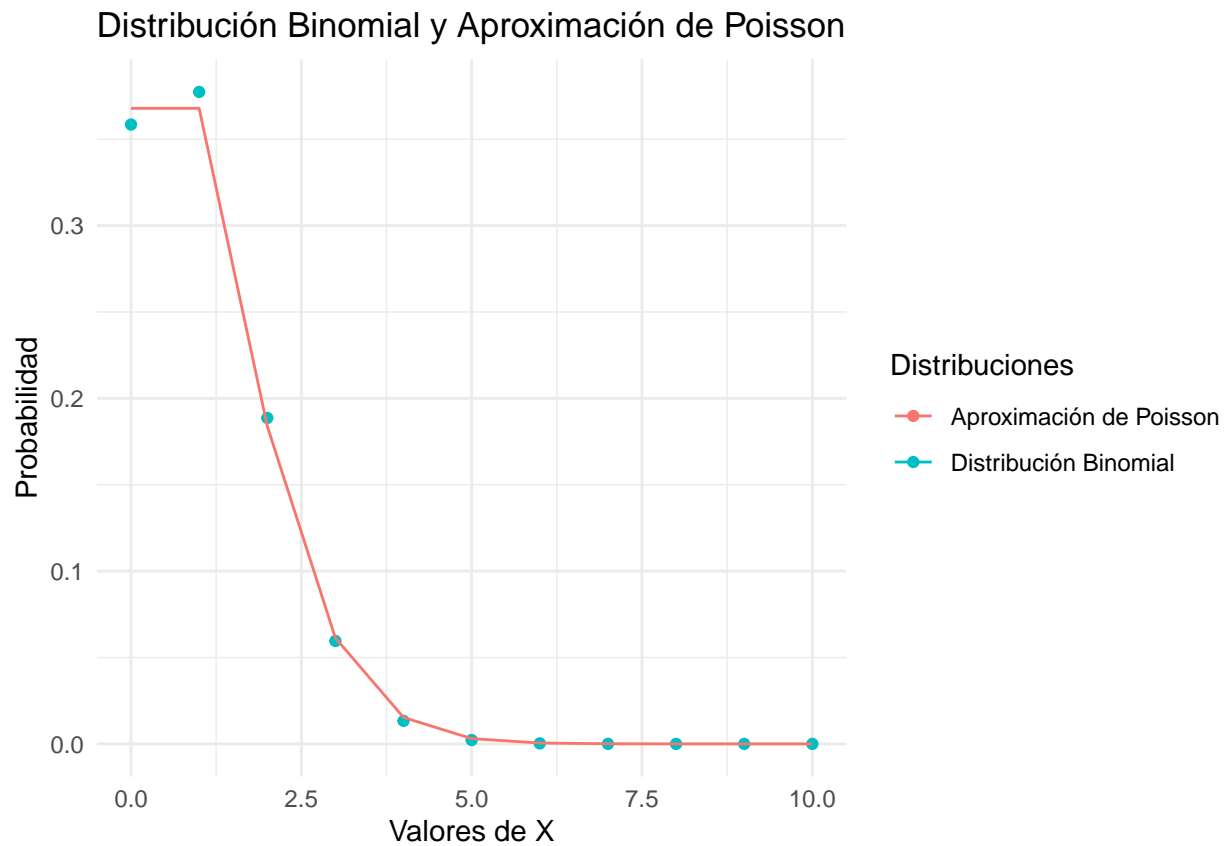
# Aproximación de Poisson de parámetro lambda = n*p
lambda <- n * p

# Crear un dataframe con los valores de x e y
x <- seq(0, 10, 1)
y <- dbinom(x, n, p)
df <- data.frame(x, y)

# Datos para la aproximación de Poisson
y_poisson <- dpois(x, lambda)

# Crear el gráfico de puntos de la distribución binomial y la aproximación de Poisson
ggplot(df, aes(x = x, y = y)) +
  geom_point(aes(y = y, colour = "Distribución Binomial")) +
  geom_line(aes(y = y_poisson, colour = "Aproximación de Poisson")) +
```

```
theme_minimal() +
labs(title = "Distribución Binomial y Aproximación de Poisson",
     x = "Valores de X",
     y = "Probabilidad",
     colour = "Distribuciones")
```



2.5.5.2 Relación entre la distribución Binomial y Normal Relación entre las ditribuciones Bino-
mial y Normal cuando $n = 20$ y $p = 0.4$

```
# Instalar y cargar el paquete ggplot2 si aún no está instalado
if(!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)

# Definir los parámetros de la distribución binomial
n <- 20
p <- 0.4

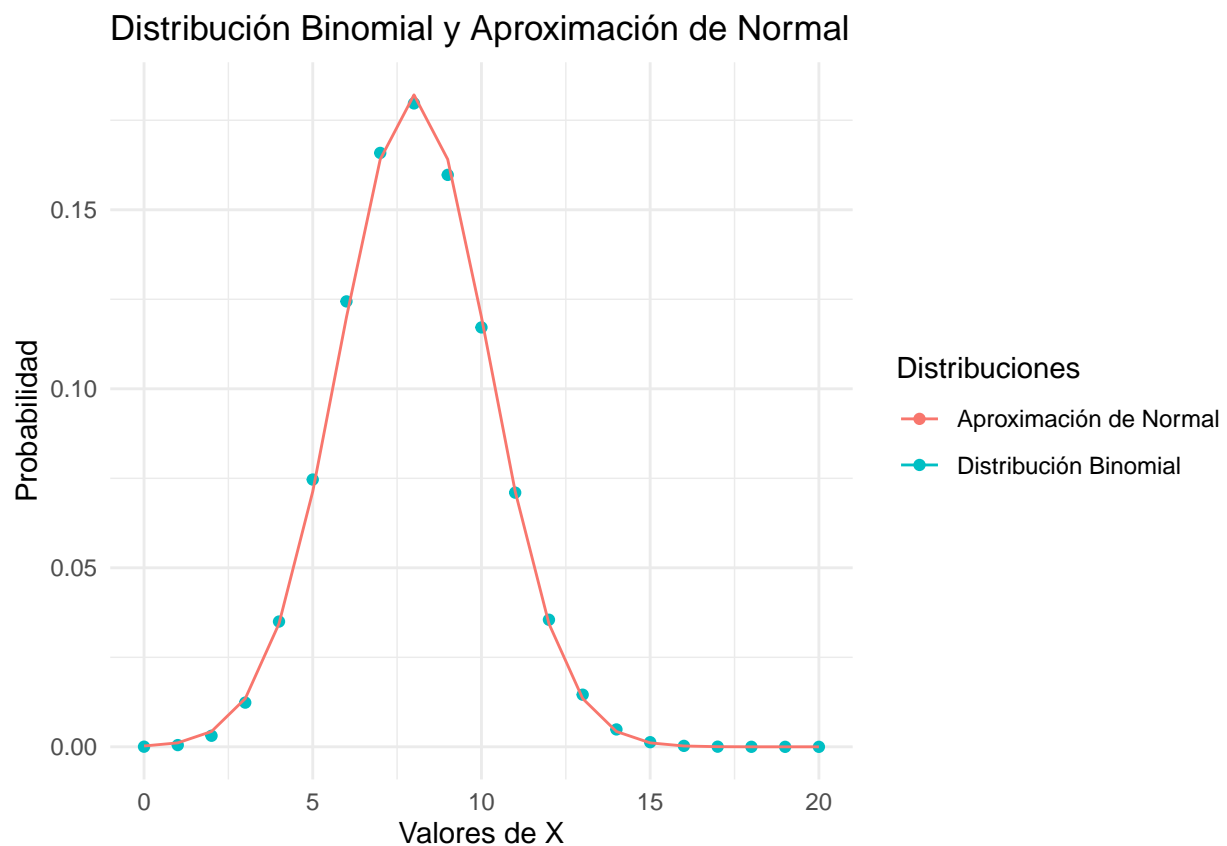
# Aproximación de Normal de parámetros mu = n*p y sigma = sqrt(n*p*(1-p))
mu <- n * p
sigma <- sqrt(n * p * (1 - p))

# Crear un dataframe con los valores de x e y
x <- seq(0, 20, 1)
y <- dbinom(x, n, p)
```

```
df <- data.frame(x, y)

# Datos para la aproximación de Normal
y_normal <- dnorm(x, mu, sigma)

# Crear el gráfico de puntos de la distribución binomial y la aproximación de Normal
ggplot(df, aes(x = x, y = y)) +
  geom_point(aes(y = y, colour = "Distribución Binomial")) +
  geom_line(aes(y = y_normal, colour = "Aproximación de Normal")) +
  theme_minimal() +
  labs(title = "Distribución Binomial y Aproximación de Normal",
       x = "Valores de X",
       y = "Probabilidad",
       colour = "Distribuciones")
```



2.5.5.3 Relación entre la distribución de Poisson y Normal

Relación entre las distribuciones Poisson y Normal cuando $\lambda = 5$

```
# Instalar y cargar el paquete ggplot2 si aún no está instalado
if(!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)

# Definir los parámetros de la distribución de Poisson
lambda <- 5
```

```

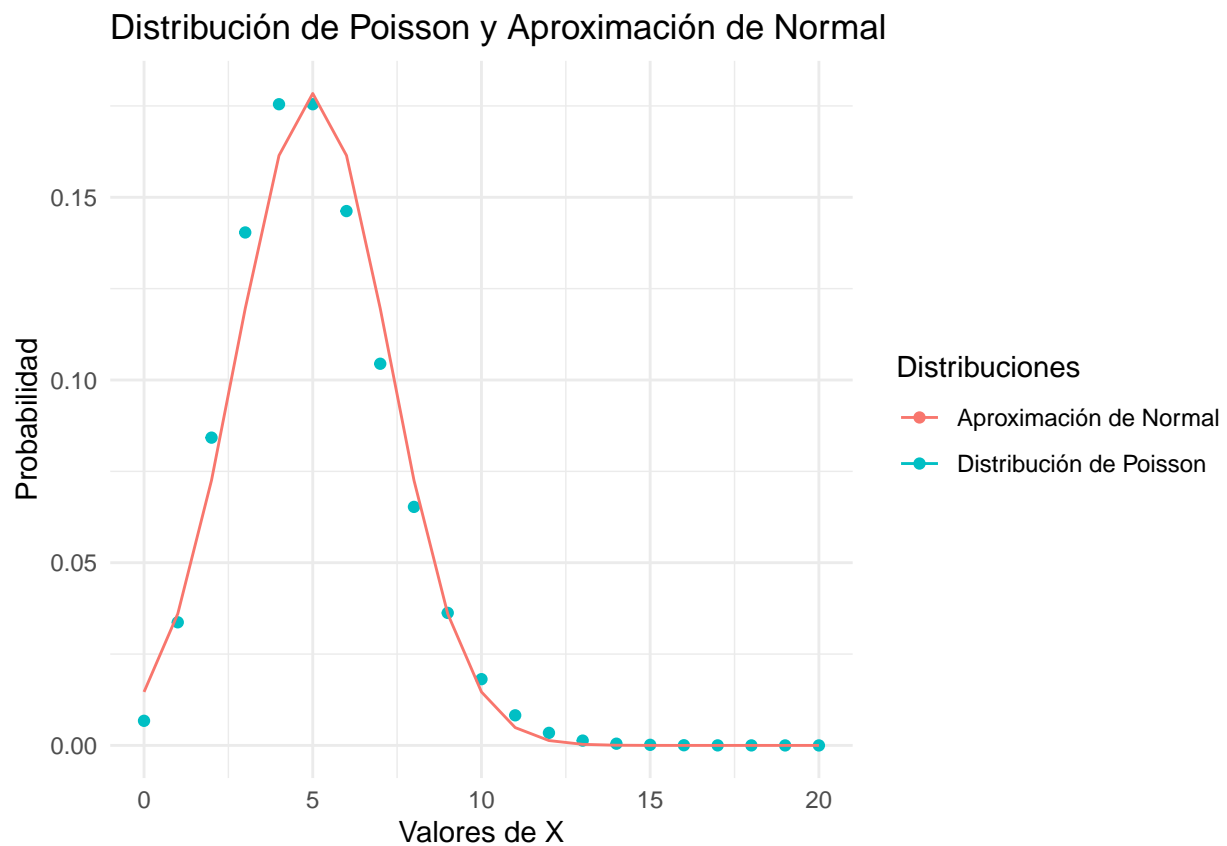
# Aproximación de Normal de parámetros  $\mu = \lambda$  y  $\sigma = \sqrt{\lambda}$ 
mu <- lambda
sigma <- sqrt(lambda)

# Crear un dataframe con los valores de x e y
x <- seq(0, 20, 1)
y <- dpois(x, lambda)
df <- data.frame(x, y)

# Datos para la aproximación de Normal
y_normal <- dnorm(x, mu, sigma)

# Crear el gráfico de puntos de la distribución de Poisson y la aproximación de Normal
ggplot(df, aes(x = x, y = y)) +
  geom_point(aes(y = y, colour = "Distribución de Poisson")) +
  geom_line(aes(y = y_normal, colour = "Aproximación de Normal")) +
  theme_minimal() +
  labs(title = "Distribución de Poisson y Aproximación de Normal",
       x = "Valores de X",
       y = "Probabilidad",
       colour = "Distribuciones")

```



2.6 Distribución t de Student

La distribución t de Student es una de las distribuciones fundamentales en estadística, especialmente útil para la estimación de parámetros y la realización de pruebas de hipótesis cuando se trabaja con tamaños de muestra pequeños y/o cuando la desviación estándar de la población no es conocida. La distribución t es similar a la distribución normal, pero con mayor área en los extremos y menor en el centro para muestras pequeñas, lo que significa que es más propensa a producir valores que se desvían significativamente de la media.

2.6.1 Conceptos

- **Forma de la distribución:** la distribución t es simétrica y tiene forma de campana, al igual que la distribución normal.
- **Grados de libertad:** los grados de libertad de una distribución t se denotan como df y determinan la forma de la distribución. Cuantos más grados de libertad tenga una distribución t, más se parecerá a una distribución normal.
- **Pruebas de hipótesis:** la distribución t se utiliza para realizar pruebas de hipótesis sobre la media de una población cuando la desviación estándar de la población no es conocida y el tamaño de la muestra es pequeño ($n < 30$).

2.6.2 Propiedades

- **Media:** $\mu = 0$ para $df > 1$ y $\mu = \infty$ para $df = 1$
- **Varianza:** $\sigma^2 = \frac{df}{df-2}$ para $df > 2$ y $\sigma^2 = \infty$ para $df \leq 2$
- **t:** $\frac{\bar{x} - \mu}{s/\sqrt{n}}$ donde \bar{x} es la media muestral, μ es la media poblacional, s es la desviación estándar muestral y n es el tamaño de la muestra.

2.6.3 Representación gráfica y comparación con la distribución normal

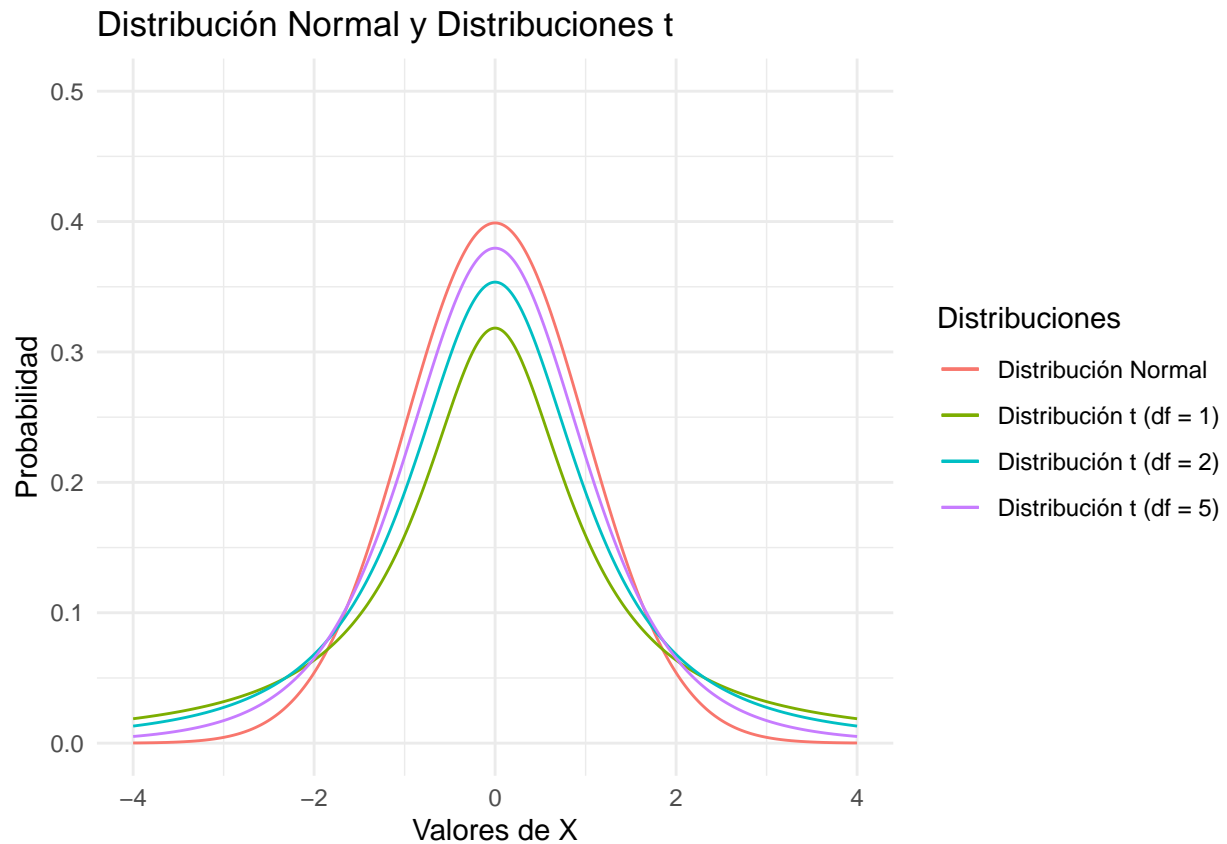
```
# Instalar y cargar el paquete ggplot2 si aún no está instalado
if(!require(ggplot2)) install.packages("ggplot2")
library(ggplot2)

# Crear un dataframe con los valores de x e y
x <- seq(-4, 4, 0.01)
y_normal <- dnorm(x, 0, 1)
y_t_1 <- dt(x, 1)
y_t_2 <- dt(x, 2)
y_t_5 <- dt(x, 5)
df <- data.frame(x, y_normal, y_t_1, y_t_2, y_t_5)

# Crear el gráfico de la distribución normal y las distribuciones t
ggplot(df, aes(x = x)) +
  geom_line(aes(y = y_normal, colour = "Distribución Normal")) +
  geom_line(aes(y = y_t_1, colour = "Distribución t (df = 1)")) +
  geom_line(aes(y = y_t_2, colour = "Distribución t (df = 2)")) +
  geom_line(aes(y = y_t_5, colour = "Distribución t (df = 5)")) +
  theme_minimal() +
  labs(title = "Distribución Normal y Distribuciones t",
       x = "Valores de X",
```



```
y = "Probabilidad",
colour = "Distribuciones") +
ylim(0, 0.5)
```



2.6.4 Ejemplos

2.6.4.1 Ejemplo 1: Prueba de Hipótesis para la media de una muestra. Supongamos que eres un investigador estudiando el efecto de un nuevo fármaco. Tienes una muestra de 15 pacientes y deseas saber si el fármaco tiene un efecto significativo en los niveles de cierto biomarcador en la sangre.

Datos hipotéticos:

- Tamaño de la muestra: $n = 15$
- Media muestral: $\bar{x} = 2.3$
- Desviación estándar muestral: $s = 0.5$
- Media poblacional: $\mu = 2.0$

3 Teoremas y propiedades

Si sumamos m variables con distribución $\text{Ber}(p)$ independientes obtenemos una variable aleatoria X con distribución binomial con parámetros m y p