

# EST-24107: Simulación

**Profesor:** Alfredo Garbuno Iñigo — Primavera, 2022 — Números no uniformes.  
**Objetivo:** Que veremos.  
**Lectura recomendada:** Referencia.

## 1. GENERACIÓN DE VARIABLES NO UNIFORMES

R, por ejemplo, tiene distintos generadores de variables aleatorias. Un catálogo nos muestra que podemos generar mucho mas variables que solamente una variable uniforme.

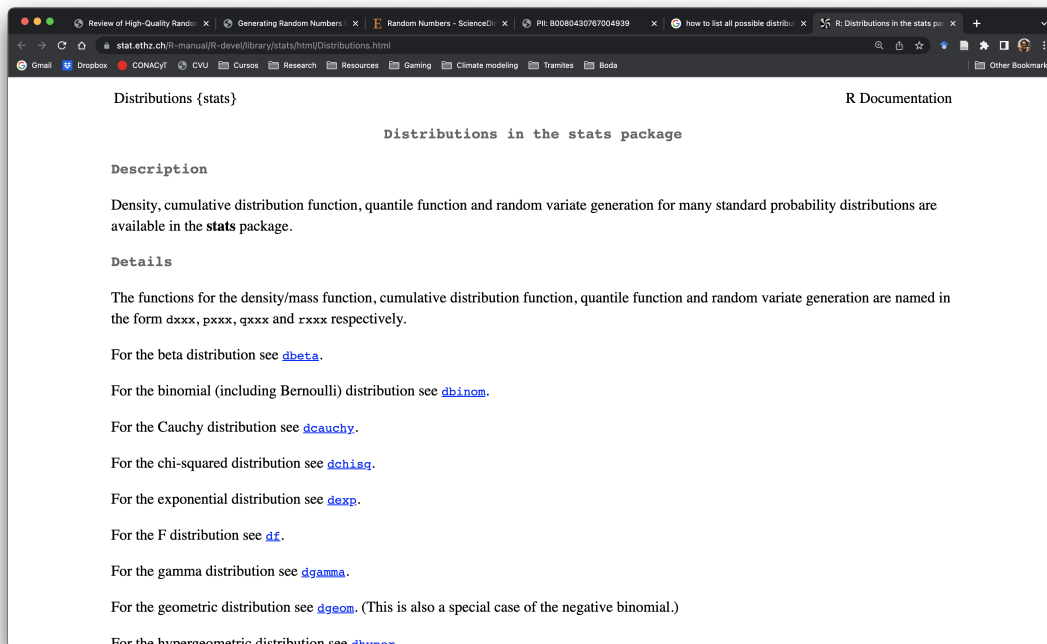


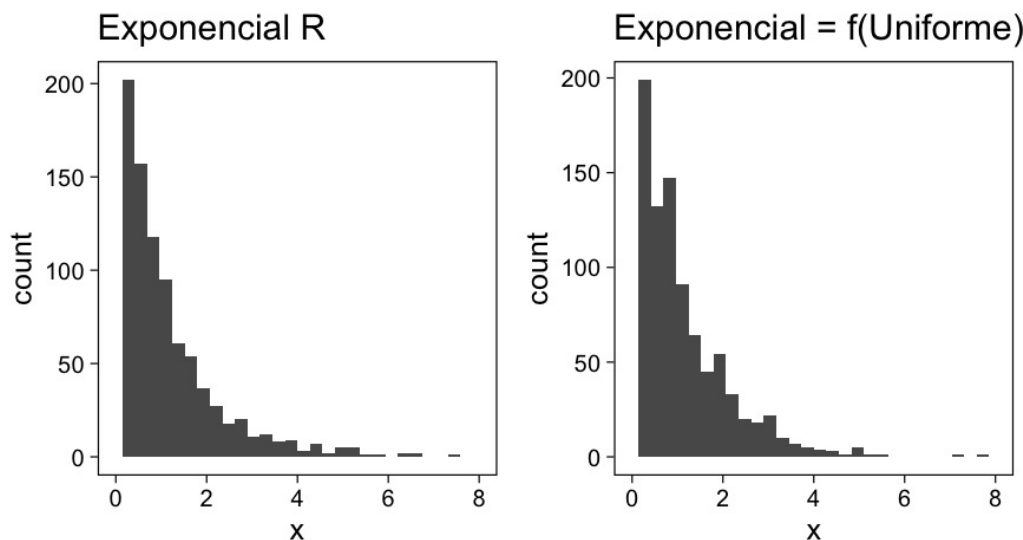
FIGURA 1. Catálogo de distribuciones en R.

En esta sección exploraremos los mecanismos tradicionales para generar números pseudo-aleatorios de distribuciones un poco mas generales que una distribución uniforme.

### 1.1. Método de la transformada inversa

De su curso de cálculo de probabilidades se acordarán que si tenemos una variable aleatoria  $X$  con función de acumulación  $\mathbb{P}$  y utilizamos  $U = \mathbb{P}(X)$ , entonces  $U$  es una variable aleatoria con distribución  $U(0, 1)$ .

**1.1.1. Ejercicio:** Considera  $X \sim \text{Exp}(1)$ , de tal forma que  $\mathbb{P}(x) = 1 - e^{-x}$ . Si sólo sabemos generar números uniformes, ¿cómo generarías números de  $X$ ?



1.1.2. *Tarea:* Considera las siguientes distribuciones:

1. Logística con  $\mu, \beta$  y función de acumulación

$$\mathbb{P}(x) = \frac{1}{1 + e^{-(x-\mu)/\beta}}. \quad (1)$$

2. Cauchy con parámetros  $\mu, \beta$  y función de acumulación

$$\mathbb{P}(x) = \frac{1}{2} + \frac{1}{\pi} \arctan((x - \mu)/\beta). \quad (2)$$

Genera números aleatorios con valores  $\mu = 1$ ,  $\beta = 2$  y comenta el rol de cada parámetro en cada distribución.

## 2. DISTRIBUCIONES CONTINUAS

### 2.1. Prueba KS

## 3. DISTRIBUCIONES DISCRETAS

### 4. PRUEBA $\chi^2$

Podemos usar otro mecanismo para probar estadísticamente si nuestros números pseudo aleatorios siguen la distribución que deseamos.

Podemos pensar en esta alternativa como la versión **discreta** de la prueba KS.

Lo que estamos poniendo a prueba es

$$H_0 : \mathbb{P}(x) = \mathbb{P}_0(x) \quad \forall x \quad \text{contra} \quad H_1 : \mathbb{P}(x) \neq \mathbb{P}_0(x) \quad \text{para alguna } x. \quad (3)$$

#### 4.1. Procedimiento de la prueba $\chi^2$

1. Hacemos una partición del rango de la distribución supuesta en  $k$  subintervalos con límites  $\{a_0, a_1, \dots, a_k\}$ , y definimos  $N_j$  como el número de observaciones (de nuestro generador de pseudo-aleatorios) en cada subintervalo.
2. Calculamos la proporción esperada de observaciones en el intervalo  $(a_{j-1}, a_j]$  como

$$p_j = \int_{a_{j-1}}^{a_j} d\mathbb{P}(x). \quad (4)$$

3. La estadística de prueba es

$$\chi^2 = \sum_{j=1}^k \frac{(N_j - p_j)^2}{np_j}. \quad (5)$$

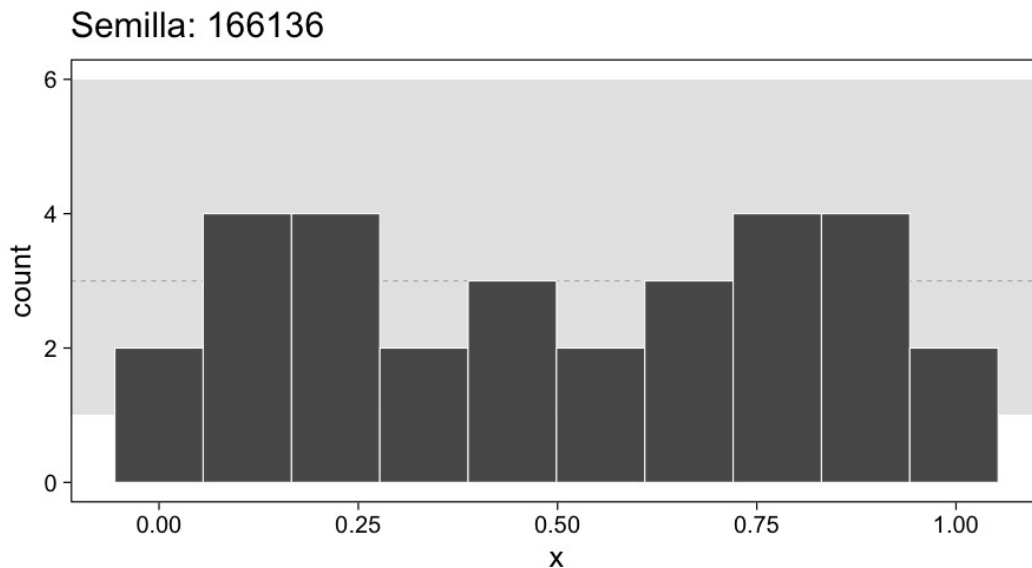
Nota que estamos comparando dos histogramas. El histograma observado que construimos a partir de nuestros números pseudo-aleatorios contra el histograma que esperaríamos de la distribución. ¿Puedes pensar en algún problema con esta prueba?

La visualización correspondiente sería lo siguiente. Utilizamos nuestro generador para obtener muestras.

```

1  ## Esto es para poner a prueba un pseudo generador
2  rpseudo.uniform <- function(nsamples, seed = 108727){
3    x0 <- seed; a <- 7**5; m <- (2**31)-1;
4    x <- x0;
5    for (jj in 2:nsamples){
6      x[jj] <- (a * x[jj-1]) %% m
7    }
8    x/m
9  }

```



4.1.1. *Pregunta:* ¿Qué esperaríamos de nuestro estadístico  $\chi^2$  si nuestro generado de pseudo-aleatorios es incorrecto?

## 4.2. Aplicación de la prueba

```

1  nsamples <- 30; nbreks <- 10
2  samples <- data.frame(x = rpseudo.uniform(nsamples))
3
4  Fn <- hist(samples$x, breaks = nbreks, plot = FALSE)$counts/nsamples
5  F0 <- 1/nbreks
6
7  X2.obs <- (nsamples*nbreks)*sum((Fn - F0)**2)

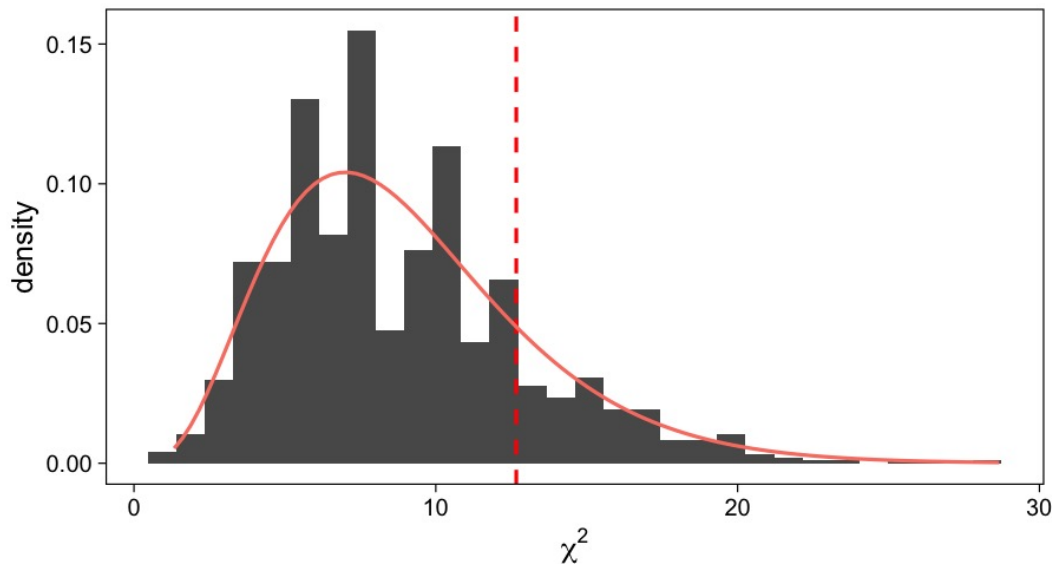
```

```

1  ## Esto es para generar datos observados de la distribución que queremos
2  experiment <- function(nsamples){
3    nbreaks <- 10
4    samples <- data.frame(x = runif(nsamples))
5    Fn <- hist(samples$x, breaks = nbreaks, plot = FALSE)$counts/nsamples
6    F0 <- 1/nbreaks
7    X2 <- (nsamples*nbreaks)*sum((Fn - F0)**2)
8    return(X2)
9  }
10
11  X2 <- c()
12  for (jj in 1:1000){
13    X2[jj] <- experiment(nsamples)
14  }

```

En la Section 4.2 se muestra el histograma de las réplicas del estadístico  $\chi^2$  bajo el generador uniforme (lo tomamos como la distribución de la hipótesis nula) y comparamos contra el observado (línea punteada). Adicional, se incorpora la densidad de una  $\chi^2_{k-1}$  (léase ji-cuadrada con  $k - 1$  grados de libertad) que es la distribución asintótica del estadístico.



Por lo tanto, la probabilidad de haber observado un estadístico  $\chi^2$  tan extremo como el que observamos si el generador hubiera sido el que suponemos es:

```

1  [1] "Estadístico: 12.6667, Probabilidad: 0.177"

```

Que podemos comparar contra el que obtenemos de una prueba "tradicional":

```

1  counts.obs <- Fn*nsamples
2  chisq.test(counts.obs, p = rep(1, nbreaks)/nbreaks, simulate.p.value = TRUE)

```

```

1
2      Chi-squared test for given probabilities with simulated p-value (based
3      on 2000 replicates)
4
5  data:  counts.obs
6  X-squared = 13, df = NA, p-value = 0.2

```

- La prueba  $\chi^2$  pues usualmente no es buena cuando el número de observaciones es menor a 50.
- La prueba KS tiene mejor potencia que la prueba  $\chi^2$ :

```
1 ks.test(samples$x, "punif")
```

```
1  
2         Exact one-sample Kolmogorov-Smirnov test  
3  
4 data:  samples$x  
5 D = 0.16, p-value = 0.4  
6 alternative hypothesis: two-sided
```

### 4.3. Aplicación de pruebas

En la práctica se utiliza una colección de pruebas pues cada una es sensible a cierto tipo de desviaciones. La batería de pruebas mas utilizada es la colección de pruebas DieHARD que desarrolló [George Marsaglia](#) y que se ha ido complementando con los años.

## 5. REFERENCES

### REFERENCIAS