MPEI 2018-2019

Aula 6

Caracterização de variáveis aleatórias (continuação) : Distribuições

Distribuições - Motivação

- As funções de massa de probabilidade e de densidade de probabilidade (para o caso contínuo) podem assumir as mais variadas formas
- Mas existe um conjunto de “formas” (distribuições) que aparecem repetidamente em muitos e variados problemas
  - Formam um conjunto de ferramentas base que é muito útil conhecer …

Existem muitas distribuições

- Discretas
  - **Bernoulli**
  - **Binomial**
  - **Poisson**
  - **Geométrica**
  - …
- Contínuas
  - **Uniforme**
  - **Normal**
  - Qui-quadrado
  - T de Student …

- Ver Wikipedia
  - https://en.wikipedia.org/wiki/List_of_probability_distributions

Distribuições Discretas

Variável de Bernoulli

- Distribuição directamente relacionada com as experiências de Bernoulli
- Seja A um acontecimento relacionado com o resultado de uma experiência aleatória
- A variável de Bernoulli define-se como
- $I_A(\omega) = \begin{cases} 1 & se\ \omega \in A \\ 0 & caso\ contrário \end{cases}$

Variável de Bernoulli

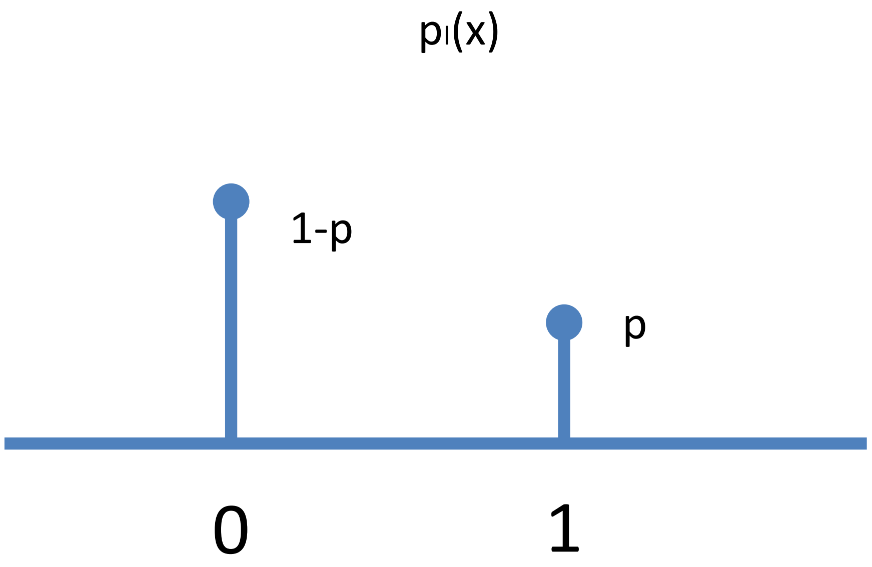
- O I que usamos para a designar resulta de ser usada muitas vezes como indicadora da ocorrência/não ocorrência de um evento

- Quando o evento ocorre a variável aleatória I assume o valor 1; caso contrário o valor 0

Variável de Bernoulli

- $S_I = \{0,1\}$
- p= Pr(A)

- $p_I(1)$ =p
- $p_I(0)$ =1-p



- Valor esperado E[I] ?
- Var(I) = ?

Variável de Bernoulli

- $\mathrm{E}[I] = \sum_i x_i p(x_i)$
  $$= 0 \times (1 - p) + 1 \times p$$
  $$= \boldsymbol{p}$$
- $Var(I) = E[I^2] - (E[I])^2$

- $E[I^2]$=$0^2 \times (1 - p) + 1^2 \times p = \mathrm{p}$

- $Var(I) = p - p^2 = \boldsymbol{p\,(1 - p)}$

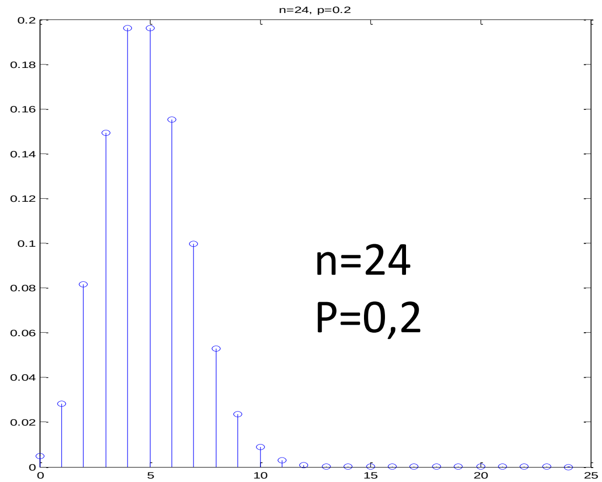
Variável Binomial

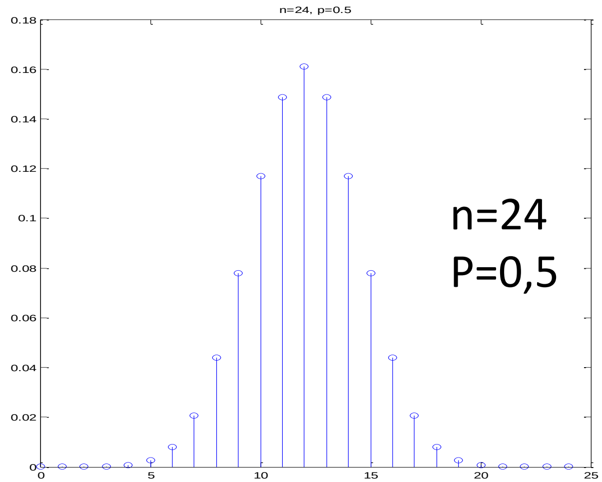
- Directamente relacionada com a Lei Binomial
- Seja X o número de vezes que um acontecimento A ocorre em n experiências de Bernoulli
  - isto é, X representa o número de sucessos em n experiências (observações)

- $X = \sum_{j=1}^{n} I_j$ → $S_X = \{0,1,2,..,n\}$

Variável Binomial

- $p_X(k) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$





- $F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}$

Variável Binomial – Média e Variância

- Fácil derivar usando o facto de termos **n variáveis de Bernoulli** independentes, que designamos por X_i

- $E[X] = E[\sum X_i] = \sum E[X_i] \quad pq\,?$
$$= p + p + \cdots + p = \boldsymbol{n}\,p$$

- De forma similar

$\mathrm{Var}(X) =$
$$Var(\sum X_i) = \sum Var(X_i) = \cdots = \boldsymbol{n}\,\underbrace{p\,(1-p)}$$

Variável Binomial - Exemplos

- Têm distribuição Binomial, por exemplo:
  - Número de peças defeituosas num lote de um determinado tamanho (ex: 50 peças)
  - Número de respostas certas num exame de verdadeiro falso
  - Número de clientes que efectuaram compras em 100 que entraram numa loja

Distrib. Binomial - Áreas de aplicação

- A distribuição surge em muitas áreas cientifico-tecnológicas:
  - Engenharia de produção: Muitas vezes as medidas de controlo de qualidade são baseadas na distribuição binomial
    - O caso Binomial aplica-se a qualquer situação industrial em que o resultado é binário e os resultados de ensaio são independentes e com probabilidades constantes
  - Medicina: Por exemplo os resultados "cura" ou "não cura" são importantes na indústria farmacêutica
  - Indústria Militar: "acerta" "falha" é muitas vezes a interpretação do lançamento de um míssil ou de uma missão
  - **Informática**: "acerto" e "falha" é uma interpretação possível para detectores de SPAM, testes a métodos/funções de um programa, procura de informação na web ...

Exemplo de aplicação 1: Transmissão digital

- Um sistema de transmissão digital envia um pacote de 1 kByte através de canal com ruído sendo a probabilidade de erro de cada bit 10^{-3} (ou seja 1 bit em cada mil).
- Considerando que os erros são independentes, determine:
  - Probabilidade de haver 1 erro ?
  - Probabilidade de haver erro ?

Exemplo 2 – segurança de aviões

- Considere que um motor de avião pode falhar com probabilidade p e que as falhas em motores distintos são independentes.
- Se um avião se despenha quando mais do que 50% dos motores falham, é mais seguro voar num avião de 4 motores ou de 2 motores ?
- Faz parte do guião Prático 3
- Como resolver ?

- Sugestão: calcular a probabilidade de cair um avião com 2 motores, repetir para o de 4 motores e comparar os resultados (será função da probabilidade de falha de um avião)

Possível resolução

- O de 2 motores despenha-se se os 2 motores falharem. Qual a probabilidade de 2 falhas em 2 motores ?
- $p2 = p_X(2, n = 2) = \binom{2}{2} p^2 (1-p)^{2-2} = p^2$

- O de 4 despenha-se se 3 ou 4 falharem. Qual a probabilidade ?
- $p4 = p_X(3, n = 4) + p_X(4, n = 4)$
- $= \binom{4}{3} p^3 (1-p)^{4-3} + \binom{4}{4} p^4 (1-p)^{4-4}$
- = $4\,p^3(1-p) \;\; + p^4 \;\; = 4\,p^3 - 3\,p^4$

- Relação entre $p2$ e $p4$
- $\frac{p4}{p2} = 4p - 3p^2 = p\,(4-3p)$
  - NOTE que depende de p

…

p	p2	p4	p4/p2
0,01	0,0001	0,00000397	0,0397
0,02	0,0004	0,00003152	0,0788
0,03	0,0009	0,00010557	0,1173
0,04	0,0016	0,00024832	0,1552
0,05	0,0025	0,00048125	0,1925
0,06	0,0036	0,00082512	0,2292
0,07	0,0049	0,00129997	0,2653
0,08	0,0064	0,00192512	0,3008
0,09	0,0081	0,00271917	0,3357
0,1	0,01	0,0037	0,37
0,2	0,04	0,0272	0,68
0,3	0,09	0,0837	0,93
0,4	**0,16**	**0,1792**	**1,12**
0,5	0,25	0,3125	1,25
0,6	0,36	0,4752	1,32
0,7	0,49	0,6517	1,33
0,8	0,64	0,8192	1,28
0,9	0,81	0,9477	1,17

p	p2	p4	p4/p2
0,3	0,09	0,0837	0,93
0,31	0,0961	0,091458	0,9517
0,32	0,1024	0,099615	0,9728
0,33	0,1089	0,10817	0,9933
0,34	**0,1156**	**0,117126**	**1,0132**
0,35	0,1225	0,126481	1,0325
0,36	0,1296	0,136236	1,0512
0,37	0,1369	0,146387	1,0693
0,38	0,1444	0,156934	1,0868
0,39	0,1521	0,167873	1,1037

O que significam p4/p2 < 1 ?

é mais seguro voar num avião de 4 motores ou de 2 motores ?

Exemplo de aplicação III

- According to the U.S. Census Bureau, approximately 6% of all workers in Jackson, Mississippi, are unemployed.
- In conducting a random telephone survey in Jackson, what is the probability of getting two or fewer unemployed workers in a sample of 20?

- De: Business Statistics, Ken Black, 6th ed, John Willey & Sons (cap 5)

Resolução

- 6% desempregado => $p = 0{,}06$
- Tamanho da amostra é 20 => $n = 20$
- 94% têm emprego => $1 - p = 0{,}94$
- x é o número de sucessos que se pretende

- Qual é a probabilidade de termos 2 ou menos desempregados na amostra de 20 ?

- Neste tipo de problemas o importante e muitas vezes o mais difícil é identificar o p, n e x

Resolução

$$n = 20$$
$$p = 0{,}06$$
$$q = 1 - p = 0{,}94$$
$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= 0{,}2901 + 0{,}3703 + 0.2246 = 0{,}8850$$

$$P(X = 0) = \frac{20!}{0!(20-0)!}\left(0{,}06\right)^{0}\left(0{,}94\right)^{20-0} = (1)(1)(0{,}2901) = 0{,}2901$$

$$P(X = 1) = \ldots$$

$$P(X = 2) = \ldots.$$

Variável Geométrica

- Seja X o número de vezes que é necessário repetir uma experiência de Bernoulli até obter um sucesso
  - Prob. Sucesso: p   prob. Falha = 1-p

- $p_X(k) = \boldsymbol{p}(1-p)^{k-1}, k = 1,2,3, \ldots$

  Porque teremos k-1 insucessos e depois sucesso

- $F_X(x) = \sum_{k=0}^{\lfloor x \rfloor} p(1-p)^{k-1}$

Exemplo de aplicação – Helpdesk UA

- Problema:
- Considere o serviço de atendimento via telefone do Helpdesk da UA.
- Supondo que a probabilidade de se conseguir contactar o suporte é p=0,1 (só ao fim de 10 tentativas ☹).
- Determine a probabilidade de necessitar de menos de 3 chamadas até conseguir expor o seu problema ?

- Solução:
- $\Pr(nº\ chamadas < 3) = \Pr(1\ chamada\ OU\ 2\ chamadas)$
- $= p(1-p)^{1-1} + p(1-p)^{2-1} = p(2-p) = 0{,}19$

Variável Geométrica – Média e Variância

- Demontra-se que:

- $E[X] = \frac{1}{p}$
  - Resultado de $\sum_{i=1}^{\infty} i \;\; p(1-p)^{i-1}$
  - Intuitivo: no exemplo do Helpdesk, por exemplo, quanto mais provável atenderem menos chamadas teremos que fazer (em média)

- $Var(X) = (1-p)/p^2$

Binomial para valores de n elevados

- Consideremos o seguinte cenário:
- Num conjunto de programas a probabilidade de haver pelo menos um erro ao analisar um conjunto de **1.000 linhas** de código é **p** $(p<1)$
  - Não nos interessa o número de erros, apenas se existe algum ou não
- Se o número total de linhas dos programas for **N x 1000** e os dividirmos em blocos de 1000 linhas a probabilidade de **k** blocos terem erros segue a distribuição **Binomial com parâmetros N e p**

(continuação)

- Se quisermos analisar células de 100 linhas, e considerarmos que a distribuição dos erros é uniforme, a probabilidade desce para p/10
  - Teríamos então uma Binomial com parâmetros 10 N e p/10
  - Teoricamente temos a forma de cálculo mas basta N ser um número moderado e 10N começa a ser elevado e os cálculos complicados [mesmo em computador]

- Exemplo: Blocos de 100 linhas; 1000 blocos ; p=0,98/10
- Qual a probabilidade de blocos com erro ser inferior ou igual a 100 ?

$$P = F_x(100) = \sum_{k=0}^{100} \binom{1000}{k} 0{,}098^k 0{,}902^{1000-k}$$

☹

…

- As coisas ainda se complicam mais de reduzirmos mais o tamanho dos blocos (100 linhas, 10 linhas ..)

- Será que conseguimos arranjar maneira(s) eficiente(s) de calcular quando o tamanho é muito pequeno ?

- No limite teremos apenas um bloco que vai ter ou não um erro
  - Número de blocos “infinitesimais” com erro = número de erros

Distribuição de Poisson

- Considere-se que temos uma variável Binomial, n cresce e p decresce por forma a $np \rightarrow \lambda > 0$

- Para n grande pode fazer-se as seguintes aproximações: $p \cong \frac{\lambda}{n}$ e $1 - p \cong 1 - \frac{\lambda}{n}$

- $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ → Binomial

- $\lim\limits_{n \to \infty} P(X = k) = \lim\limits_{l \to \infty} \binom{n}{k} p^k (1-p)^{n-k}$ =
- = ... = $\frac{\lambda^k e^{-\lambda}}{k!}$
- $p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$

é a função de massa de probabilidade da distribuição de Poisson , com k=0, 1, 2, …

Demonstração

- $\lim_{n\to\infty} \binom{n}{k} p^k (\mathbf{1}-\boldsymbol{p})^{\boldsymbol{n-k}}$
- $= \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$
- $= \lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{k!} \frac{\lambda^k}{n^k} \left(1-\frac{\lambda}{n}\right)^n \left(1-\frac{\lambda}{n}\right)^{-k}$
- $= \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n(n-1)\dots(n-k+1)}{n^k} \left(\mathbf{1}-\frac{\boldsymbol{\lambda}}{\boldsymbol{n}}\right)^{\boldsymbol{n}} \left(1-\frac{\lambda}{n}\right)^{-k}$
- $= \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{\boldsymbol{n}}{\boldsymbol{n}} \frac{\boldsymbol{n-1}}{\boldsymbol{n}} \frac{\boldsymbol{n-2}}{\boldsymbol{n}} \dots \frac{\boldsymbol{n-k+1}}{\boldsymbol{n}} \left(\mathbf{1}-\frac{\boldsymbol{\lambda}}{\boldsymbol{n}}\right)^{\boldsymbol{n}} \left(1-\frac{\lambda}{n}\right)^{-k}$
- $= \frac{\lambda^k}{k!} \lim_{n\to\infty} \mathbf{1} \left(\mathbf{1}-\frac{\mathbf{1}}{\boldsymbol{n}}\right)\left(\mathbf{1}-\frac{\mathbf{2}}{\boldsymbol{n}}\right)\dots\left(\mathbf{1}-\frac{\boldsymbol{k-1}}{\boldsymbol{n}}\right) \left(\mathbf{1}-\frac{\boldsymbol{\lambda}}{\boldsymbol{n}}\right)^{\boldsymbol{n}} \left(1-\frac{\lambda}{n}\right)^{-k}$

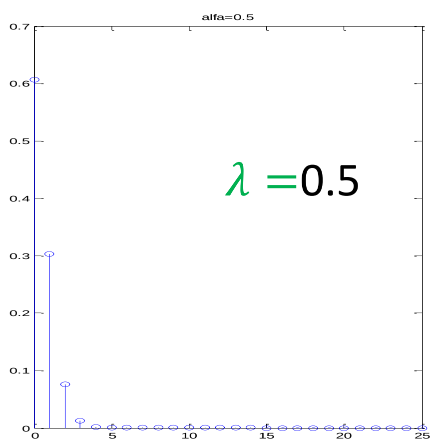
- como $\boldsymbol{\lim_{n\to\infty}} \left(\mathbf{1}-\frac{\boldsymbol{\alpha}}{\boldsymbol{n}}\right)^{\boldsymbol{n}} = e^{-\alpha}$, $\boldsymbol{\lim_{n\to\infty}} \left(\mathbf{1}-\frac{\boldsymbol{j}}{\boldsymbol{n}}\right) = 1$ e o limite do produto é o produto dos limites temos:
- $= \frac{\lambda^k}{k!} \times 1 \times e^{-\lambda} \times 1$

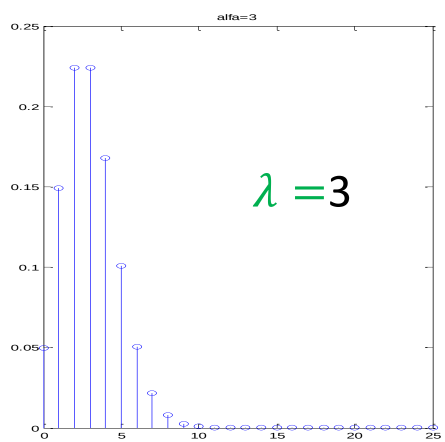
Distribuição para vários valores do parâmetro λ

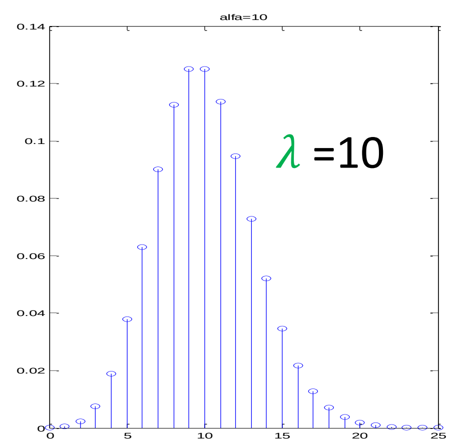
- Função de probabilidade:

$$p_X(k) = \Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0,1,2,3\ \ldots$$

- Tem apenas um parâmetro, o lambda







Variável de Poisson – Média e Variância

- $E[X] = \lambda$
  - Relembre que λ é aproximado por np e o valor esperado da Binomial é np

- $Var(X) = \lambda$

Variável de Poisson

- A distribuição de Poisson foca-se apenas no número de ocorrências (discreto) num intervalo de tempo contínuo (ou região do espaço).

- Esta distribuição não tem um número de experiências (n) como na Binomial
  - As ocorrências são independentes das outras ocorrências

Aproximação de Poisson à distribuição Binomial

- Problemas envolvendo a distribuição Binomial em que **n é grande e o valor de p é pequeno**, gerando desta forma **eventos raros**, são os candidatos à utilização da distribuição de Poisson

- Regra prática ("rule of thumb") :
  - Se n>20 e np <=7   a aproximação de Poisson é suficientemente próxima para ser usada em vez da Binomial

Aproximação de Poisson à distribuição Binomial

- Procedimento para aproximar a Binomial por Poisson:
    1. Calcular a média da Binomial **$\mu = np$**

    2. Como μ é o valor esperado da Binomial, passa a ser o **λ** (=E[X]) de Poisson

    3. Usar a fórmula de Poisson (ou uma tabela)

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Aplicações da distribuição Poisson

- As distribuições de Poisson surgem em experiências onde se verificam as seguintes propriedades:
  - O número de resultados que ocorrem num determinado intervalo de tempo ou região é independente do número que ocorre em qualquer outro intervalo temporal ou região espacial disjunta
  - A probabilidade que um resultado ocorra durante um intervalo ou região infinitesimal é proporcional ao comprimento do intervalo ou dimensão da região e não depende das ocorrências fora desse intervalo ou região
  - A probabilidade de haver mais que um resultado numa região infinitesimal é desprezável

Exemplo de aplicação

- Bank customers arrive randomly on weekday afternoons at an average of 3.2 customers every 4 minutes.
- What is the probability of having more than 7 customers in a 4-minute interval on a weekday afternoon?

- De: Business Statistics, Ken Black, 6th ed, John Willey & Sons (cap 5)

Resolução

- Consideremos o número de clientes (em intervalos de 4 minutos) como a variável aleatória X

- Pretendemos

  P("X > 7 clientes /4 minutos")

- λ= ?

- λ = 3,2   [clientes em 4 minutos]

Resolução (continuação)

- A solução requer que calculemos para k = 8, 9, 10, 11, 12, 13, 14, ... até o valor ser aproximadamente zero
  - Ou usemos o complemento e calculemos k =0,1,2,3,4,5,6,7

- Depois é só somar as probabilidades

- O resultado (0,0169) mostra que é pouco provável que um banco que tem em média 3,2 clientes a cada 4 minutos receba 7 clientes num período de 4 minutos
  - TPC: confirmar este valor

Resolução (continuação)

- Este tipo de probabilidades são muito úteis para os gestores de Bancos (e outras instituições com atendimento ao público) dimensionarem o número de pessoas e postos de atendimento

- A distribuição de Poisson é também muito útil na modelação da chegada de mensagens (ou outros tipos de eventos) em redes de computadores