3.5 A distribuição uniforme discreta

Definição: X tem distribuição uniforme discreta se cada um dos valores possíveis, $x_1,\ldots,x_n$, tiver função de probabilidade

$$P\left(X = x_i\right) = \begin{cases} \frac{1}{n}, & i = 1,\ldots,n \\ 0, & c.c. \end{cases}$$

e representa-se por

$$X \sim Unif\left\{x_1,\ldots,x_n\right\}.$$

Se X tem distribuição *Unif{1,…,n}* , então

$$P(X = i) = \begin{cases} \frac{1}{n}, & i = 1,\ldots,n \\ 0, & c.c. \end{cases}$$

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

$$V(X) = E\left(X^2\right) - E(X)^2 = \sum_{i=1}^{n} i^2\frac{1}{n} - \left(\frac{n+1}{2}\right)^2 =$$

$$= \frac{n^2-1}{12}$$

Exemplo: Seja X uma variável aleatória que conta o número de pontos obtidos no lançamento de um dado equilibrado, então *X~Unif{1,…,6}*

$$P(X = i) = \begin{cases} \frac{1}{6}, & i = 1,\ldots,6 \\ 0, & c.c. \end{cases}$$

$$E(X) = \frac{6+1}{2} = 3.5$$

$$V(X) = \frac{6^2-1}{12} = \frac{35}{12}$$

3.6 A distribuição binomial

Definição: uma experiência ou <u>prova de Bernoulli</u> é uma experiência aleatória só com dois resultados possíveis (um deles chamado "sucesso" e o outro "insucesso").

Seja P(sucesso) = p, $\quad 0 < p < 1$

Exemplos:

a) Lançamento de uma moeda. Podemos convencionar sucesso=cara. $p = 0.5$, se a moeda for perfeita.

b) Numa caixa estão 100 peças, das quais 10 são defeituosas. Extracção de uma peça ao acaso. Sucesso=defeituosa. $p = 0.1$

<u>Variável aleatória de Bernoulli</u> (ou com distribuição de Bernoulli) de parâmetro p:

X - número de sucessos observados quando se realiza uma prova de Bernoulli com P(sucesso) = p

$$\begin{array}{l} P(X = 0) = 1 - p \\ P(X = 1) = p \end{array} \qquad 0 < p < 1$$

ou $\qquad P(X = x) = p^x(1-p)^{1-x}, \quad x = 0,1$

$$E(X) = 0 \times (1-p) + 1 \times p = p$$

$$E\left(X^2\right) = 0^2 \times (1-p) + 1^2 \times p = p$$

$$V(X) = E\left(X^2\right) - E^2(X) = p - p^2 = p(1-p)$$

Representa-se $\quad X \sim Ber\,(p)$

Variável aleatória Binomial (ou com distribuição Binomial) de parâmetros *n* e *p*:

X - número de sucessos em n provas de Bernoulli realizadas de forma independente e com P(sucesso)= p em cada prova

$$X \sim Bin(n, p)$$

Valores possíveis de X: $0,1,2,\ldots,n$

Exemplos:

a) Lançamento de uma moeda perfeita 4 vezes consecutivas.

X - número de caras observado

F - cara (sucesso) C - coroa (insucesso)

É razoável admitir que os lançamentos são independentes.

Então $X \sim Bin\left(4, \frac{1}{2}\right)$

$$P(X=0) = P(C,C,C,C) = \left(\frac{1}{2}\right)^4$$

$$P(X=1) = P(F,C,C,C) + P(C,F,C,C) + \\ + P(C,C,F,C) + P(C,C,C,F) = 4 \times \left(\frac{1}{2}\right)^4$$

$$P(X=2) = P(F,F,C,C) + P(F,C,F,C) + \\ + P(F,C,C,F) + P(C,F,F,C) + \\ + P(C,F,C,F) + P(C,C,F,F) = \\ = 6 \times \left(\frac{1}{2}\right)^4 = \binom{4}{2} \times \left(\frac{1}{2}\right)^4$$

$$P(X=3) = 4 \times \left(\frac{1}{2}\right)^4 \qquad P(X=4) = \left(\frac{1}{2}\right)^4$$

Verificar que $P(X=0) + \ldots + P(X=4) = 1$

b) Lançamento de um dado perfeito 10 vezes consecutivas

X - número de vezes que sai a face 6

Sucesso=sair a face 6, $p = \frac{1}{6}$

Admitindo que os lançamentos são independentes

$$X \sim Bin\left(10, \frac{1}{6}\right)$$

$$P(X=0) = \left(\frac{5}{6}\right)^{10} \qquad P(X=1) = 10 \times \frac{1}{6} \times \left(\frac{5}{6}\right)^9$$

$$P(X=2) = \binom{10}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^8 \qquad \ldots\ldots\ldots$$

$$P(X=x) = \binom{10}{x} \times \left(\frac{1}{6}\right)^x \times \left(\frac{5}{6}\right)^{10-x}, \quad x = 0,1,\ldots,10$$

Em geral: Se $X \sim Bin(n, p)$ então

$$P(X=x) = f_X(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \\ x = 0,1,\ldots,n$$

Pode mostrar-se que

$$E(X) = \sum_{x=0}^{n} x P(X=x) = np$$

$$V(X) = \sum_{x=0}^{n} x^2 P(X=x) - E^2(X) = np(1-p)$$

Cálculo de probabilidades:

n pequeno: usar a expressão

n grande:

- programas em computador ou calculadora

- tabelas. Em geral o que está tabelado é a função de distribuição, mas

$$P(X = k) = P(X \le k) - P(X \le k-1) =$$

$$= F_X(k) - F_X(k-1)$$

- Também há métodos aproximados (a ver posteriormente)

3.7 A distribuição geométrica

Definição: Considere-se uma sucessão de provas de Bernoulli <u>independentes</u>, com P(sucesso)=p (constante). Seja

X - número de provas realizadas até à obtenção do <u>primeiro sucesso</u> (inclusive)

Diz-se que X tem distribuição geométrica de parâmetro p,

$$X \sim Geo(p)$$

Valores possíveis de X: $1, 2, 3, \ldots$

$P(X = 1) = p$ $\qquad$ $P(X = 2) = (1-p)p$

$P(X = 3) = (1-p)^2 p$ $\qquad$ ......

$$P(X = x) = f_X(x; p) = (1-p)^{x-1} p, \quad x = 1, 2, 3, \ldots$$

Pode mostrar-se que se $X \sim Geo(p)$, então

$$\mu_X = E(X) = \sum_{x=1}^{+\infty} x(1-p)^{x-1} p = \frac{1}{p}$$

$$\sigma_X^2 = V(X) = \frac{1-p}{p^2}$$

Exemplo: Problema das 10 chaves, com reposição.

X - nº de tentativas necessárias para encontrar a chave certa

$X \sim Geo(0.1)$, porque as tentativas são provas de Bernoulli independentes e com P(sucesso)=0.1 (constante).

$E(X) = 1/0.1 = 10$ e $V(X) = 0.9/0.01 = 90$

(comparar com a situação sem reposição!)

3.8 A distribuição hipergeométrica

Exemplo: Caixa com 100 peças das quais 10 são defeituosas. São efectuadas 5 extracções ao acaso, <u>sem reposição</u>. Seja

X - número de peças defeituosas encontradas

(10 def. e 90 não def.)

Valores possíveis: $0, 1, 2, 3, 4, 5$

$$P(X = 0) = \frac{\binom{90}{5}}{\binom{100}{5}} \cong 0.584$$

$$P(X = 1) = \frac{\binom{10}{1}\binom{90}{4}}{\binom{100}{5}} \cong 0.339$$

$$P(X=2)=\frac{\binom{10}{2}\binom{90}{3}}{\binom{100}{5}}\cong 0.070$$

$P(X=3)=\ldots\cong 0.006$

$P(X=4)=\ldots\cong 2.5\times 10^{-4}$

$P(X=5)=\ldots\cong 3.3\times 10^{-6}$

$$P(X=x)=\frac{\binom{10}{x}\binom{90}{5-x}}{\binom{100}{5}},\quad x=0,1,2,3,4,5$$

Nota: Se as extracções forem efectuadas <u>com reposição</u> (o que em termos práticos é estranho!...) têm-se 5 provas de Bernoulli independentes com $P(\text{sucesso})$ constante e igual a 0.1, logo $X \sim Bin(5,\,0.1)$.

Comparação das probabilidades:

	Sem reposição:	**Com** reposição:
$P(X=0)\cong$	0.584	0.590
$P(X=1)\cong$	0.339	0.328
$P(X=2)\cong$	0.070	0.073
$P(X=3)\cong$	0.006	0.008
$P(X=4)\cong$	2.5×10^{-4}	4.5×10^{-4}
$P(X=5)\cong$	3.3×10^{-6}	1×10^{-5}

A situação **com reposição**, embora irrealista, dá probabilidades próximas das calculadas para o caso **sem reposição** (mais adiante será indicado em que casos a aproximação é razoável).

Definição: Considere-se um conjunto de N objectos em que

K são do tipo 1 (ou sucessos)

N-K são do tipo 2 (ou insucessos)

são realizadas n extracções ao acaso e sem reposição (com $K \le N$ e $n \le N$). Seja

X - número de objectos do tipo 1 encontrados

Então X tem distribuição hipergeométrica

$$X \sim Hip(N,K,n)$$

$$P(X=x)=f_X(x;N,K,n)=\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

Valores possíveis:

Se $n \le K$ e $n \le N-K$, $\quad x=0,1,2,\ldots,n$

Se $n > K$ e $n \le N-K$, $\quad x=0,1,2,\ldots,K$

Se $n \le K$ e $n > N-K$, $\quad x=n-(N-K),\ldots,n$

Se $n > K$ e $n > N-K$, $\quad x=n-(N-K),\ldots,K$

ou, englobando todos os casos,

$$x=\max\left(0,\,n-(N-K)\right),\ldots,\min(K,n)$$

Pode mostrar-se que se $X \sim Hip(N,K,n)$, então

$$E(X)=np \qquad \text{e} \qquad V(X)=np(1-p)\frac{N-n}{N-1}$$

onde $\quad p=\dfrac{K}{N}$

Se N grande e n pequeno em relação a N $(n < 0.1N)$, pode usar-se a distribuição binomial para calcular valores aproximados das probabilidades:

$$P(X = x) \cong \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,1,\ldots,n$$

com $p = \dfrac{K}{N}$.

Pode representar-se por $\quad X \overset{a}{\sim} Bin(n, p)$

3.9 A distribuição de Poisson

A distribuição de Poisson pode ser vista como um limite "especial" da distribuição binomial: Suponha-se que n aumenta e que p diminui de tal forma que $E(X) = np$ se mantém constante, seja λ essa constante, então tem-se

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} =$$

$$= \frac{n(n-1)\cdots(n-x+1)}{x!}\left(\frac{np}{n}\right)^x \left(1-\frac{np}{n}\right)^n \left(1-\frac{np}{n}\right)^{-x} =$$

$$= \frac{n(n-1)\cdots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} =$$

$$= \frac{n(n-1)\cdots(n-x+1)}{n^x}\frac{\lambda^x}{x!} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-x} =$$

$$= \underset{\substack{\downarrow \\ 1}}{\frac{n}{n}}\;\underset{\substack{\downarrow \\ 1}}{\frac{n-1}{n}}\cdots\underset{\substack{\downarrow \\ 1}}{\frac{n-x+1}{n}}\;\frac{\lambda^x}{x!}\underset{\substack{\downarrow \\ e^{-\lambda}}}{\left(1-\frac{\lambda}{n}\right)^n}\;\underset{\substack{\downarrow \\ 1}}{\left(1-\frac{\lambda}{n}\right)^{-x}}$$

logo

$$\lim_{n\to\infty, np=\lambda} P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2,\ldots \; (\lambda > 0)$$

Definição: Dado um intervalo de números reais suponha-se que certas ocorrências surgem aleatoriamente ao longo do intervalo. Se o intervalo puder ser subdividido em subintervalos de comprimento suficientemente pequeno de modo a que se verifique:

1) A probabilidade de mais do que uma ocorrência num subintervalo é zero.

2) A probabilidade duma ocorrência num subintervalo é constante e proporcional ao comprimento do subintervalo.

3) As ocorrências nos diversos subintervalos são independentes.

(A uma "experiência" nestas condições chama-se <u>Processo de Poisson</u>)

Se o número médio de ocorrências no intervalo total for $\lambda > 0$ e se

X - número de ocorrências no intervalo total

então X tem distribuição de Poisson de parâmetro λ,

$$X \sim Poisson(\lambda)$$

e

$$P(X = x) = f_X(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0,1,2,\ldots$$

Verificação:

$$\sum_{x=0}^{\infty} P(X = x) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$$

Pode mostrar-se que:

$$E(X) = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \lambda$$

$$E(X^2) = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda} \lambda^x}{x!} = \lambda^2 + \lambda$$

$$V(X) = E(X^2) - E^2(X) = \lambda$$

Exemplos de situações em que <u>pode</u> ser aplicada a distribuição de Poisson:

- número de acidentes por semana num determinado cruzamento ou secção de estrada (não tendo em conta os acidentes em cadeia...)
- número de clientes que chegam a uma loja ou serviço num determinado intervalo de tempo (não tendo em conta as chegadas em grupo...)

- número de defeitos em peças ou materiais produzidos continuamente (tecidos, fios, etc.)

Cuidado com as unidades:

Exemplo: sabemos (por observação anterior) que o número médio de defeitos por m^2 de um certo tipo de tecido é 2. Se admitirmos que estamos nas condições do Processo de Poisson, então:

nº. de defeitos em 10 m^2 ~ Poisson $(\lambda = 20)$
nº. de defeitos em 100 m^2 ~ Poisson $(\lambda = 200)$
etc.

Qual a probabilidade de não haver defeitos numa peça com 5 m^2? E pelo menos 10 defeitos?

X - nº. de defeitos em 5 m^2 $\quad X \sim Poisson\,(10)$

$$P(X = 0) = e^{-10} \cong 0$$

$$\begin{aligned} P(X \geq 10) &= 1 - P(X < 10) = \quad (\text{complementar}) \\ &= 1 - P(X \leq 9) = \quad (\text{discreta em inteiros}) \\ &= 1 - F_X(9) = \quad (\text{definição de } F_X(x)) \\ &= 0.5421 \qquad (\text{tabelas}) \end{aligned}$$

Nota: dado o resultado com que começamos (a distribuição de Poisson como limite da distribuição Binomial) podemos também usar a distribuição de Poisson para calcular probabilidades aproximadas da distribuição binomial quando n grande e p pequeno (regra prática: $n > 20$ e $p < 0.1$).