Control Estadístico de Procesos

Gustavo Ahumada

Distribción hipergeométrica

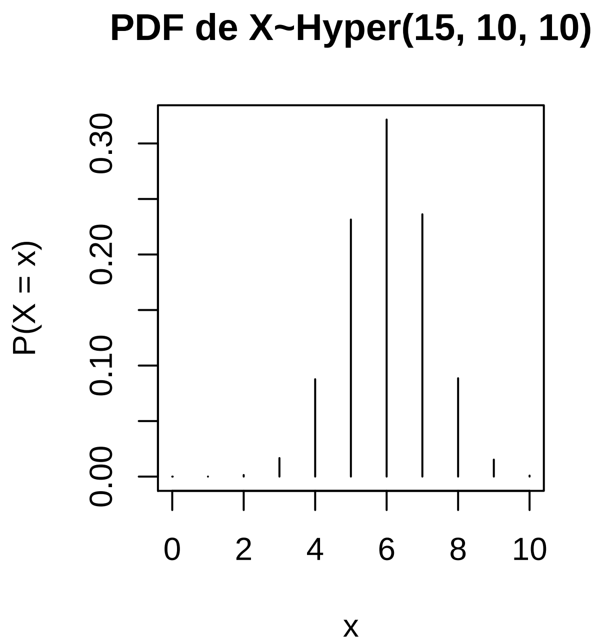
La distribución hipergeométrica no requiere independencia y se basa en el muestre que se realiza **sin reemplezo**. Las aplicaciones de la distribución hipergeométrica se encuentran en muchas áreas, con gran uso en muestreo de aceptación y garantía de calidad. Obviamente, para muchos de estos campos de muestreo se realiza a expensas del artículo de que se prueba. Es decir, el artículo se destruye y por ello no se puede reemplazar en la muestra.

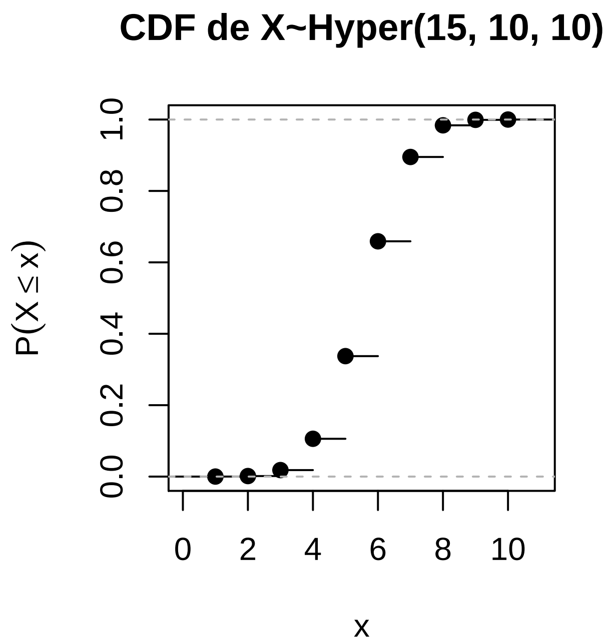
En general, nos interesa la probabilidad de seleccionar x éxitos de los k artículos considerados como éxito y $n - k$ fracasos de los $N - k$ artículos que se consideran fracasos cuando se selecciona se selecciona una muestra aleatoria de tamaño n de N artículos. Esto se conoce como **experimento hipergeométrico**.

Definición. La dist. de probabilidad de la variable aleatoria hipergeométrica X, el número de éxtios en una muestra aleatoria de tamaño n que se selecciona de N artículos de los que k son **éxito** y $N - K$ **fracaso**, es

$$h(x; N, n, k) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad x = 0, 1, 2..., n.$$

```r
opar <- par(no.readonly = TRUE)
par(mfrow = c(1, 2), pty = "s")
x <- 0:10
m <- 15
n <- 10 # N = m+n
k<- 10
px <- dhyper(x, m,n,k) # ?dhyper para mas información de esta distribución
plot(x, px, type = "h", xlab = "x", ylab = "P(X = x)",
    main = "PDF de X~Hyper(15, 10, 10)")
xs <- rep(x, round(dhyper(x,m,n,k)*100000, 0))
plot(ecdf(xs), main = "CDF de X~Hyper(15, 10, 10)",
    ylab = expression(P(X<=x)), xlab = "x" )
```






```r
par(opar)
```

Distribuciones de probabilidad univaridas continuas

Si los posibles resultados de una solo experimento se encuentran en un rango de valores, una distribución univariada será necesaria para modelas los resultados.

Distribución uniforme continua

Una distribuciones continuas más simple en la estadística es la **distribución uniforme continua**. Esta distribución se carcateriza por una función de densidad que es "plana", y por lo tanto la probabilidad es uniforme en un intervalo cerrado, digamos $[a, b]$.

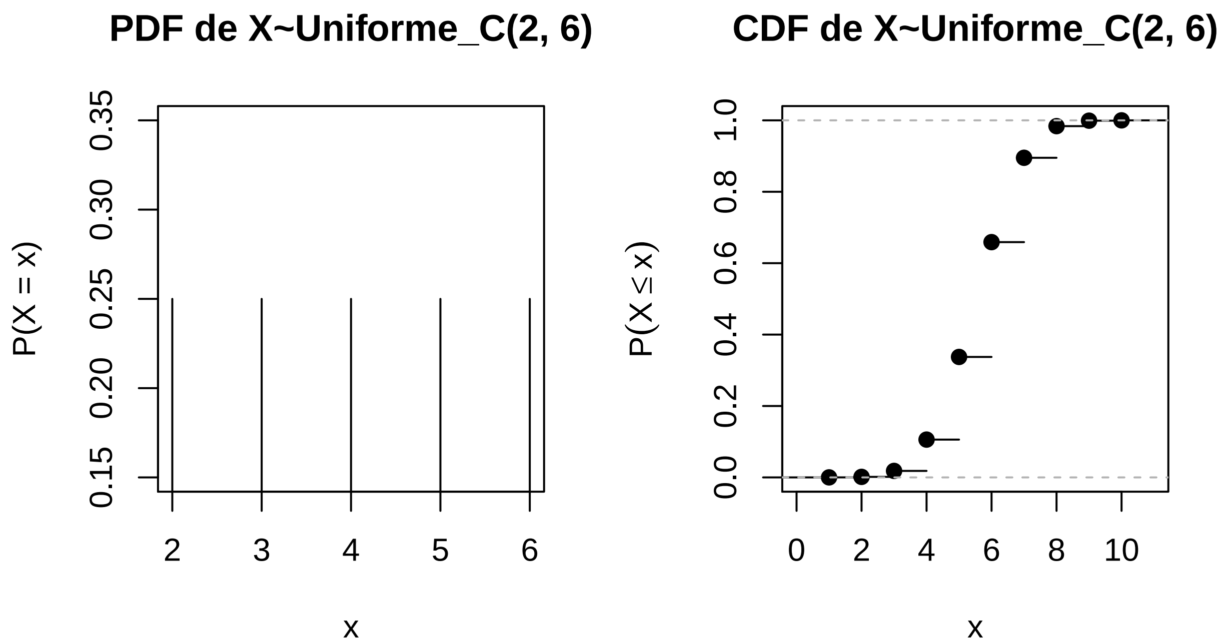
Definición. La función de densidad de la variable aleatoria uniforme continua X en el intervalo $[a, b]$ es

$$f(x; a, b) = \frac{1}{b - a}, \quad a \leq x \leq b$$

Teorema. La media y la varianza de la distribución uniforme son

$$\mu = \frac{a + b}{2} \quad y \quad \sigma^2 = \frac{(b - a)^2}{12}$$

```r
opar <- par(no.readonly = TRUE)
par(mfrow = c(1, 2), pty = "s")
x <- 2:6
px <- dunif(x, min = 2, max = 6)
plot(x, px, type = "h", xlab = "x", ylab = "P(X = x)",
    main = "PDF de X~Uniforme_C(2, 6)")
#xs <- rep(2:6, round(dunif(x, min = 2, max = 6)*100000, 0,))
plot(ecdf(xs), main = "CDF de X~Uniforme_C(2, 6)",
    ylab = expression(P(X<=x)), xlab = "x")
```




```r
par(opar)
```

Distribución Normal (Gaussiana)

La distribución continua de probabilidad más importante en todo el campo de la estadística es la **distribucón bnormal**. Esto es debido al hecho que muchas poblaciones numericas tiene distribuciones que pueden ser aproximadas con una distribución normal. Ejemplo de las distribuciones que siguen una aproximación normal incluyen las caracteristicas fisicas como el peso y la estatura de una especia particular.

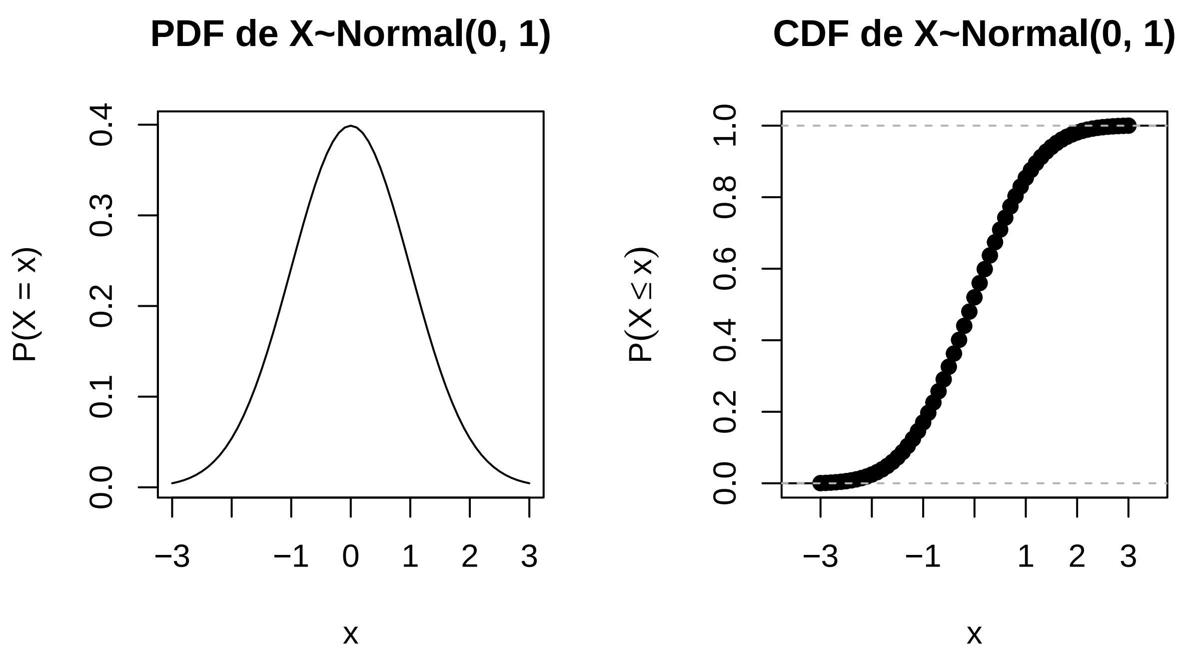
Definición. La función de densidad de la variable aleatoria normal X, con media μ y varianza σ^2, es

$$n(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma}}e^{-(1/2)[(x-\mu)/\sigma]^2} \qquad -\infty < x < \infty$$

$$donde \quad \pi = 3.14159... \quad y \quad e = 2.71828....$$

$$E[X] = \mu \quad y \quad Var[X] = \sigma^2$$

```r
opar <- par(no.readonly = TRUE)
par(mfrow = c(1, 2), pty = "s")
x <- seq(-3, +3, by=.1)
px <- dnorm(x, mean = 0, sd = 1)
plot(x, px, type = "l", xlab = "x", ylab = "P(X = x)",
    main = "PDF de X~Normal(0, 1)")
xs <- rep(x, round(dnorm(x, mean = 0, sd = 1)*100000))
plot(ecdf(xs), main = "CDF de X~Normal(0, 1)",
    ylab = expression(P(X<=x)), xlab = "x")
```




```r
par(opar)
```

La **cdf** para una variable aleatoria normal, X, con media, μ, y desviación estandar, σ, es

$$F(x) = P(X \leq x) = \frac{1}{\sqrt(2\pi\sigma^2)} \int_{-\infty}^{x} e^{-}(x-\mu)^2 2\sigma^2 dx$$

Una variable normal con $\mu = 0$ y $\sigma = 1$ frecuentemente es denotada Z, es llamada variable aleatoria **normal estándar**. La **cdf** para la distribución estándar es calculada para la primera estandarización de la variable aleatoria X, donde $X \ N(\mu, \sigma)$, empleando

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

$$F(x) = P(X \leq x) = P\left(Z \leq \frac{X-\mu}{\sigma}\right) = \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\frac{x-\mu}{\sigma}} e^{-}\frac{x^2}{2}dx.$$

Ejemplo Los resultados en una prueba estandarizada sigue una distribución normal con una media de 100 y una desviación estándar de 10. a. ¿Cual es la probabilidad que un individuo seleccionado aleatoriamente tenga un resultado entre 90 y 115? b. ¿Cual es el resultado que uno necesita para estar en el 10 con mejores notas de la clase? c. Encontrar la constante c tal que $P(100 \leq X \leq c) = 0.10$.

Solución a. Tenemos que $P(90 \leq X \leq 115) = P(X \leq 115) - P(X \leq 90)$. Para encontrar $P(X \leq 115)$ y $P(X \leq 90)$, uno puede estándarizar empleando la definición anterior. Esto es,

$$P(X \leq 115) = P\left(Z \leq \frac{115-100}{10}\right) P(X \leq 1.5)$$

y

$$P(X \leq 90) = P\left(Z \leq \frac{90-100}{10}\right) P(X \leq 1.0)$$

Consecuentemente tenemos,

$$\begin{aligned} P(90 \leq X \leq 115) &= P(-1.0 \leq Z \leq 1.5) \\ &= P(X \leq 1.5) - P(X \leq -1.0) \\ &= 0.9332 - 0.1587 = 0.7745. \end{aligned}$$

La probabilidad de seleccionar un individuo con un resultado entre 90 y 115 es 0.7754.

```
pnorm(115, 100, 10) - pnorm(90, 100, 10)
```

```
## [1] 0.7745375
```

b. Encontrar el valor c tal que 90% del área este a su ozquierda es equivalente a encontar el valor c tal que el 10% de sus área está a la derecha. Esto es, el valor c que satisface $P(X \leq c) = 0.9$ es equivalente a encontrar el valor c tal que $P(X \geq c) = 0.10$ desde que la función $qnorm()$ se refiere al área la izquierda de un valor dado por defecto, resolvemos

$$P(X \leq c) = P\left(Z = \frac{X-100}{10} \leq \frac{Xc-100}{10}\right) = 0.90 \quad para \quad c.$$

Empleando $qnorm(0.9)$, encontramos el valor Z (1.2816) tal que 90% del área en la distribución este a la izquierda de ese valor. Consecuentemente, estar en el 10% superior, uno necesita estar a más de 1.2816 desviaciones estándar por encima de la media:

$$\frac{c-100}{10} = 1.2816 \Longrightarrow c = 112.816.$$

Para estar en el 10%, uno necesita una calificación igual o superior 112.8155.

```
qnorm(0.90, 100, 10)
```

```
## [1] 112.8155
```

c. $P(105 \leq X \leq c) = 0.10$ es la igual a decir

$$P(X \leq c) = 0.10 + P(X \leq 105) = 0.10 + P\left(Z \leq \frac{105 - 100}{10}\right)$$

Empleando $pnorm(0.5)$,

$$P\left(Z \leq \frac{105 - 100}{10}\right) = P(Z \leq 0.5) = 0.6915.$$

Esto sigue que $P(X \leq c) = 0.7915$. Empleando $pnorm(0.7915)$ da 0.8116:

$$P(X \leq c) = P\left(Z \leq \frac{X - 100}{10} \leq \frac{c - 100}{10}\right) = 0.7915$$

$$esto \quad se \quad encuentra \quad resolviendo \quad \frac{c - 100}{10} = 0.8116 \Longrightarrow c = 108.116.$$

Notar que un valor Z de 0.8116 tiene un 79.15% de su área a la izquierda de dicho valor.

```
qnorm(0.1 + pnorm(105, 100, 10), 100, 10)
```

```
## [1] 108.1151
```