

Estadística Descriptiva e Introducción a la Probabilidad

Juan Antonio Maldonado Jurado

Departamento de Estadística e Investigación Operativa
Universidad de Granada



Doble Grado en Ingeniería Informática y Matemáticas

Tema 6. Algunos modelos de distribuciones discretas

Introducción

Distribución degenerada

Distribución uniforme discreta

Distribución de Bernoulli

Distribución Binomial

Distribución Binomial Negativa

Distribución Hipergeométrica

Distribución de Poisson

Introducción: Modelos de distribuciones y sus aplicaciones

Una vez expuesta la teoría general sobre variables aleatorias y sus distribuciones de probabilidad, vamos a describir algunas distribuciones particulares que han demostrado, empíricamente, ser modelos apropiados para situaciones que ocurren en la vida real. A pesar de ello tales distribuciones presentan un carácter teórico en el sentido de que sus funciones de probabilidad o de densidad se deducen matemáticamente en base a ciertas hipótesis que se suponen válidas para los fenómenos aleatorios.

La elección de una distribución de probabilidad para representar un fenómeno de interés práctico debe estar motivada tanto por la comprensión de la naturaleza del fenómeno en sí, como por la posible verificación de la distribución seleccionada a través de la evidencia empírica. En todo momento debe evitarse aceptar de manera tácita una determinada distribución de probabilidad como modelo de un problema práctico.

Una distribución de probabilidad está caracterizada, de forma general, por una o más cantidades que reciben el nombre de parámetros de la distribución. Un parámetro puede tomar cualquier valor de un conjunto dado y, en ese sentido, se define una familia de distribuciones de probabilidad que tendrán la misma función genérica de probabilidad o función de densidad.



Distribución degenerada

La distribución discreta más sencilla es la correspondiente a una variable aleatoria degenerada o constante, es decir, la asociada a un experimento aleatorio que da lugar siempre al mismo resultado. Por tanto, dicha variable aleatoria tomará un único valor c.

Su función masa de probabilidad es

$$P[X = x] = \begin{cases} 1 & x = c \\ 0 & x \neq c \end{cases}$$

Algunas de sus características son:

Función de distribución:

$$F(x) = P[X \le x] = \begin{cases} 0 & x < c \\ 1 & x \ge c \end{cases}$$

Distribución degenerada

Momentos no centrados:

$$m_k = E[X^k] = c^k P[X = c] = c^k, \qquad k = 1, 2, \cdots$$

Y, en particular, la MEDIA

$$E[X] = c$$
.

Momentos centrados:

$$\mu_k = E[(X-c)^k] = 0, \qquad k = 1, 2, \cdots$$

Y, en particular, la VARIANZA

$$Var[X] = 0.$$

Esta propiedad caracteriza a las distribuciones degeneradas; es decir, una variable aleatoria tiene varianza cero si y solamente si es degenerada en un punto (Propiedades de la varianza).

► Función generatriz de momentos:

$$M(t) = E[e^{tX}] = e^{tc} \quad \forall t \in \mathbb{R}.$$



Distribución uniforme discreta

Esta es la distribución de probabilidad de una variable aleatoria d
valores que son equiprobables y se utiliza para modelizar variables
aleatorios que tienen un número finito de posibles resultados equi
probabilidad es Esta es la distribución de probabilidad de una variable aleatoria discreta que toma un número finito de valores que son equiprobables y se utiliza para modelizar variables aleatorias asociadas a experimentos aleatorios que tienen un número finito de posibles resultados equiprobables. Su función masa de

$$P[X = x_i] = \begin{cases} \frac{1}{n} & i = 1, 2, \dots, n \\ 0 & \text{en el resto} \end{cases}$$

Se dice entonces que la variable aleatoria X se distribuye uniformemente sobre los puntos x_1, x_2, \dots, x_n y se notará $X \sim U(x_1, x_2, \dots, x_n)$.

Algunas de sus características son:

► Función de distribución:

$$F(x) = \frac{1}{n}(\text{Número de valores } x_i \le x) = \begin{cases} 0 & \text{si } x < x_1 \\ \frac{1}{n} & \text{si } x_1 \le x < x_2 \end{cases}$$

$$\vdots$$

$$\frac{n-1}{n} & \text{si } x_{n-1} \le x < x_n \\ 1 & \text{si } x \ge x_n \end{cases}$$



Distribución uniforme discreta

Momentos no centrados:

$$m_k = E[X^k] = \frac{1}{n} \sum_{i=1}^n x_i^k, \qquad k = 1, 2, \cdots$$

Y, en particular, la MEDIA

$$E[X] = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}.$$

Momentos centrados:

$$\mu_k = E[(X - EX)^k] = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, \qquad k = 1, 2, \dots$$

Y, en particular, la VARIANZA

$$Var[X] = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

► Función generatriz de momentos:

$$M(t) = E[e^{tX}] = \frac{1}{n} \sum_{i=1}^{n} e^{tx_i} \quad \forall t \in \mathbb{R}.$$

Distribución uniforme discreta

En el caso particular $x_i = i, i = 1, 2, ..., n$

$$E[X] = \frac{1}{n} \sum_{i=1}^{n} i = \frac{1}{n} \frac{n(n+1)}{2} = \frac{n+1}{2}$$

$$E[X^2] = \frac{1}{n} \sum_{i=1}^{n} i^2 = \frac{1}{n} \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

$$Var[X] = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{n^2 - 1}{12}$$

Distribución de Bernoulli

Supongamos un experimento aleatorio que da lugar, únicamente, a dos posibles resultados que son mutuamente excluyentes y exhaustivos, que denotaremos por éxito (E), que será el suceso objeto de estudio, y fracaso (F), que es el complementario de E.

A este tipo de experimentos aleatorios se les llama **experimentos o pruebas de Bernoulli**. Asociado a un experimento o prueba de Bernoulli y a su correspondiente espacio muestral $\Omega = \{E, F\}$, se define la variable aleatoria con distribución de Bernoulli como

$$X = \begin{cases} 1 & \text{si ocurre el suceso } E \\ 0 & \text{si no ocurre el suceso } E \text{ (ocurre } F \text{)} \end{cases}$$

Si se denota por p a la probabilidad del suceso éxito (E), su función masa de probabilidad es

$$P[X = 1] = p$$

 $P[X = 0] = 1 - p$

o bien,
$$P[X=x]=p^{x}(1-p)^{1-x}; \quad x=0,1; \quad 0< p<1$$
 y se notará $X\sim B(1,p)$. $\mathbb{E}[X]=p, \ \mathbb{E}[X^2]=p, \ \mathrm{Var}[X]=\mathbb{E}[X^2]-(\mathbb{E}[X])^2=p-p^2=p(1-p)$

Distribución binomial

Se dice que una variable aleatoria X sigue una **distribución binomial** de parámetros n y p, $n \in \mathbb{N}, p \in (0,1)$ si modela el número de éxitos en n repeticiones independientes de un ensayo de Bernoulli con probabilidad p de éxito, manteniéndose ésta constante en las n repeticiones del experimento. Su función masa de probabilidad es

$$P[X = x] = \binom{n}{x} p^{x} (1-p)^{n-x}, \qquad x = 0, 1, \dots, n.$$

Se notará como $X \sim B(n, p)$.

$$\mathsf{E}[X] = np; \ \mathsf{Var}[X] = np(1-p)$$

Propiedad de simetría.- Si $X \sim B(n,p)$, entonces la variable aleatoria que contabiliza el número de fracasos, $Y = n - X \sim B(n,1-p)$ y, además

$$P[X = x] = P[Y = n - x]$$



Distribución binomial

Modelos probabilísticos discretos

Distribución Binomial

Distribución Binomial B(n,p)

Valores: $P\{X = k\} = {n \choose k} p^k (1-p)^{k+1}; k = 0,1,2,3,...,n$

	k p	0.81	0.05	0.19	8.15	6.20	6.25	0.30	113	9.35	8.40	8.45	0.49	0.50
2		1901	1968	2.5 600	877295	6.6900	4.6024	6.400	0.888	3.42%	8.3804	43005	6.364	
	1	1000	1000	9.700	8.2594 8.3035	6.3000 6.3000	6.9926	6.4000 6.6000	0.000	9.420E	1.4900	8.4950 8.2020	6.2604	150
		0990	936N	9 7 290	0.5041	4.5100	6.4(79	0.000	0.290	5 123K	8.1906 9.2986	8.7004	6.7001	
3	:	0.000	0.00%	9.750	8,3201	1300	5 (21)	1400	0.000	24636	1.4320	1.004	6.000	0.05
		1000	93671	1007	0.3674	1.1900	5.7006	0.790	0.150	9.250	1 2500	83041	6,0004	
	1	0.000	0.0001	1000	1.30N	£ 1080	6.0100	6.600	0.0079	10639	1,2660	B3011	6.1176	0.09
4		1908		0.0001	9.5291	1.4090	0.0054	0.001	0.005		8.1296	1.7645		
			01466		0.0675		6.2504							0.0%
	3	0.000	0.000	0.00%	0.24 81	6.1056	E 4906	0.076	0.088	0.1115	8.1636	B 2000	6,3900	0.30
		0.000	9-3000	9-0001	93000	£30mb	6.9064	6.084	04105	1-1-69	8.3054	0.1410	6.1626	
•		0.80	0.758	2.5905	14427	6,3077	6,300	6 991	0.001	3.1168	1309	8-38GE	6.6043	(4)
		0.000	9.2686	9.5360	8.3045	6.4006	63006	6.960	0.880	9.5434	9.2548	R3066	6.982	0.88
	1	0.000	9.00%	0.0039	8.1863	0.2048	0.2037	0.3007	0.030	3.3304	1.3400	8.3369 8.2057	6,3795	
	:	0000	0.0000	9-900H	0.3052	6.3004	5.5079	0.006	0.998	2.1011	8.2904	8.2007 8.1526	5.7970	0.00
	:	0.000	1000	1000	83001	£ 3004	5.1040	0.004	0.000	1000	1.3000	8.1126 8.0005	6.000	0.00
	÷	0900	9.580	2.55 H	8,3571	1,7851	6.1790	0.1100	0.0679	1059	1.3467	8.3077	6.0076	
	1	0.00%	9201	1350	1.3061	6.3962	5.3500	0.005	0.00%	3 2607	1.1366	8.100g	6.1014	0.00
		000%	0.000	2-2694	5.1762	1,361	0.2500	0.000	0.130	2.220	1,251	1,2790	1,301	0.29
		0000	0.0001	0.000	9.3009	6,000	5.5000	0.000	0000	2-2611	5.1262	8.1901	0.239	0.29
,			2-890	24302	11204	1000	4.134	6.0004	0000	2-3466	8.3044		4.6000	
	1	0.0009	0.259	0.1528	8.3966	6.3890	6,3075	6.347	0.2949	3.7541	5.7006	8.3642	6.8906	0.00
	3	0.000	0.0400	0.1361	0.2067	6.20%	113111	THE	0.30%	3 2005	11,2013	11,2140	6.0790	0.30
	9	0.000	0.008	9-0099	8.3647	6.1547	6.1930	0.3399	0.29m	9.20%	1 2901	8.2546	0.2790	0.379
	4	0.000	0.0002	9.0036	9.0004	63067	6.1677	6.007	0.030	9.3442	1,1965	9.2586	0.3879	0.25
	5	1000	0.000	1000	0.0042	6.3543	1.0275	6.650	0.00M	10400	83654	8.16.63	6.593	0.394
		0000	0.000	1000	0.3001 0.3000	£ 1006	£3063	0.000	0.000	2-300H	8.3KT2	8 3020 8 3067	E 2004	600
		1000	188	1400	12521	6,979	6.1004	1007	0.000	1019	1,3041	8.3084	5.0040	
		0.634	0.170	0.000	0.3567	6 1005	6.7520	0.007	0.190	3 1570	1 3000	1.1541	4.005	
	9	0008	0.000		12376	1,7900	52115	0.360	0.750	3.2507	1,2080	1.1001	6.7783	0.31
		1000	0.0054	0.0621	0.0654	E 365	8.2006	0.294	0.356	9.25m	12567	1 2504	4.055	024
		0.000	0.0004	0.0066	5.2085	6.1400	5.7965	0.581	0.150	2.7675	1.2152	1.2027	0,270	
		0.000	0.000	0.0000	8.3000	6.3000	6.1000	0.0000	0.0002	1000	8.3007	83007	6,8003	0.00
•		0965	9.550	3.56N	9.25.86	6.1(4)	6.9054	6.004	0430	1-100	8.3404	8.3045	6.0005	
		0.000	0.295	5.36%	8.3659	8.3000	6.2054	0.700	0.1175	3.5004	8.3009	8.3009	6.8000	0.01
	2	0.00K	0.0038	9 (1)(2)	9.2567	63000	63006	4308	0.380	9.2402	8.1643	8,10.00	6.6716	4.6%
	9	1000	9.007	1046	9.7069 9.3063	C981	6.2504	0.200	0.255	9.21W 9.21W	1.2504	1215	6.1/79	0.39
	;	0.000	0.000	0-00W	9.3062	£.1961	E 1066	6 676	0.308	9.7761	8.7692	9.2900 9.2529	0.2006	0.34
		1000	1200	1000	1300	6.000	6.1067	1670	0.000	20636	1350	8.1960	6.792	0.39
		1000	1000	1-3001	8 3000	8.3006	6.10e)	5.009	0000	P-2006	83613	8.5807	6.6005	
		0.000	1.000	1000	1.3000	6.1000	6.0001	0.0004	0.000	1000	1,308	E3003	6.6153	660
		1.000	0.000	1000	8,3000	6.1000	6,0000	6.000	0.000	3-3001	0.3000	8,3000	6.6000	100
4		0.000	1367	1360	1.1961	6.009	1,000	1.000	0005	1336	1386	1.835	6.6043	- (40
		0.00 K	0.5451	0.56N	0.5434	6.7564	6:007	0.1018	0.007	MARK	0.5405	8.9007	6.0114	4.00
	2	0.000	0.0066	9.1687	0.2558	6,3000	0.2696	0.3306	0.180	9.1757	1.1208	B 3053	6.000	0.00
		0-06H	0.0425	0-067N	8.1264	6.2543	6.2504	4.000	0-36m	P-2553	9.2950	8.1005	6.1067	
	4	0.000	0.000	0.010	0.0601	63881	6.1600	6,300	OIIW	9.2837	8.2904	12014	0.2100	6.38
		0-000	0.0001	1006	2906.6	6.1054	1.1594	4.609	0.186	3.1586	1.2967	8.2540	4.3454	0.34
		0.000	0.0000	0.0001	9.30CI	6.3000	6.6162	0.6068	0.0000	0.0000	0.1110	8.1000	6.1000	0.38
		0-000	9-3000	9-0008	9.3001	K.1006	6.1064	6.6000	0496	9304	0.1425	8.3546	6:500	
		0.000	0.0000	0.0000	8.3000	8,3001	6.1004	0.0014	0.000	0.000	8.3006	1.1001	6.8089	000
		6-000	9-3000	2-3000	8.3004	6.3000	£3000	6-664	0.000	1-3005	8.3046	83643	6.1068	6.66



Consideremos un experimento aleatorio consistente en repeticiones independientes de ensayos de Bernoulli con probabilidad de éxito constante, hasta que aparezca el éxito r-ésimo. Es decir, en lugar de fijar el número de ensayos y observar el número de éxitos en esas realizaciones, se repiten las realizaciones hasta obtener un número determinado de éxitos y contabilizamos los fracasos.

Definimos la variable aleatoria con **distribución binomial negativa** como aquella que modela el número de fracasos antes de que aparezca el éxito r-ésimo.

La variable aleatoria X puede tomar los valores $x=0,1,2,\ldots$ y r será un entero positivo; es decir, $r=1,2,\ldots$

la función masa de probabilidad de esta variable aleatoria es

$$P[X = x] = {x + r - 1 \choose x} (1 - p)^x p^r, \qquad x = 0, 1, 2, \cdots \quad (r = 1, 2, \cdots; \quad 0$$

y se notará como $X \sim BN(r, p)$.

$$\mathsf{E}[X] = rac{r(1-p)}{p}; \quad \mathsf{Var}[X] = rac{r(1-p)}{p^2}$$

5.6 Distribución Hipergeométrica Supongamos una población de N individuos divididos individuos cada uno. Se elige una muestra de n indivio o simultáneamente). La variable aleatoria X que cont categoría en la muestra se dice que tiene distribución Supongamos una población de N individuos divididos en dos categorías de N_1 y $N_2 = N - N_1$ individuos cada uno. Se elige una muestra de n individuos de la población (sin reemplazamiento o simultáneamente). La variable aleatoria X que contabiliza el número de individuos de la 1^a categoría en la muestra se dice que tiene distribución hipergeométrica de parámetros N, N1 y n v se nota

$$X \sim H(N, N_1, n),$$
 $n, N_1, N \in \mathbb{N} - \{0\}, n, N_1 \le N$

La función masa de probabilidad de esta variable aleatoria es

$$\mathsf{P}[X=x] = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}} \qquad \mathsf{máx}(0,n-(N-N_1)) \leq x \leq \mathsf{min}(n,N_1)$$

$$\mathsf{E}[X] = \mathit{np}; \ \mathsf{Var}[X] = \frac{\mathit{n}(\mathit{N}-\mathit{n})\mathit{N}_1(\mathit{N}-\mathit{N}_1)}{\mathit{N}^2(\mathit{N}-1)}$$

Distribución de Poisson

Representa el número de ocurrencias de un determinado suceso durante un periodo de tiempo fijo o en una región fija del espacio, cuando el número de ocurrencias sigue unas determinadas pautas:

- ► El número de ocurrencias en un intervalo o región especificada debe ser independiente del número de ocurrencias en cualquier otro intervalo o región.
- Si se considera un intervalo de tiempo muy pequeño (o una región muy pequeña), la probabilidad de una ocurrencia es proporcional a la longitud del intervalo (al volumen de la región) y la probabilidad de dos o más ocurrencias es prácticamente nula (despreciable).

Con estas consideraciones, si por X_t denotamos el número de ocurrencias del suceso en un intervalo de longitud t (región de volumen t), puede probarse que

$$P\{X_t = k\} = \frac{e^{-\lambda t} (\lambda t)^k}{k!}$$
 $k = 0, 1, 2, ..., \lambda > 0$

La colección de variables aleatorias $\{X_t;\ t\geq 0\}$ constituyen lo que se denomina un PROCESO DE POISSON y cada variable X_t se dice que tiene una distribución de Poisson de parámetro λt .

En particular, la variable aleatoria que específica el número de ocurrencias en un intervalo de longitud unidad (o región de volumen unidad), se dice que tiene una distribución de Poisson de parámetro λ y se nota $X \sim \mathcal{P}(\lambda)$. Su función masa de probabilidad es

$$P[X = x] = e^{-\lambda} \frac{\lambda^x}{x!}, \qquad x = 0, 1, \dots, \ \lambda > 0$$

$$E[X] = \lambda; \quad Var[X] = \lambda$$

67 Distribución de Poisson

Distribución de Poisson $P(\lambda)$

