DGI
2019

5.1. Distribuição Binomial [$\mathrm{B}(n, p)$]

5.1.1. Introdução

Os processos mais simples de recolha de dados consistem na ***contagem do número de vezes que um determinado acontecimento ocorre*** ao longo de uma série de experiências (ex. contagem de peças defeituosas num posto de inspecção).

Frequentemente tais experiências apresentam as seguintes características:

- A cada experiência tem apenas *dois resultados possíveis* (sucesso ou insucesso).
- A *probabilidade de ocorrência* de cada resultado mantém-se *inalterada* de experiência para experiência
- Os *resultados* associados a diferentes experiências são *independentes* (a probabilidade de sucesso p não é afectada pelo possível conhecimento dos resultados obtidos até então).

Experiências nestas condições dizem-se ***experiências de Bernoulli***.

Uma variável aleatória binomial Y representa o número de vezes que, no decurso de n experiências de Bernoulli, ocorre um dos resultados possíveis. A variável aleatória Y pode tomar os valores 0, 1, 2, ..., n.

$$Y \rightsquigarrow \mathrm{B}(n, p)$$

- Trata-se de uma distribuição muito utilizada em amostragem, em situações em que se conhece a dimensão da amostra e se sabe quantas vezes um acontecimento ocorreu.
- A Distribuição Binomial está para as distribuições discretas de probabilidade, assim como a distribuição normal está para as distribuições contínuas.

5.1.2. Função de Probabilidade

$$p(y) = \left(\begin{array}{c} n \\ y \end{array} \right) \times p^y \times q^{(n-y)} = C_y^n \times p^y \times q^{(n-y)}$$

$p(y)$ representa a probabilidade de, em n tentativas, se obterem y sucessos e $(n$ - $y)$ insucessos.

Exemplo

De uma linha de produção, retiram-se 3 peças (cada uma de um turno diferente). Admitindo que, para todos os turnos, a percentagem de peças defeituosas é de 10%, calcular a probabilidade de entre as três peças recolhidas haver 0, 1, 2 ou 3 peças defeituosas.

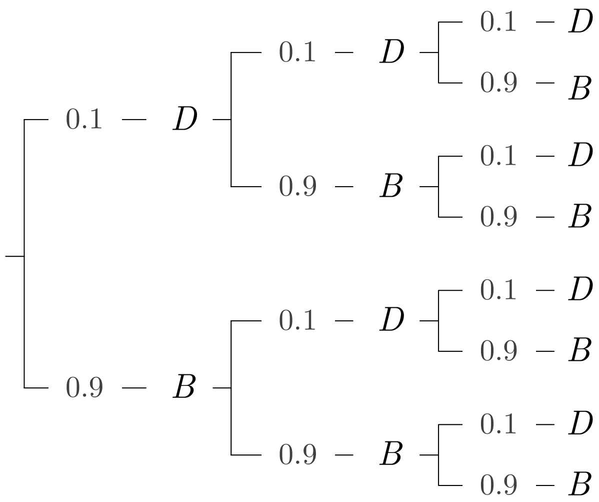
sendo: Y– número de peças defeituosas de entre as três seleccionadas

p – probabilidade de saída de peça defeituosa *(sucesso)*

q – probabilidade de saída de peça não defeituosa

DGI
2019

representação
dos acontecimentos
(diagrama em árvore)



Função de Probabilidade

$y=0$	$p(0)=1\times(0.1)^0\times(0.9)^3=0.7290$
$y=1$	$p(1)=3\times(0.1)^1\times(0.9)^2=0.2430$
$y=2$	$p(2)=3\times(0.1)^2\times(0.9)^1=0.0270$
$y=3$	$p(3)=1\times(0.1)^3\times(0.9)^0=0.0010$

$$p(y)=\mathrm{C}_y^n\times p^y\times q^{(n-y)}$$

$$\downarrow$$

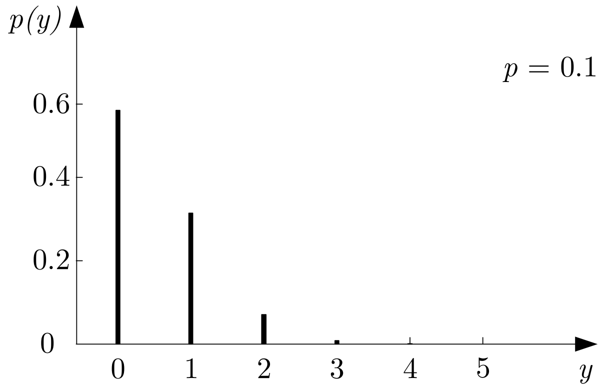
$$Y \rightsquigarrow B(n=3,p=0.1)$$

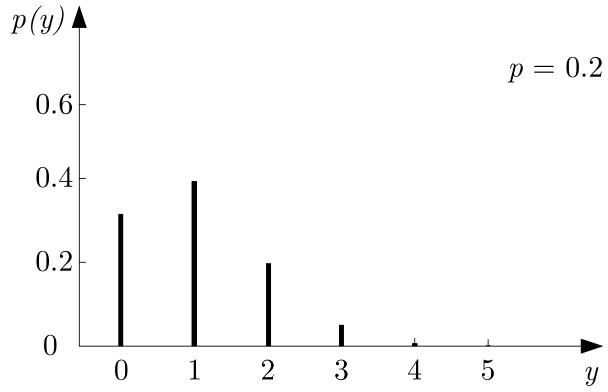
5.1.3. *Simetria*

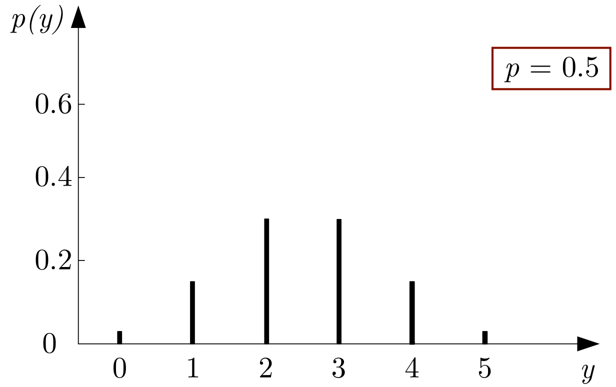
Considere-se a distribuição associada à variável aleatória Y:

$$Y \to B(n = 5,\ p)$$

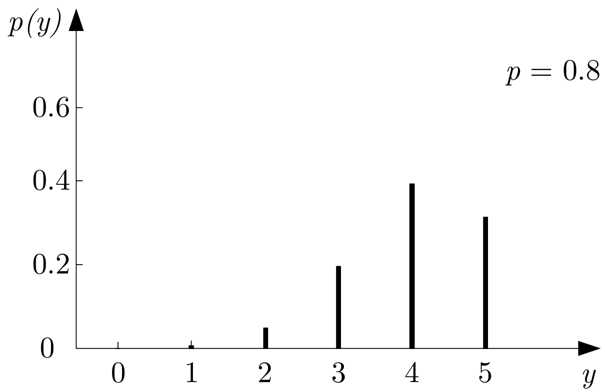
y	$p(y)$				
	$p = 0.1$	$p = 0.2$	$p = 0.5$	$p = 0.8$	$p = 0.9$
0	0.5905	0.3277	0.0313	0.0003	0
1	0.3281	0.4096	0.1563	0.0064	0.0005
2	0.0729	0.2048	0.3125	0.0512	0.0081
3	0.0081	0.0512	0.3125	0.2048	0.0729
4	0.0005	0.0064	0.1563	0.4096	0.3281
5	0	0.0003	0.0313	0.3277	0.5905
$\sum$	1	1	1	1	1

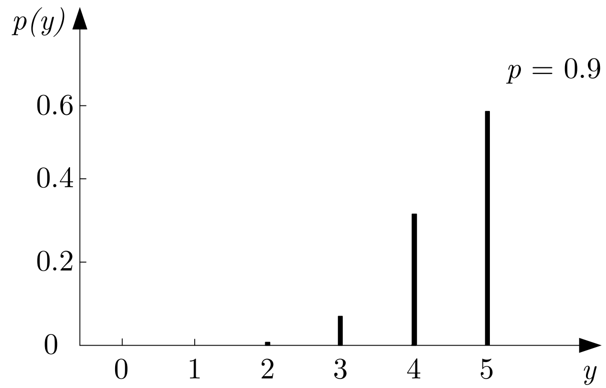






$p\,(y)$ é simétrica se $p = 0.5$





5.1.4. Parâmetros

Valor esperado	$\mu = n \times p$
Variância	$\sigma^2 = n \times p \times q$
Desvio Padrão	$\sigma = \sqrt{n \times p \times q}$

5.1.5. *Valores Tabelados*

Os valores tabelados correspondem às probabilidades

$$p(y) = \mathrm{C}_y^n \times p^y \times q^{(n-y)}$$

Exemplo

$Y \rightsquigarrow B(n=3,\ p=0.1)$

$n=2$

y		p	**0.05 (0.95)**	**0.10 (0.90)**	**0.15 (0.85)**	**0.20 (0.80)**	**0.25 (0.75)**	**0.30 (0.70)**	**0.35 (0.65)**	**0.40 (0.60)**	**0.45 (0.55)**	**0.50 (0.50)**
0	**(2)**		.9025	.8100	.7225	.6400	.5625	.4900	.4225	.3600	.3025	.2500
1	**(1)**		.0950	.1800	.2550	.3200	.3750	.4200	.4550	.4800	.4950	.5000
2	**(0)**		.0025	.0100	.0225	.0400	.0625	.0900	.1225	.1600	.2025	.2500

$\boxed{n=3}$

y		p	**0.05 (0.95)**	**0.10 (0.90)**	**0.15 (0.85)**	**0.20 (0.80)**	**0.25 (0.75)**	**0.30 (0.70)**	**0.35 (0.65)**	**0.40 (0.60)**	**0.45 (0.55)**	**0.50 (0.50)**
0	**(3)**		.8574	.7290	.6141	.5120	.4219	.3430	.2746	.2160	.1664	.1250
1	**(2)**		.1354	.2430	.3251	.3840	.4219	.4410	.4436	.4320	.4084	.3750
2	**(1)**		.0071	.0270	.0574	.0960	.1406	.1890	.2389	.2880	.3341	.3750
3	**(0)**		.0001	.0010	.0034	.0080	.0156	.0270	.0429	.0640	.0911	.1250

5.2. Distribuição Binomial Negativa $[\ \mathrm{BN}(r,p)\ ]$

5.2.1. *Introdução*

Esta distribuição permite descrever o comportamento de uma variável aleatória que, embora associada à *repetição de experiências de Bernoulli*, ***envolve um processo de contagem distinto daquele que é descrito pela distribuição Binomial***.

Numa *sequência infinita de experiências de Bernoulli*, a variável Y seguirá uma distribuição Binomial Negativa se representar ***o número de insucessos até ocorrer o r-ésimo sucesso (inclusive)***.

$$Y \rightsquigarrow \mathrm{BN}(r,p)$$

5.2.2. *Função de Probabilidade*

$$p(y) = \mathrm{C}_y^{y+r-1} \times p^r \times q^y$$

$p(y)$ representa a probabilidade de ocorrerem y insucessos até ocorrer o r-ésimo sucesso.

Definição da distribuição Binomial Negativa a partir da distribuição Binomial

Y – número de insucessos até ao r-ésimo sucesso

p – probabilidade de sucesso

Número total de tentativas $= Y + r$

$$\underbrace{\boxed{}\boxed{}\boxed{\;\cdots\;}\boxed{}\boxed{}}_{y+r-1} + \underset{\substack{1 \\ \text{sucesso}}}{\boxed{}}$$

$$\left(\begin{array}{cc} r-1 & \text{sucessos} \\ y & \text{insucessos} \end{array}\right)$$

$\rightarrow$

$$p(y) = \left(C_y^{y+r-1} \times p^{r-1} \times q^y\right) \times p$$

$$= C_y^{y+r-1} \times p^r \times q^y$$

5.2.3. Parâmetros

Valor esperado	$\mu = \dfrac{r \times q}{p}$
Variância	$\sigma^2 = \dfrac{r \times q}{p^2}$
Desvio Padrão	$\sigma = \sqrt{\dfrac{r \times q}{p^2}}$

5.2.4. Caso Particular - Distribuição Geométrica ($r = 1$) $[\ \mathrm{G}(p)\ ]$

Y representa o número de experiências de Bernoulli realizadas até à ocorrência do primeiro sucesso (inclusive).

$$Y \rightsquigarrow \mathrm{G}(p)$$

5.2.4.1. Função de Probabilidade

$$\underbrace{\boxed{}\boxed{}\boxed{\cdots}\boxed{}\boxed{}}_{y \text{ insucessos}} + \underbrace{\boxed{}}_{\substack{1 \\ \text{sucesso}}} \quad \rightarrow \quad p(y) = q^y \times p = p \times q^y$$

5.2.4.2. Parâmetros

Valor esperado	$\mu = \dfrac{q}{p}$
Variância	$\sigma^2 = \dfrac{q}{p^2}$

5.3. Distribuição Hipergeométrica $[\,\mathrm{H}(M \times p, M \times q, n)\,]$

5.3.1. Introdução

Considere-se uma ***população finita constituída por M elementos de dois tipos*** (ex. peças defeituosas e peças boas). Sejam p e $q = 1 - p$ as proporções nas quais os dois tipos de elementos figuram na população. Nestas condições teremos:

$M \times p$ elementos de um tipo (ex. peças defeituosas)

$M \times q$ elementos do outro tipo (ex. peças não defeituosas)

Admitindo que se retiram sucessivamente e sem reposição, n elementos da população.

- Uma variável aleatória Hipergeométrica Y representa o número de vezes que, no decurso das n experiências sucessivas, ocorre um dos dois resultados possíveis. Y pode tomar os valores 0, 1, 2, ..., n.

A diferença fundamental entre esta situação e a que estava na base da distribuição Binomial é que ***as experiências não são de Bernoulli***. De facto dado que a população é finita e que se tiram elementos sem reposição, os resultado de cada experiência passará a estar dependente dos resultados obtidos nas experiências anteriores. Dessa forma, ***os resultados das sucessivas experiências deixam de ser independentes***.

5.3.2. *Função de Probabilidade*

$$p(y) = \frac{\mathrm{C}_{y}^{M\times p} \times \mathrm{C}_{n-y}^{M\times q}}{\mathrm{C}_{n}^{M}}$$

$p(y)$ representa a probabilidade de, em n tentativas, se obterem y sucessos e n-y insucessos.

DGI
2019

No conjunto dos $\boldsymbol{n}$ resultados, o *número de casos possíveis* é:

$$\mathrm{C}_n^M$$

Dessas combinações, o número daquelas que contêm exactamente $\boldsymbol{y}$ elementos do primeiro tipo e consequentemente, $\boldsymbol{n}$ **-** $\boldsymbol{y}$ elementos do segundo tipo é:

$$\mathrm{C}_y^{M\times p} \times \mathrm{C}_{n-y}^{M\times q}$$

Dado que ***as diferentes combinações de elementos da população são equiprováveis***, a função de probabilidade $\boldsymbol{p(y)}$ vem dada por:

$$p(y) = \frac{\text{nº casos favoráveis}}{\text{nº casos possíveis}} = \frac{\mathrm{C}_y^{M\times p} \times \mathrm{C}_{n-y}^{M\times q}}{\mathrm{C}_n^M}$$

5.3.3. Parâmetros

Valor esperado	$\mu = n \times p$
Variância	$\sigma^2 = n \times p \times q \times \dfrac{M-n}{M-1}$
Desvio Padrão	$\sigma = \sqrt{n \times p \times q \times \dfrac{M-n}{M-1}}$

5.3.4. Relação entre as Distribuições Hipergeométrica e Binomial

Quando $M >> n$, o facto de não haver reposição dos elementos retirados ***não afecta substancialmente as probabilidades*** associadas aos resultados das experiências subsequentes. Dessa forma:

$$\mathrm{H}(M \times p, M \times q, n) \approx \mathrm{B}(n, p)$$

Critério de aproximação (regra prática): $M \geq 10 \times n$

Exemplo

y	$p(y)$			
	H(10,90,10)	B(10,0.1)	H(10,90,5)	B(5,0.1)
0	0.330	0.349	0.584	0.590
1	0.408	0.387	0.339	0.328
2	0.202	0.194	0.070	0.073
3	0.052	0.057	0.006	0.008
4	0.008	0.011	0.000	0.000
5	0.001	0.002	0.000	0.000
$\sum$	1.000	1.000	1.000	1.000
μ	1.000	1.000	0.500	0.500
σ^2	0.818	0.900	0.432	0.450

5.4. Distribuição de Poisson [$\mathrm{P}(\lambda)$]

5.4.1. *Introdução*

A distribuição de Poisson tem na sua génese um processo de Poisson e permite descrever certos tipos de fenómenos ou ***acontecimentos cuja ocorrência se repete no tempo ou no espaço***.

Exemplos:

- O número de golos marcados num jogo de futebol
- O número de avarias de uma máquina durante um dia de trabalho
- O número de erros ortográficos existentes num livro
- O número de defeitos de isolamento ao longo de um cabo eléctrico.

Para que a distribuição de Poisson tenha aplicação é suficiente que se verifiquem as seguintes condições:

- Os acontecimentos ocorrem um a um e não em grupos.
- A probabilidade de um acontecimento ocorrer num intervalo (de tempo ou de espaço) infinitamente pequeno é proporcional ao comprimento do intervalo e é independente do número de ocorrências verificadas em intervalos anteriores (***processo sem memória***).

Uma variável aleatória de Poisson Y representa o ***número ocorrências de um dado evento num intervalo temporal ou numa região espacial***. Y pode tomar os valores

$$\{0, 1, 2, 3, \ldots\} = \mathbb{N}_0$$

5.4.2. *Função de Probabilidade*

$$p(y) = \frac{e^{-\lambda} \times \lambda^{y}}{y!}$$

$\boldsymbol{p(y)}$ representa a probabilidade de se verificarem $\boldsymbol{y}$ ocorrências de um dado evento no intervalo de espaço/tempo considerado

5.4.3. *Parâmetros*

Valor esperado	$\mu = \lambda$
Variância	$\sigma^2 = \lambda$
Desvio Padrão	$\sigma = \sqrt{\lambda}$

5.4.4. Valores Tabelados

Os valores correspondem às probabilidades acumuladas: $F(y) = \sum_{u=0}^{y} \frac{e^{-\lambda} \times \lambda^u}{u!}$

	λ									
y ↓	***0.10***	***0.20***	***0.30***	***0.40***	***0.50***	***0.60***	***0.70***	***0.80***	***0.90***	***1.00***
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	0.3679
1	0.9953	0.9825	0.9631	0.9384	0.9098	0.8781	0.8442	0.8088	0.7725	0.7358
2	0.9998	0.9989	0.9964	0.9921	0.9856	0.9769	0.9659	0.9526	0.9371	0.9197
3	1.0000	0.9999	0.9997	0.9992	0.9982	0.9966	0.9942	0.9909	0.9865	0.9810
4	1.0000	1.0000	1.0000	0.9999	0.9998	0.9996	0.9992	0.9986	0.9977	0.9963
5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9997	0.9994
6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999

Exemplo

$\lambda = 1$

$P(y = 3) = ?$

$$P(y = 3) = \frac{e^{-\lambda} \times \lambda^y}{y!} = \frac{e^{-1} \times 1^3}{3!} = 0.0613$$ *(da função de probabilidade)*

$$P(y = 3) = F(3) - F(2) = 0.9810 - 0.9197 = 0.0613$$ *(da tabela)*

5.4.5. Relação entre a Distribuição de Poisson e a Distribuição Binomial

Na prática, interessa aproximar a distribuição Binomial pela de Poisson pois *o cálculo da função de probabilidade desta última é mais fácil*.

Critério de aproximação (regra prática):

- A *dimensão* da amostra é *elevada* $(n \geq 20)$
- A *distribuição* binomial é *assimétrica* $(n \times p < 7)$

Se a distribuição Binomial for simétrica, torna-se mais prático aproximá-la à distribuição normal (*distribuição simétrica*).

Se uma distribuição Hipergeométrica poder ser aproximada por uma distribuição Binomial e esta, por uma distribuição de Poisson, então deve recorrer-se a essa última distribuição para calcular $p(y)$.

DGI
2019

Exemplo

y	$p(y)$			
	B(10,0.1)	B(50,0.02)	B(100,0.01)	P(1)
0	0.349	0.364	0.366	0.368
1	0.387	0.372	0.370	0.368
2	0.194	0.186	0.185	0.184
3	0.057	0.061	0.061	0.061
4	0.011	0.015	0.015	0.015
5	0.002	0.003	0.003	0.003