eman ta zabal zazu

Universidad del País Vasco — Euskal Herriko Unibertsitatea

Material de estudio

OCW 2019: *Curso práctico para el análisis e inferencia estadística con Mathematica*

Tema 5. Variable aleatoria discreta

Equipo docente del curso
Arrospide Zabala, Eneko
Martín Yagüe, Luis
Unzueta Inchaurbe, Aitziber
Soto Merino, Juan Carlos
Durana Apaolaza, Gaizka
Bikandi Irazabal, Iñaki

Departamento de Matemática Aplicada
Escuela de Ingeniería de Bilbao, Edificio II-I

TEMA 5. VARIABLE ALEATORIA DISCRETA

Introducción

Definición

Se considera un experimento aleatorio en cuyo espacio muestral, Ω, está definida una función de probabilidad, P.

Una variable aleatoria, X, es una función que hace corresponder un número real a cada uno de los sucesos elementales del espacio muestral Ω.

$$\boxed{X : \Omega \longrightarrow \mathbb{R}}$$

Dominio

Una variable aleatoria es discreta si el conjunto de todos los valores que puede tomar es numerable; es decir, puede tomar un número finito o infinito numerable de valores.

Distribución de probabilidad discreta

Definición

Conjunto de todos los posibles valores que toma una variable discreta asociada a cierto experimento aleatorio junto con las probabilidades para cada uno de esos valores.

Puede darse en diferentes formas: tabla, fórmula, lista ó gráfico.

Ejemplo. Distribución de probabilidad del número de caras que aparecen al lanzar tres veces una moneda equilibrada.

- espacio muestral: $\Omega = \{CCC, CCX, CXC, XCC, CXX, XCX, XXC, XXX\}$
- el resultado CCX significa que sale cara en las dos primeras tiradas y cruz en la tercera, CCC tres caras, etc.
- variable aleatoria Z : "*número de caras obtenidas al lanzar 3 veces una moneda equilibrada*"

z_i	0	1	2	3
$P(Z = z_i)$	0.125	0.375	0.375	0.125

- **ProbabilityDistribution[*pdf*,{*x*,x_{min},x_{max},*dx*}]**. Representa la distribución de probabilidad de la variable discreta x, que toma valores entre x_{min} y x_{max}, siendo *pdf* la función que asigna la probabilidad de cada valor de x. La función *pdf* toma el valor cero para $x < x_{min}$ y $x > x_{max}$.
- **Piecewise[{{val_1, $cond_1$},{val_2, $cond_2$},...},*val*]**. Representa una función definida a trozos que toma los valores val_i en las regiones definidas por las condiciones $cond_i$. Usa *val* como valor cuando no se cumple ninguna de las condiciones; por defecto, ese valor es cero.
- **EmpiricalDistribution[{x_1, x_2, ..., x_k}]**. Representa una distribución empírica basada en los valores de los datos x_i.
- **EmpiricalDistribution[{w_1, w_2, ..., w_k}→{d_1, d_2, ... d_k}]**. La distribución empírica se basa en los valores de los datos d_i que ocurren con los pesos w_i.

Función de masa de probabilidad

La función de masa de probabilidad (ó función de probabilidad) de una variable aleatoria discreta X se define como:

$$\boxed{p(x_i) = P(X = x_i) \quad \forall\, x_i\ (i = 1, 2, \ldots, n)}$$

Es la función que asocia a cada suceso del espacio muestral Ω (valores de la variable aleatoria discreta) su probabilidad de ocurrencia.

Una función de masa $p(x)$ verifica las siguientes condiciones:

- $p(x_i) \geq 0 \quad \forall\, x_i$
- $\sum_{i=1}^{n} p(x_i) = 1$

Se representa gráficamente con un diagrama de barras.

- **PDF[*dist,x*]**. Da la función de masa de probabilidad de una variable x que sigue una distribución de probabilidad *dist*.
- **Boole[*expr*]**. Devuelve el valor 1 si *expr* es cierta y el valor 0 si *expr* es falsa.
- **DiscretePlot[*expr*,{*x*,x_{min},x_{max},*dx*}]**. Genera la gráfica de los valores de *expr* cuando x toma valores entre x_{min} y x_{max}, con incremento dx.

Función de distribución acumulada

La función de distribución acumulada (ó función de probabilidad acumulada) de una variable aleatoria discreta X se define como:

$$\boxed{F(x) = P(X \leq x) \quad \forall\, x \in \mathbb{R}}$$

Es la función que asocia a cada valor real x la probabilidad de que una variable aleatoria discreta X tome valores menores o iguales que x.

- **CDF[*dist,x*]**. Da la función de distribución acumulada de una variable x que sigue una distribución de probabilidad *dist*.

Siendo $a<b$: $\boxed{P(a < x \leq b) = F(b) - F(a)}$

- **Probability[*pred,x*≈*dist*]**. Da la probabilidad de un suceso que satisface el predicado *pred* en el supuesto de que la variable aleatoria x sigue una distribución de probabilidad *dist*.
- **Distributed[*x,dist*]** o ***x*≈*dist***. Indica que la variable aleatoria x sigue una distribución de probabilidad *dist*.
- **Conditioned[*expr,cond*]** o ***expr*⊳*cond***. Representa una expresión *expr* condicionada por el predicado *cond*.

Nota. Se recomienda copiar los símbolos ≈ y ⊳ de la *Ayuda* del programa. Se indican alias para su obtención por teclado: `<Esc>+dist+<Esc>` (≈) y `<Esc>+cond+<Esc>` (⊳).

Valor esperado

El valor esperado de una variable aleatoria discreta X es una media ponderada de los posibles valores de X en la que el peso de un valor determinado coincide con la probabilidad de que X tome ese valor.

Por tanto, se define como:

$$\boxed{E[X] = \sum_{i=1}^{n} x_i \cdot p(x_i)}$$

- **Mean[*dist*]**. Da la media de la distribución de probabilidad *dist*.
- **Expectation[*expr*,*x*≈*dist*]**. Da el valor esperado de *expr* en el supuesto de que la variable aleatoria *x* siga una distribución de probabilidad *dist*.
- **Moment[*dist*,1]**. Da el primer momento de la distribución de probabilidad *dist*.

Varianza

La varianza de una variable aleatoria discreta X es el valor esperado del cuadrado de las desviaciones repecto de la media de X.

Se define como:

$$\boxed{Var[X] = E\left[(X - E[X])^2\right]}$$

- **Variance[*dist*]**. Da la varianza de la distribución de probabilidad *dist*.
- **CentralMoment[*dist*,2]**. Da el segundo momento central de la distribución de probabilidad *dist*.

Cuantiles

- **Median[*dist*]**. Da la mediana de la distribución de probabilidad *dist*.
- **Quantile[*dist*,*q*]**. Da un cuantil de la distribución de probabilidad *dist*.
- **CDF[*dist*,*x*]**. Da la función de distribución acumulada de una distribución de probabilidad *dist* evaluada para el valor x.

Forma y simetría

- **Skewness[*dist*]**. Da el coeficiente de asimetría de la distribución de probabilidad *dist*.
- **Kurtosis[*dist*]**. Da el coeficiente de apuntamiento de la distribución de probabilidad *dist*.

Ejemplo

Se considera una variable aleatoria discreta X cuya distribución de probabilidad viene dada por la función:

$$\begin{cases} \frac{1}{3} - \frac{x}{30} & si\ 4 \leqslant x \leqslant 10 \\ \frac{x}{20} & si\ 0 \leqslant x < 4 \end{cases}$$

- función de masa de probabilidad

```
dist1 = ProbabilityDistribution[
   Piecewise[{{x / 20, 0 ≤ x < 4}, {1 / 3 - x / 30, 4 ≤ x ≤ 10}}], {x, 0, 10, 1}];
```

```
PDF[dist1, x]
```

$$\begin{cases} \frac{1}{3} - \frac{x}{30} & 4 \le x \le 10 \\ \frac{x}{20} & 0 \le x < 4 \\ 0 & \text{True} \end{cases}$$

- distribución de probabilidad de la variable aleatoria X

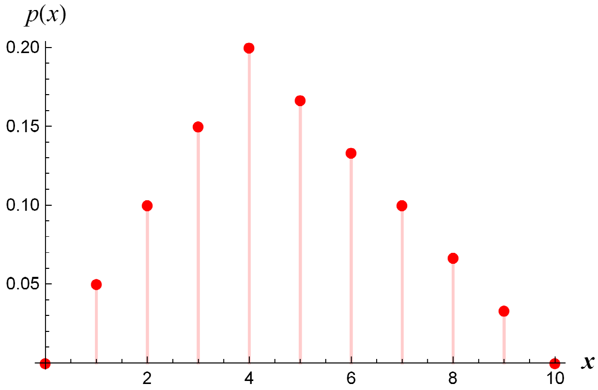
```
valores = Table[i, {i, 0, 10}]; PDF[dist1, valores]
```

$$\left\{0, \frac{1}{20}, \frac{1}{10}, \frac{3}{20}, \frac{1}{5}, \frac{1}{6}, \frac{2}{15}, \frac{1}{10}, \frac{1}{15}, \frac{1}{30}, 0\right\}$$

x_i	0	1	2	4	4	5	6	7	8	9	10
$P(X = x_i)$	0	$\frac{1}{20}$	$\frac{1}{10}$	$\frac{3}{20}$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{2}{15}$	$\frac{1}{10}$	$\frac{1}{15}$	$\frac{1}{30}$	0

- representación gráfica de la función de masa

```
DiscretePlot[PDF[dist1, x], {x, 0, 10},
 AxesLabel → {HoldForm["x"], HoldForm["p(x)"]}, PlotMarkers → Automatic, PlotStyle → Red]
```



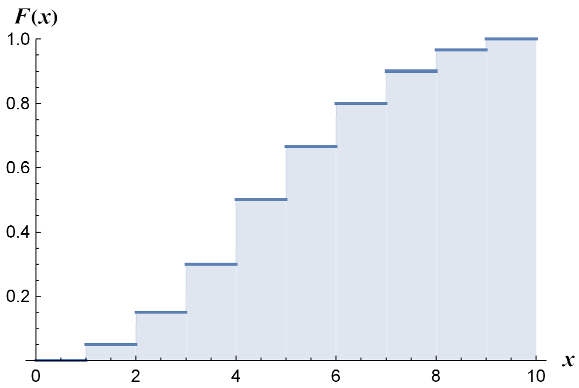
- función de distribución acumulada

```
CDF[dist1, x]
```

$$\begin{cases} \frac{1}{20} & 1 \le x < 2 \\ \frac{3}{20} & 2 \le x < 3 \\ \frac{3}{10} & 3 \le x < 4 \\ \frac{1}{2} & 4 \le x < 5 \\ \frac{2}{3} & 5 \le x < 6 \\ \frac{4}{5} & 6 \le x < 7 \\ \frac{9}{10} & 7 \le x < 8 \\ \frac{29}{30} & 8 \le x < 9 \\ 1 & x \ge 9 \\ 0 & \text{True} \end{cases}$$

- representación gráfica

```
Plot[CDF[dist1, x], {x, 0, 10}, Filling → Axis,
 PlotRange → {0, 1}, AxesLabel → {HoldForm["x"], HoldForm["F(x)"]}]
```



- cálculo de probabilidades
  - $P(x \leq 4)$

```
CDF[dist1, 4] // N
```

0.5

  - $P(x \leq 9)$

```
CDF[dist1, 9]
```

1

  - $P(x > 3) = 1 - P(x \leq 3)$

```
1 - CDF[dist1, 3] // N
```

0.7

  - $P(3 \leq x < 7)$

```
Probability[3 ≤ x < 7, x ≈ dist1]
```

$\frac{13}{20}$

  - $P(5 < x < 10)$

```
Probability[5 < x < 10, x ≈ dist1]
```

$\frac{1}{3}$

  - $P(x > 5 \mid x < 8)$

```
Probability[Conditioned[x > 5, x < 8], x ≈ dist1]
```

$\frac{7}{27}$

```
Probability[x > 5  x < 8, x ≈ dist1]
```

$\frac{7}{27}$

  - $P(x > 5 \mid x < 8) = \frac{P(x > 5 \bigcap x < 8)}{P(x < 8)} = \frac{P(5 < x < 8)}{P(x < 8)}$

```
Probability[5 < x < 8, x ≈ dist1] / Probability[x < 8, x ≈ dist1]
```

$\frac{7}{27}$

- valor esperado

```
{Mean[dist1], Expectation[x, x ≈ dist1], Moment[dist1, 1]}
```

$\left\{\frac{14}{3}, \frac{14}{3}, \frac{14}{3}\right\}$

- varianza

```
{Variance[dist1], Expectation[(x - Mean[dist1])^2, x ≈ dist1], CentralMoment[dist1, 2]}
```

$\left\{\frac{73}{18}, \frac{73}{18}, \frac{73}{18}\right\}$

- cuantiles
  - mediana

```
Median[dist1]
```

4

  - cuartiles

```
Quantile[dist1, {0.25, 0.50, 0.75}]
```

{3, 4, 6}

  - percentil 30

```
Quantile[dist1, 0.30]
```

3

- simetría y forma

```
Skewness[dist1] // N (* asimetría a la derecha *)
```

0.20316

```
Kurtosis[dist1] // N (*distribución platicúrtica*)
```

2.35068

Ejemplo

Distribución de probabilidad de la variable aleatoria usada para contar el número de caras que salen al lanzar tres veces una moneda equilibrada.

- espacio muestral: $\Omega = \{CCC, CCX, CXC, XCC, CXX, XCX, XXC, XXX\}$
- variable aleatoria Z : "*número de caras obtenidas al lanzar 3 veces una moneda equilibrada*"
- función de masa de probabilidad

```
dist2 = EmpiricalDistribution[{1/8, 3/8, 3/8, 1/8} -> {0, 1, 2, 3}];
```

```
PDF[dist2, x]
```

$\frac{1}{8}$ Boole[0 == x] + $\frac{3}{8}$ Boole[1 == x] + $\frac{3}{8}$ Boole[2 == x] + $\frac{1}{8}$ Boole[3 == x]

  - distribución de probabilidad de la variable aleatoria Z

```
valores2 = Table[i, {i, 0, 3}]; PDF[dist2, valores2]
```

$\left\{\frac{1}{8}, \frac{3}{8}, \frac{3}{8}, \frac{1}{8}\right\}$

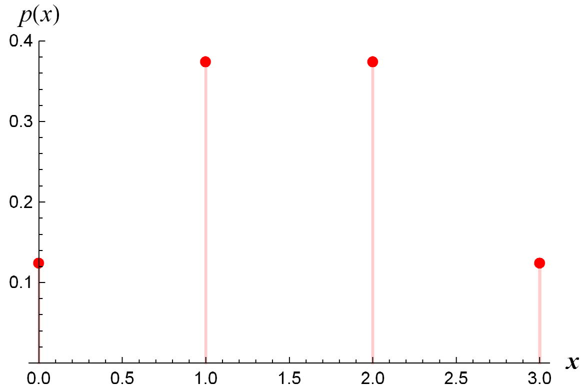
```
PDF[dist2, valores2] // N
```

{0.125, 0.375, 0.375, 0.125}

z_i	0	1	2	3
$P(Z = z_i)$	0.125	0.375	0.375	0.125

- representación gráfica de la función de masa

```
DiscretePlot[PDF[dist2, x], {x, 0, 3}, AxesLabel → {HoldForm["x"], HoldForm["p(x)"]},
 PlotMarkers → Automatic, PlotStyle → Red, PlotRange → {0, 0.4}]
```



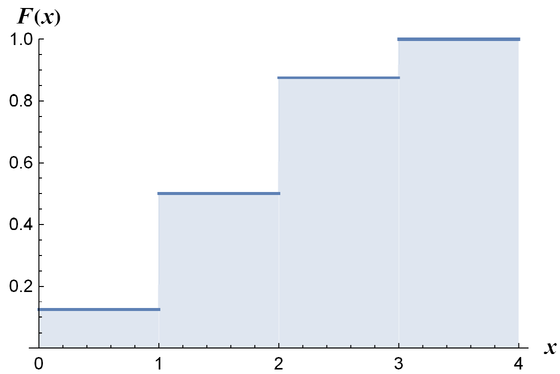
- función de distribución acumulada

```
CDF[dist2, x]
```

$\frac{1}{8}$ Boole[0 ≤ x] + $\frac{3}{8}$ Boole[1 ≤ x] + $\frac{3}{8}$ Boole[2 ≤ x] + $\frac{1}{8}$ Boole[3 ≤ x]

- representación gráfica

```
Plot[CDF[dist2, x], {x, 0, 4}, Filling → Axis,
 PlotRange → {0, 1}, AxesLabel → {HoldForm["x"], HoldForm["F(x)"]}]
```



- cálculo de probabilidades
  - $P(x \leq 2)$

```
CDF[dist2, 2]
```

$\frac{7}{8}$

  - $P(x \leq 3)$

```
CDF[dist2, 3]
```

1

  - $P(1 \leq x < 2)$

```
Probability[1 ≤ x < 2, x ≈ dist2]
```

$\frac{3}{8}$

- valor esperado

```
{Mean[dist2], Expectation[x, x≈dist2], Moment[dist2, 1]}
```

$\left\{\frac{3}{2}, \frac{3}{2}, \frac{3}{2}\right\}$

- varianza

```
{Variance[dist2], Expectation[(x - Mean[dist2])^2, x≈dist2], CentralMoment[dist2, 2]}
```

$\left\{\frac{3}{4}, \frac{3}{4}, \frac{3}{4}\right\}$

- cuantiles
  - mediana

```
Median[dist2]
```

1

  - cuartiles

```
Quantile[dist2, {0.25, 0.50, 0.75}]
```

{1, 1, 2}

  - percentil 30

```
Quantile[dist2, 0.30]
```

1

- simetría y forma

```
Skewness[dist2] // N (* simétrica *)
```

0.

```
Kurtosis[dist2] // N (*distribución platicúrtica*)
```

2.33333

Modelos de distribución de probabilidad

Introducción

Para determinar la distribución de probabilidad de una variable aleatoria basta con conocer la función de masa. Esto, a priori, no siempre es posible.

Se presentan una serie de modelos teóricos de distribución de probabilidad cuyas funciones de masa pueden resultar adecuadas para determinadas variables aleatorias discretas.

Se considera, en cada caso y de forma genérica, una variable aleatoria discreta X que puede tomar los valores $\{x_i \,/\, i = 1, 2, \ldots, n\}$.

Uniforme discreta

Una variable aleatoria X sigue una distribución uniforme discreta si cada uno de los n valores que puede tomar tiene la misma probabilidad.

Notación: $\boxed{X \sim UD(n)}$

Considerando los valores mínimo, x_{min}, y máximo, x_{max}, que puede tomar la variable se tiene:

$$\boxed{X \sim UD\,(x_{min},\, x_{max})}$$

- **DiscreteUniformDistribution[{x_{min},x_{max}}]**. Representa una distribución uniforme discreta definida sobre los enteros desde x_{min} hasta x_{max}.

- función de masa de probabilidad

```
PDF[DiscreteUniformDistribution[{xmin, xmax}], x] // TraditionalForm
```

$$\begin{cases} \frac{1}{x_{max}-x_{min}+1} & x_{min} \le x \le x_{max} \\ 0 & \text{True} \end{cases}$$

- función de distribución acumulada

```
CDF[DiscreteUniformDistribution[{xmin, xmax}], x] // TraditionalForm
```

$$\begin{cases} \frac{\lfloor x \rfloor - x_{min}+1}{x_{max}-x_{min}+1} & x_{min} \le x < x_{max} \\ 1 & x \ge x_{max} \end{cases}$$

- valor esperado

```
Mean[DiscreteUniformDistribution[{xmin, xmax}]] // TraditionalForm
```

$$\frac{1}{2}(x_{max} + x_{min})$$

```
Expectation[x, x ≈ DiscreteUniformDistribution[{xmin, xmax}]]
```

$$\frac{1}{2}(x_{max} + x_{min})$$

- varianza

```
Variance[DiscreteUniformDistribution[{xmin, xmax}]] // TraditionalForm
```

$$\frac{1}{12}\left((x_{max} - x_{min} + 1)^2 - 1\right)$$

Ejemplo. Se considera el experimento consistente en observar la puntuación obtenida en el lanzamiento de un dado equilibrado.

- variable aleatoria X : "*puntuación obtenida al lanzar un dado equilibrado*"
- espacio muestral: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- función de masa de probabilidad

```
distUD = DiscreteUniformDistribution[{1, 6}];
```

```
PDF[distUD, x] // TraditionalForm
```

$$\begin{cases} \frac{1}{6} & 1 \le x \le 6 \\ 0 & \text{True} \end{cases}$$

- distribución de probabilidad de la variable aleatoria X

```
valoresUD = Table[i, {i, 1, 6}]; PDF[distUD, valoresUD]
```

$$\left\{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\right\}$$

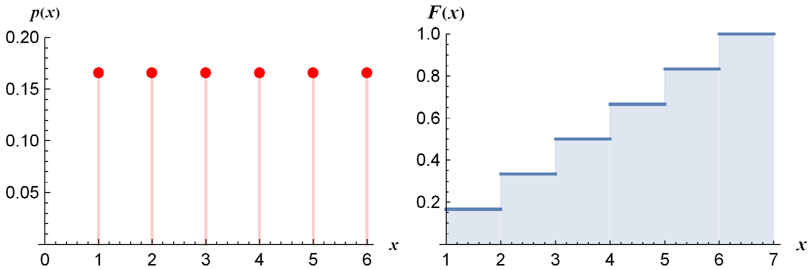
- función de distribución acumulada

```
CDF[distUD, x]
```

$$\begin{cases} \frac{\text{Floor}[x]}{6} & 1 \le x < 6 \\ 1 & x \ge 6 \\ 0 & \text{True} \end{cases}$$

- representación gráfica

```
Grid[{{
   DiscretePlot[PDF[distUD, x], {x, 1, 6}, AxesLabel → {HoldForm["x"], HoldForm["p(x)"]},
    PlotMarkers → Automatic, PlotStyle → Red, PlotRange → {0, 0.20}],
   Plot[CDF[distUD, x], {x, 1, 7}, Filling → Axis, PlotRange → {0, 1},
    AxesLabel → {HoldForm["x"], HoldForm["F(x)"]}]}}]
```



- cálculo de probabilidades
  - $P(X = 2)$

```
PDF[distUD, 2]
```

$\frac{1}{6}$

  - $P(X \le 2)$

```
CDF[distUD, 2]
```

$\frac{1}{3}$

Binomial

Se considera un experimento aleatorio consistente en la realización de n pruebas independientes de Bernoulli. Cada una sólo tiene dos posibles resultados, denominados éxito (S) y fracaso (F).

Las n pruebas de Bernoulli son independientes con lo que la probabilidad de éxito, $p = P(S)$, permanece constante.

La variable aleatoria X:*"número de éxitos en las n pruebas"* sigue un modelo de distribución binomial con parámetros n y p.

Notación: $\boxed{X \sim B(n, p)}$

- **BinomialDistribution[n,p]**. Representa una distribución binomial en la que se realizan n ensayos con una probabilidad de éxito p.

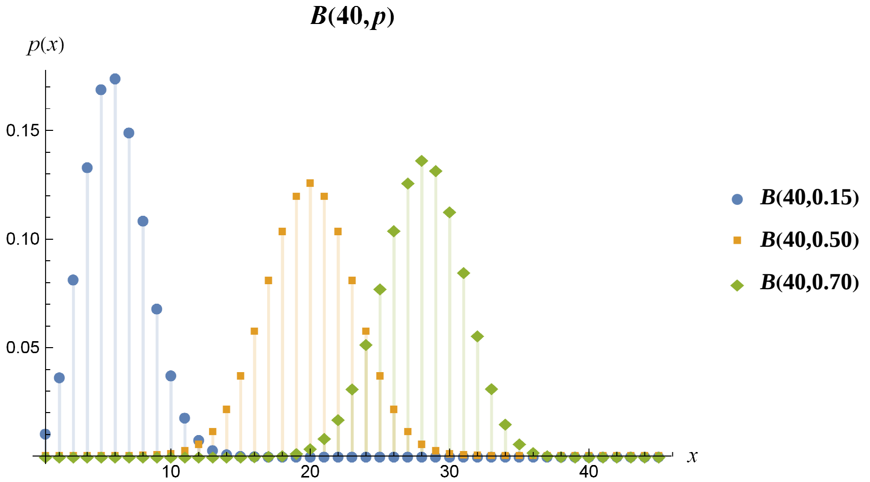
- función de masa de probabilidad

```
PDF[BinomialDistribution[n, p], x] // TraditionalForm
```

$$\begin{cases} p^x \binom{n}{x} (1-p)^{n-x} & 0 \le x \le n \\ 0 & \text{True} \end{cases}$$

- representación gráfica de la función de masa para $n = 40$ y diferentes valores de p

```
DiscretePlot[Table[PDF[BinomialDistribution[40, p], k], {p, {0.15, 0.5, 0.7}}] // Evaluate,
 {k, 45}, PlotRange → All, PlotMarkers → Automatic,
 PlotLabel → "B(40,p)", AxesLabel → {HoldForm["x"], HoldForm["p(x)"]},
 PlotLegends → {"B(40,0.15)", "B(40,0.50)", "B(40,0.70)"}]
```



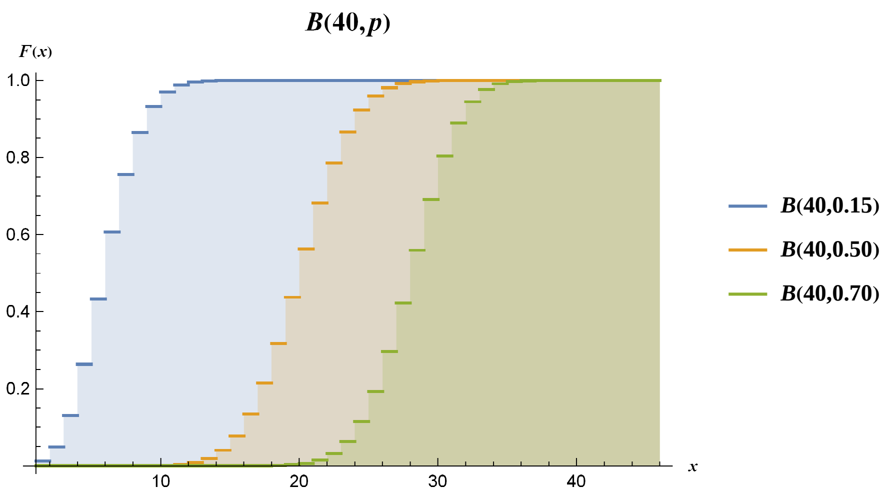
- función de distribución acumulada

```
CDF[BinomialDistribution[n, p], x] // TraditionalForm
```

$$\begin{cases} I_{1-p}(n - \lfloor x \rfloor, \lfloor x \rfloor + 1) & 0 \le x < n \\ 1 & x \ge n \end{cases}$$

- representación gráfica de la función de distribución para $n = 40$ y diferentes valores de p

```
DiscretePlot[Table[CDF[BinomialDistribution[40, p], k], {p, {0.15, 0.5, 0.7}}] // Evaluate,
 {k, 45}, ExtentSize → Right, PlotLabel → "B(40,p)",
 AxesLabel → {HoldForm["x"], HoldForm["F(x)"]},
 PlotLegends → {"B(40,0.15)", "B(40,0.50)", "B(40,0.70)"}]
```



- valor esperado

```
Mean[BinomialDistribution[n, p]] // TraditionalForm
```

$n\,p$

```
Expectation[x, x ≈ BinomialDistribution[n, p]] // TraditionalForm
```

$n\,p$

- varianza

```
Variance[BinomialDistribution[n, p]] // TraditionalForm
```

$n\,(1-p)\,p$

Ejemplo. Se considera el experimento aleatorio consistente en lanzar siete veces un dado equilibrado y contar el número de veces que sale el número seis.

- variable aleatoria X : "*número de veces que sale el 6 al lanzar un dado equilibrado 7 veces*"
- espacio muestral: $\Omega = \{0, 1, 2, 3, 4, 5, 6, 7\}$
- función de masa de probabilidad

```
distB = BinomialDistribution[7, 1/6];
```

```
PDF[distB, x] // TraditionalForm
```

$$\begin{cases} \frac{5^{7-x}\binom{7}{x}}{279\,936} & 0 \le x \le 7 \\ 0 & \text{True} \end{cases}$$

  - distribución de probabilidad de la variable aleatoria X

```
valoresB = Table[i, {i, 0, 7}]; PDF[distB, valoresB] // N
```

{0.279082, 0.390714, 0.234429, 0.0781429, 0.0156286, 0.00187543, 0.000125029, 3.57225×10^{-6}}

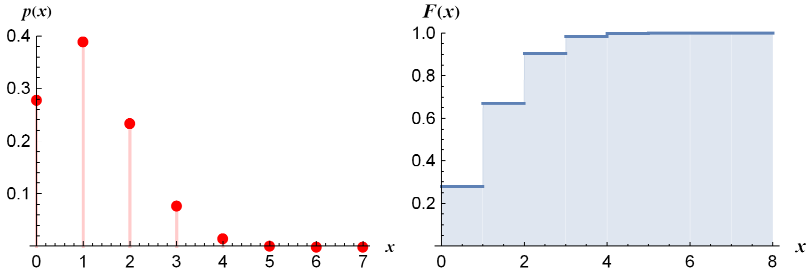
- función de distribución acumulada

```
CDF[distB, x] // TraditionalForm
```

$$\begin{cases} I_{\frac{5}{6}}(7-\lfloor x\rfloor, \lfloor x\rfloor+1) & 0 \le x < 7 \\ 1 & x \ge 7 \end{cases}$$

- representación gráfica

```
Grid[{{
   DiscretePlot[PDF[distB, x], {x, 0, 7}, AxesLabel → {HoldForm["x"], HoldForm["p(x)"]},
    PlotMarkers → Automatic, PlotStyle → Red, PlotRange → {0, 0.40}],
   Plot[CDF[distB, x], {x, 0, 8}, Filling → Axis, PlotRange → {0, 1},
    AxesLabel → {HoldForm["x"], HoldForm["F(x)"]}]}}]
```



- cálculo de probabilidades
  - $P(X = 2)$

```
PDF[distB, 2] // N
```

0.234429

  - $P(X \leq 2)$

```
CDF[distB, 2] // N
```

0.904225

Binomial negativa

Se considera un experimento aleatorio consistente en la realización de n pruebas independientes de Bernoulli. Cada una sólo tiene dos posibles resultados, denominados éxito (S) y fracaso (F). Como las n pruebas son independientes la probabilidad de éxito, $p = P(S)$, permanece constante. Las pruebas se realizan hasta que se observan un total de r éxitos.

La variable aleatoria Y:*“número de fracasos hasta la obtención del r-ésimo éxito”* sigue un modelo de distribución binomial negativa con parámetros r y p.

Notación: $\boxed{Y \sim BN(r, p)}$

  - **NegativeBinomialDistribution[*r*,*p*]**. Representa una distribución binomial negativa de parámetros r y p.

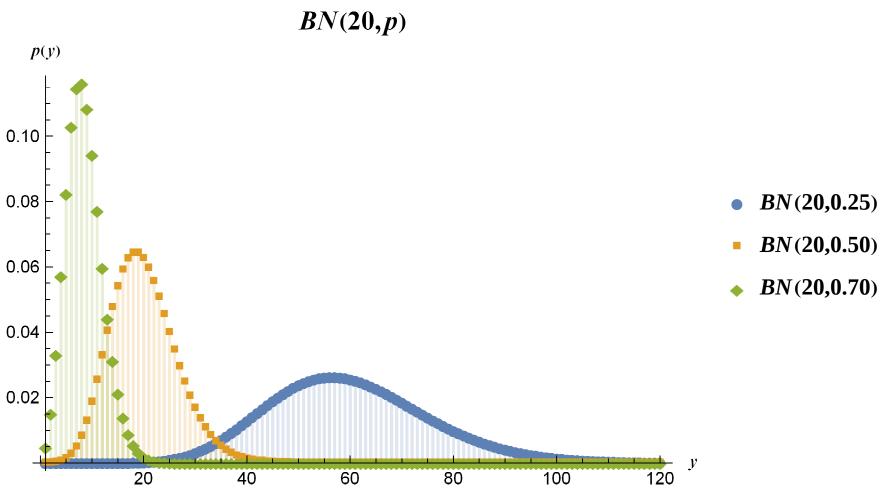
- función de masa de probabilidad

```
PDF[NegativeBinomialDistribution[r, p], y] // TraditionalForm
```

$$\begin{cases} p^r (1-p)^y \binom{r+y-1}{r-1} & y \geq 0 \\ 0 & \text{True} \end{cases}$$

- representación gráfica de la función de masa para $r = 20$ y diferentes valores de p

```
DiscretePlot[Table[PDF[NegativeBinomialDistribution[20, p], k], {p, {0.25, 0.5, 0.7}}] //
   Evaluate, {k, 120}, PlotRange → All, PlotMarkers → Automatic,
  PlotLabel → "BN(20,p)", AxesLabel → {HoldForm["y"], HoldForm["p(y)"]},
  PlotLegends → {"BN(20,0.25)", "BN(20,0.50)", "BN(20,0.70)"}]
```



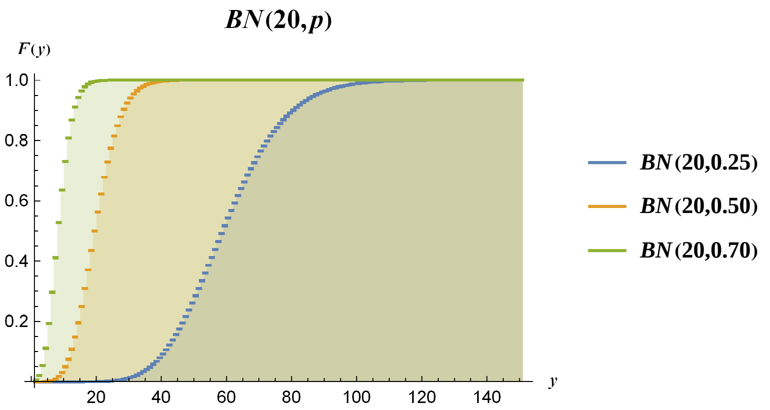
- función de distribución acumulada

```
CDF[NegativeBinomialDistribution[r, p], y] // TraditionalForm
```

$$\begin{cases} I_p(r, \lfloor y \rfloor + 1) & y \geq 0 \\ 0 & \text{True} \end{cases}$$

- representación gráfica de la función de distribución para $r = 20$ y diferentes valores de p

```
DiscretePlot[
 Table[CDF[NegativeBinomialDistribution[20, p], k], {p, {0.25, 0.5, 0.7}}] // Evaluate,
 {k, 150}, ExtentSize → Right,
 PlotLabel → "BN(20,p)", AxesLabel → {HoldForm["y"], HoldForm["F(y)"]},
 PlotLegends → {"BN(20,0.25)", "BN(20,0.50)", "BN(20,0.70)"}]
```



- valor esperado

```
Mean[NegativeBinomialDistribution[r, p]] // TraditionalForm
```

$$\frac{(1-p)\,r}{p}$$

```
Expectation[y, y ≈ NegativeBinomialDistribution[r, p]] // TraditionalForm
```

$$-\frac{(p-1)\,r}{p}$$

- varianza

```
Variance[BinomialDistribution[n, p]] // TraditionalForm
```

$$n\,(1-p)\,p$$

Ejemplo. Se considera el experimento aleatorio consistente en lanzar un dado equilibrado y contar el número de veces que no sale el número seis hasta que dicho número aparece dos veces.

- variable aleatoria Y: "*número de veces que no sale el 6 al lanzar un dado equilibrado hasta que aparece dos veces*"
- espacio muestral: $\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, \ \ldots\}$
- función de masa de probabilidad

```
distBN = NegativeBinomialDistribution[2, 1/6];
```

```
PDF[distBN, y] // TraditionalForm
```

$$\begin{cases} 5^y \times 6^{-y-2}\,(y+1) & y \geq 0 \\ 0 & \text{True} \end{cases}$$

- distribución de probabilidad de la variable aleatoria Y

```
valoresBN = Table[i, {i, 0, 15}]; N[PDF[distBN, valoresBN], 3]
```

```
{0.0278, 0.0463, 0.0579, 0.0643, 0.0670, 0.0670, 0.0651,
 0.0620, 0.0581, 0.0538, 0.0493, 0.0449, 0.0405, 0.0363, 0.0325, 0.0288}
```

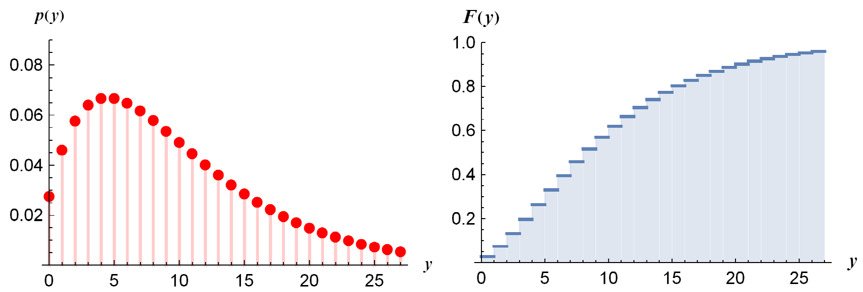
- función de distribución acumulada

```
CDF[distBN, y] // TraditionalForm
```

$$\begin{cases} I_{\frac{1}{6}}(2, \lfloor y \rfloor + 1) & y \geq 0 \\ 0 & \text{True} \end{cases}$$

- representación gráfica

```
Grid[{{
   DiscretePlot[PDF[distBN, y], {y, 0, 27}, AxesLabel → {HoldForm["y"], HoldForm["p(y)"]},
    PlotMarkers → Automatic, PlotStyle → Red, PlotRange → {0, 0.09}],
   Plot[CDF[distBN, y], {y, 0, 27}, Filling → Axis, PlotRange → {0, 1},
    AxesLabel → {HoldForm["y"], HoldForm["F(y)"]}]}}]
```



- cálculo de probabilidades
  - $P(Y = 2)$

```
PDF[distBN, 2] // N
```

```
0.0578704
```

  - $P(Y \leq 2)$

```
CDF[distBN, 2] // N
```

```
0.131944
```

Geométrica

Es un caso particular de la binomial negativa cuando $r = 1$. Es decir, las pruebas de Bernoulli se realizan hasta que se observa el primer éxito.

La variable aleatoria Y:*"número de fracasos hasta la obtención del primer éxito"* sigue un modelo de distribución geométrica con parámetro p.

Notación: $\boxed{Y \sim G(p)}$

  - **GeometricDistribution[p]**. Representa una distribución geométrica de parámetro p.

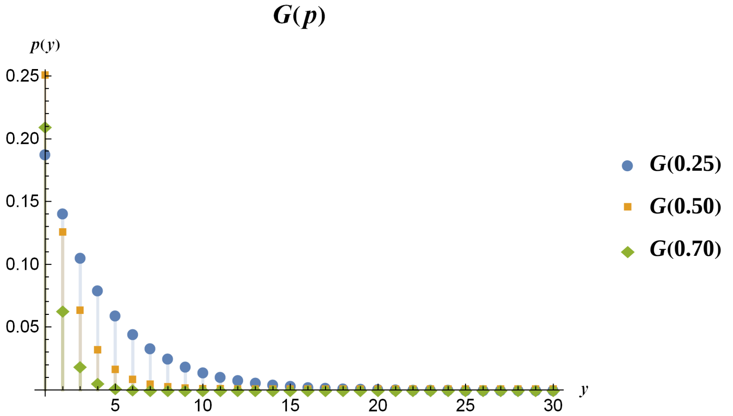
- función de masa de probabilidad

```
PDF[GeometricDistribution[p], y] // TraditionalForm
```

$$\begin{cases} p(1-p)^y & y \geq 0 \\ 0 & \text{True} \end{cases}$$

- representación gráfica de la función de masa para diferentes valores de p

```
DiscretePlot[Table[PDF[GeometricDistribution[p], k], {p, {0.25, 0.5, 0.7}}] // Evaluate,
 {k, 30}, PlotRange → All, PlotMarkers → Automatic, PlotLabel → "G(p)",
 AxesLabel → {HoldForm["y"], HoldForm["p(y)"]}, PlotLegends → {"G(0.25)", "G(0.50)", "G(0.70)"}]
```



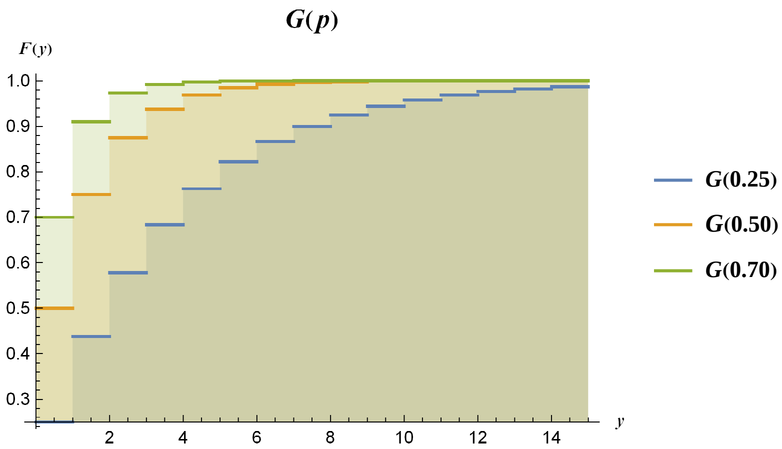
- función de distribución acumulada

```
CDF[GeometricDistribution[p], y] // TraditionalForm
```

$$\begin{cases} 1-(1-p)^{\lfloor y\rfloor+1} & y \geq 0 \\ 0 & \text{True} \end{cases}$$

- representación gráfica de la función de distribución acumulada para diferentes valores de p

```
DiscretePlot[Table[CDF[GeometricDistribution[p], k], {p, {0.25, 0.5, 0.7}}] // Evaluate,
 {k, 0, 14}, ExtentSize → Right, PlotLabel → "G(p)",
 AxesLabel → {HoldForm["y"], HoldForm["F(y)"]}, PlotLegends → {"G(0.25)", "G(0.50)", "G(0.70)"}]
```



- valor esperado

```
Mean[GeometricDistribution[p]] // TraditionalForm
```

$$\frac{1}{p} - 1$$

```
Expectation[y, y ≈ GeometricDistribution[p]] // TraditionalForm
```

$$\frac{1-p}{p}$$

- varianza

```
Variance[GeometricDistribution[p]] // TraditionalForm
```

$$\frac{1-p}{p^2}$$

Ejemplo. Se considera el experimento aleatorio consistente en lanzar un dado equilibrado y contar el número de veces que no sale el número seis hasta que dicho número aparece por primera vez.

- variable aleatoria Y : "*número de veces que no sale el 6 al lanzar un dado equilibrado hasta que aparece por primera vez*"
- espacio muestral: $\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, \ldots\}$
- función de masa de probabilidad

```
distG = GeometricDistribution[1/6];
```

```
PDF[distG, y] // TraditionalForm
```

$$\begin{cases} 5^y \times 6^{-y-1} & y \geq 0 \\ 0 & \text{True} \end{cases}$$

  - distribución de probabilidad de la variable aleatoria Y

```
valoresG = Table[i, {i, 0, 15}]; N[PDF[distG, valoresG], 3]
```

```
{0.167, 0.139, 0.116, 0.0965, 0.0804, 0.0670, 0.0558,
 0.0465, 0.0388, 0.0323, 0.0269, 0.0224, 0.0187, 0.0156, 0.0130, 0.0108}
```

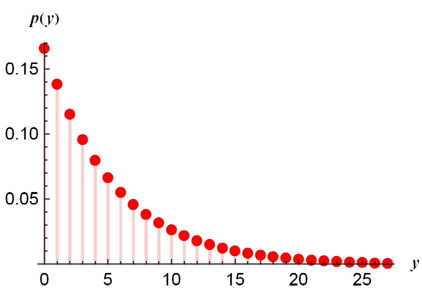
- función de distribución acumulada

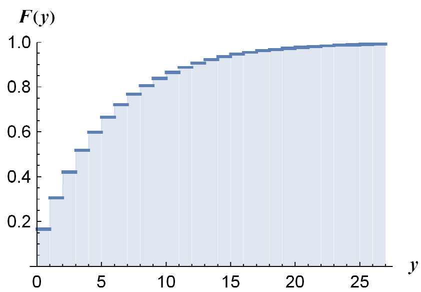
```
CDF[distG, y] // TraditionalForm
```

$$\begin{cases} 1 - \left(\frac{6}{5}\right)^{-\lfloor y \rfloor - 1} & y \geq 0 \\ 0 & \text{True} \end{cases}$$

- representación gráfica

```
Grid[{{
   DiscretePlot[PDF[distG, y], {y, 0, 27}, AxesLabel → {HoldForm["y"], HoldForm["p(y)"]},
    PlotMarkers → Automatic, PlotStyle → Red, PlotRange → {0, 0.17}],
   Plot[CDF[distG, y], {y, 0, 27}, Filling → Axis, PlotRange → {0, 1},
    AxesLabel → {HoldForm["y"], HoldForm["F(y)"]}]}}]
```





- cálculo de probabilidades
  - $P(Y = 2)$

```
PDF[distG, 2] // N
```

```
0.115741
```

- $P(Y \leq 2)$

```
CDF[distG, 2] // N
```

```
0.421296
```

Hipergeométrica

Sea un experimento consistente en seleccionar de forma aleatoria n elementos, sin reemplazamiento, de un conjunto de N elementos, catalogados como éxitos y fracasos. En el conjunto hay r éxitos y, por tanto, $N - r$ fracasos.

La variable aleatoria X:*"número de éxitos entre los n elementos extraídos"* sigue un modelo de distribución hipergeométrico con parámetros N, n y r.

Notación: $\boxed{X \sim H\left(N, n, \frac{r}{N}\right)}$

- **HypergeometricDistribution[*n*,*r*,*N*]**. Representa una distribución hipergeométrica en la que se extrae, sin reemplazamiento, de un conjunto de N elementos una muestra de tamaño n con un número inicial r de éxitos en el conjunto.

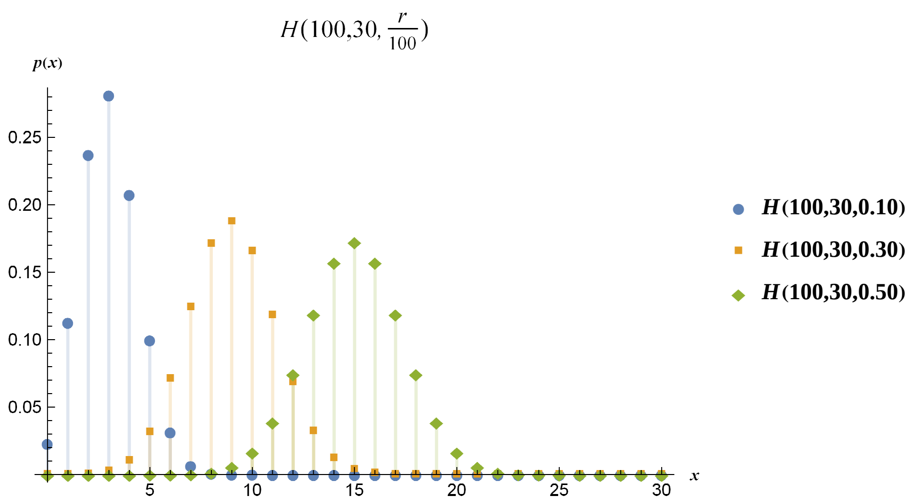
- función de masa de probabilidad

```
PDF[HypergeometricDistribution[n, r, N], x] // TraditionalForm
```

$$\begin{cases} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} & 0 \leq x \leq n \land n - N + r \leq x \leq n \land 0 \leq x \leq r \land n - N + r \leq x \leq r \\ 0 & \text{True} \end{cases}$$

- representación gráfica de la función de masa para $N = 100$, $n = 30$ y diferentes valores de r

```
DiscretePlot[Table[PDF[ HypergeometricDistribution[30, r, 100], k], {r, {10, 30, 50}}] //
  Evaluate, {k, 0, 30}, PlotRange → All, PlotMarkers → Automatic,
 PlotLabel → "H(100,30, r/100)", AxesLabel → {HoldForm["x"], HoldForm["p(x)"]},
 PlotLegends → {"H(100,30,0.10)", "H(100,30,0.30)", "H(100,30,0.50)"}]
```



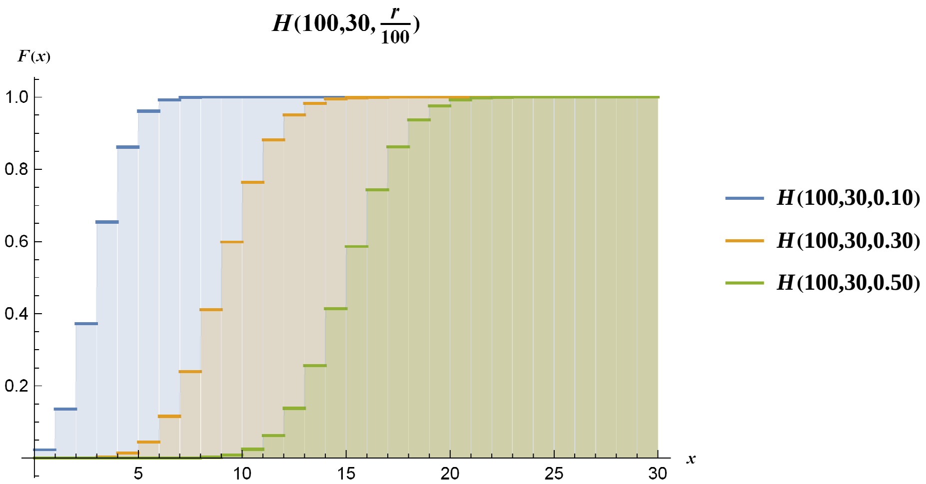
- función de distribución acumulada

```
CDF[HypergeometricDistribution[n, r, N], x] // TraditionalForm
```

$$\begin{cases} 1-\left(r!\,(N-r)!\,{}_3\tilde{F}_2(1,\,-n+\lfloor x\rfloor+1,\,-r+\lfloor x\rfloor+1;\,\lfloor x\rfloor+2,\,-n+N-r+\lfloor x\rfloor+2;\,1)\right)\Big/\left(\binom{N}{n}(-\lfloor x\rfloor+n-1)!\,(-\lfloor x\rfloor+r-1)!\right) \\ 1 \end{cases}$$

- representación gráfica de la función de masa para $N = 100$, $n = 30$ y diferentes valores de r

```
Plot[Table[CDF[ HypergeometricDistribution[30, r, 100], k], {r, {10, 30, 50}}] // Evaluate,
 {k, 0, 30}, Filling → Axis, PlotRange → All, PlotLabel → "H(100,30, r/100)",
 AxesLabel → {HoldForm["x"], HoldForm["F(x)"]},
 PlotLegends → {"H(100,30,0.10)", "H(100,30,0.30)", "H(100,30,0.50)"}]
```



- valor esperado

```
Mean[HypergeometricDistribution[n, r, N]] // TraditionalForm
```

$$\frac{n\,r}{N}$$

```
Expectation[x, x ≈ HypergeometricDistribution[n, r, N]] // TraditionalForm
```

$$\frac{n\,r}{N}$$

- varianza

```
Variance[HypergeometricDistribution[n, r, N]] // TraditionalForm
```

$$\frac{n\,r\,(N-n)\left(1-\frac{r}{N}\right)}{(N-1)\,N}$$

Ejemplo. Una urna opaca contiene 6 bolas blancas y 8 negras. Se extraen 5 bolas sin reemplazamiento y se cuenta el número de bolas blancas.

- variable aleatoria X : "*número de bolas blancas entre las 5 extraídas*"
- espacio muestral: $\Omega = \{0, 1, 2, 3, 4, 5\}$
- función de masa de probabilidad

```
distH = HypergeometricDistribution[5, 6, 14];
```

```
PDF[distH, x] // TraditionalForm
```

$$\begin{cases} \frac{\binom{6}{x}\binom{8}{5-x}}{2002} & 0 \le x \le 5 \\ 0 & \text{True} \end{cases}$$

- distribución de probabilidad de la variable aleatoria X

```
valoresH = Table[i, {i, 0, 5}]; PDF[distH, valoresH] // N
```

{0.027972, 0.20979, 0.41958, 0.27972, 0.0599401, 0.002997}

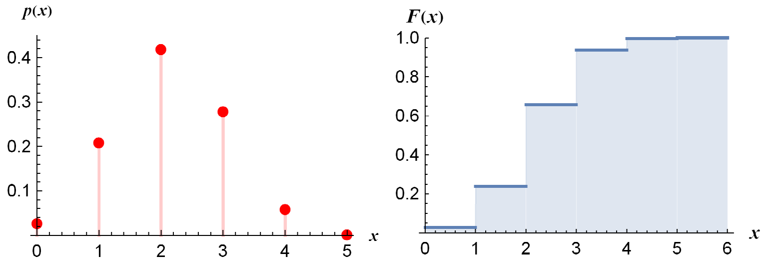
- función de distribución acumulada

```
CDF[distH, x] // TraditionalForm
```

$$\begin{cases} 1 - \left(2\,073\,600\; {}_3\tilde{F}_2(1, \lfloor x \rfloor - 4, \lfloor x \rfloor - 5; \lfloor x \rfloor + 2, \lfloor x \rfloor + 5; 1)\right) \Big/ (143\,(4 - \lfloor x \rfloor)!\,(5 - \lfloor x \rfloor)!) & 0 \le x < 5 \\ 1 & x \ge 5 \end{cases}$$

- representación gráfica

```
Grid[{{
   DiscretePlot[PDF[distH, x], {x, 0, 5}, AxesLabel → {HoldForm["x"], HoldForm["p(x)"]},
    PlotMarkers → Automatic, PlotStyle → Red, PlotRange → {0, 0.45}],
   Plot[CDF[distH, x], {x, 0, 6}, Filling → Axis, PlotRange → {0, 1},
    AxesLabel → {HoldForm["x"], HoldForm["F(x)"]}]}}]
```



- cálculo de probabilidades
  - $P(X = 2)$

```
PDF[distH, 2] // N
```

0.41958

  - $P(X \le 2)$

```
CDF[distH, 2] // N
```

0.657343

Poisson

Proporciona un modelo para la frecuencia relativa del número de sucesos que ocurren en una unidad de tiempo, área, longitud, volumen, etc.

Suele utilizarse en el cálculo de la probabilidad de ocurrencia de sucesos considerados "raros".

La variable aleatoria X:*“número de ocurrencias de un suceso en una unidad de medida”* sigue un modelo de distribución de Poisson con parámetro λ.

Notación: $\boxed{X \sim \mathcal{P}(\lambda)}$

- **PoissonDistribution[λ]**. Representa una distribución de Poisson de parámetro λ.

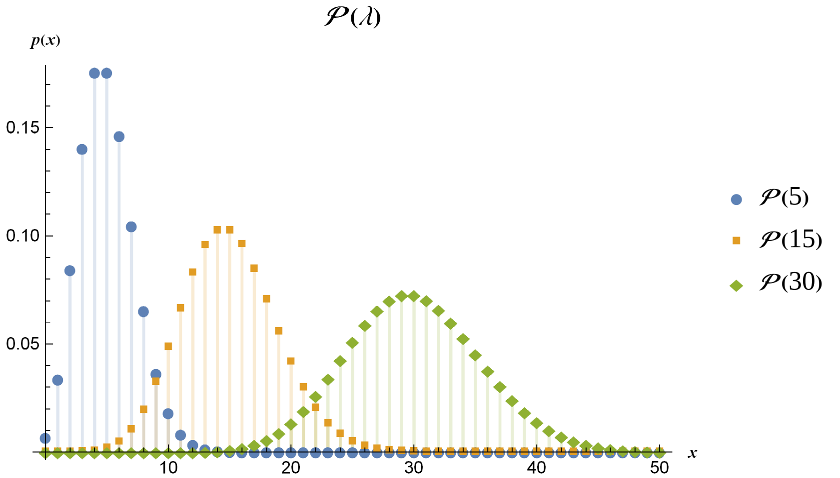
- función de masa de probabilidad

```
PDF[PoissonDistribution[λ], x] // TraditionalForm
```

$$\begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & x \geq 0 \\ 0 & \text{True} \end{cases}$$

- representación gráfica de la función de masa para diferentes valores de λ

```
DiscretePlot[Table[PDF[PoissonDistribution[λ], k], {λ, {5, 15, 30}}] // Evaluate,
 {k, 0, 50}, PlotRange → All, PlotMarkers → Automatic, PlotLabel → "𝒫(λ)",
 AxesLabel → {HoldForm["x"], HoldForm["p(x)"]}, PlotLegends → {"𝒫(5)", "𝒫(15)", "𝒫(30)"}]
```



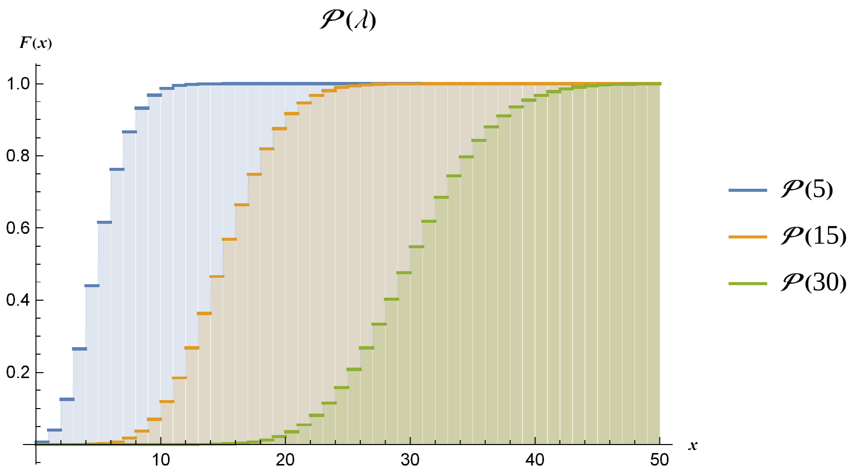
- función de distribución acumulada

```
CDF[PoissonDistribution[λ], x] // TraditionalForm
```

$$\begin{cases} Q(\lfloor x \rfloor + 1, \lambda) & x \geq 0 \\ 0 & \text{True} \end{cases}$$

- representación gráfica de la función de distribución acumulada para diferentes valores de λ

```
Plot[Table[CDF[PoissonDistribution[λ], k], {λ, {5, 15, 30}}] // Evaluate,
 {k, 0, 50}, Filling → Axis, PlotRange → All, PlotLabel → "𝒫(λ)",
 AxesLabel → {HoldForm["x"], HoldForm["F(x)"]}, PlotLegends → {"𝒫(5)", "𝒫(15)", "𝒫(30)"}]
```



- valor esperado

```
Mean[PoissonDistribution[λ]] // TraditionalForm
```

λ

```
Expectation[x, x ≈ PoissonDistribution[λ]] // TraditionalForm
```

λ

- varianza

```
Variance[PoissonDistribution[λ]] // TraditionalForm
```

λ

Ejemplo. El número de correos electrónicos recibidos en una determinada dirección sigue una distribución de Poisson con un promedio de 6 correos a la hora.

- variable aleatoria X : "*número de correos que llegan a la hora a la dirección*"
- espacio muestral: $\Omega = \{0, 1, 2, 3, \ldots\}$
- función de masa de probabilidad

```
distP = PoissonDistribution[6];
```

```
PDF[distP, x] // TraditionalForm
```

$$\begin{cases} \frac{6^x}{e^6\, x!} & x \geq 0 \\ 0 & \text{True} \end{cases}$$

  - distribución de probabilidad de la variable aleatoria X

```
valoresP = Table[i, {i, 0, 15}]; N[PDF[distP, valoresP], 3]
```

```
{0.00248, 0.0149, 0.0446, 0.0892, 0.134, 0.161, 0.161, 0.138,
 0.103, 0.0688, 0.0413, 0.0225, 0.0113, 0.00520, 0.00223, 0.000891}
```

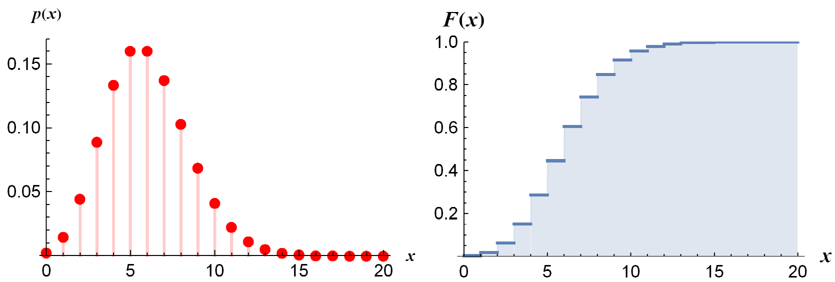
- función de distribución acumulada

```
CDF[distP, x] // TraditionalForm
```

$$\begin{cases} Q(\lfloor x \rfloor + 1,\, 6) & x \geq 0 \\ 0 & \text{True} \end{cases}$$

- representación gráfica

```
Grid[{{
   DiscretePlot[PDF[distP, x], {x, 0, 20}, AxesLabel → {HoldForm["x"], HoldForm["p(x)"]},
    PlotMarkers → Automatic, PlotStyle → Red, PlotRange → {0, 0.17}],
   Plot[CDF[distP, x], {x, 0, 20}, Filling → Axis, PlotRange → {0, 1},
    AxesLabel → {HoldForm["x"], HoldForm["F(x)"]}]}}]
```



- cálculo de probabilidades
  - $P(X = 2)$

```
PDF[distP, 2] // N
```

```
0.0446175
```

  - $P(X \leq 2)$

```
CDF[distP, 2] // N
```

```
0.0619688
```