Aula 3 - Variáveis aleatórias discretas

PhD. Wagner Hugo Bonat

Laboratório de Estatística e Geoinformação-LEG
Universidade Federal do Paraná

1/2017

Definições

- Variável aleatória (va) discreta: Assume valores num conjunto enumerável, com certa probabilidade.
- Variável aleatória contínua: Assume valores em qualquer intervalo real, o que caracteriza um conjunto não enumerável.
- Definição 3.1 - Função discreta de probabilidade (fp)

$$P(X = x_i) = p(x_i) = p_i, \quad i = 1,2,\ldots$$

- Uma função discreta de probabilidade deve satisfazer

$$0 \leq p_i \leq 1 \quad \text{e} \quad \sum_i p_i = 1.$$

- Defínção 3.2 - Função acumulada de probabilidade

$$F(x) = P(X \leq x).$$

Laboratório de Estatística e Geoinformação

Exemplo 3.1

- Com dados do último censo, a assistente social de um Centro de Saúde constatou que para as famílias da região, 20% não têm filhos, 30% têm um filho, 35% têm dois e as restantes se dividem igualmente entre três, quatro ou cinco filhos. Descreve a função de probabilidade da va N definida com número de filhos.

Laboratório de Estatística e Geoinformação

Exemplo 3.2

- Na construção de um certo prédio, as fundações devem atingir 15 metros de profundidade e, para cada 5 metros de estacas colocadas, o operador anota se houve alteração no ritmo de perfuração previamente estabelecido. Essa alteração é resultado de mudanças para mais ou para menos, na resitência do subsolo. Nos dois casos, medidas corretivas serão necessárias, encarecendo o custo da obra. Com base em avaliações geológicas, admite-se que a probabilidade de ocorrência de alterações é de 0.1 para cada 5 metros. O custo básico inicial éde 100 UPC (unidade padrão de construção) e será acrescida de $50k$, com k representando o número de alterações observadas. Como se comporta a va custo das obras de fundações?

Laboratório de Estatística e Geoinformação

Exemplo 3.3

- Considere o experimento de lançar uma certa moeda e obsevar se ocorre cara ou coroa. Descreva o comportamento da variável número de caras em dois lançamentos dessa moeda.

Exemplo 3.4

- Um jogador paga 5 fichas para participar de um jogo de dados, disputando com a banca quem tem o ponto maior. O jogador e a banca lançam cada um o seu dado e a seguinte regra de premiação é estabelecida:
  1. se o ponto do jogador é maior, ele ganha 2 vezes a diferença entre o seu ponto e o obtido pela banca;
  2. se o ponto do jogador é menor ou igual ao da banca, ele não ganha nada.

  O que você acha deste jogo?

Exemplo 3.5

- Uma população de 1000 crianças foi analisada num estudo para determinar a efetividade de uma vacina contra um tipo de alergia. No estudo, as crianças recebiam uma dose da vacina e, após um mês, passavam por um novo teste. Caso ainda tivessem tido alguma reação alérgica, recebiam outra dose da vacina. Ao fim de 5 doses todas as crianças foram consideradas imunizadas. Os resultados completados estão na tabela a seguir.

Doses	1	2	3	4	5
Freq.	245	288	256	145	66

- Para uma criança sorteado ao acaso qual a probabilidade dela ter recebido 2 doses? e até 2 doses?

Laboratório de Estatística e Geoinformação

Exemplo 3.6

- Num estudo sobrea incidência de câncer foi registrado, para cada paciente com esse diagnóstico, o número de casos de câncer em parentes próximos (pais, irmãos, tios, filhos, primos e sobrinhos). Os dados de 26 pacientes são os seguintes:

```
> x <- c(2,5,0,2,1,5,3,3,3,2,0,1,1,4,5,2,2,3,2,1,5,4,0,0,3,3)
> table(x)

x
0 1 2 3 4 5
4 4 6 6 2 4

> prop.table(table(x))

x
         0          1          2          3          4          5
0.15384615 0.15384615 0.23076923 0.23076923 0.07692308 0.15384615
```

Exemplo 3.6

- Estudos anteriores assumem que a incidência de câncer em parentes próximos pode ser teoricamente modelada pela seguinte função discreta de probabilidade:

Incidência	0	1	2	3	4	5
p_i	0.1	0.1	0.3	0.3	0.1	0.1

Os dados observados concordam com o modelo teórico?

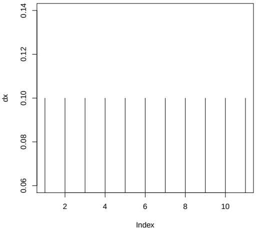
```
> esperado <- c(0.1,0.1,0.3,0.3,0.1,0.1)
> obs <- prop.table(table(x))
> cbind(round(cbind(esperado, obs), 2) , 26*cbind(esperado, obs))

  esperado  obs esperado obs
0      0.1 0.15      2.6   4
1      0.1 0.15      2.6   4
2      0.3 0.23      7.8   6
3      0.3 0.23      7.8   6
4      0.1 0.08      2.6   2
5      0.1 0.15      2.6   4
```

Modelo Uniforme Discreto

- Seja X uma va cujos possíveis valores são representados por $x_1, x2, \ldots, x_k$. Dizemos que X segue o modelo Uniforme Discreto se atribui a mesma probabilidade $1/k$ a cada um desses k valores. Sua função de probabilidade é dada por

$$P(X = x_j) = 1/k, \forall j = 1, 2, \ldots, k.$$

```
> x = c(0:10)
> dx = dunif(x = x, min = 0, max = 10)
> plot(dx, type = "h")
```



Exemplo 3.7

- Uma rifa tem 100 bilhetes numerados de 1 a 100. Tenho 5 bilhetes consecutivos numerados de 21 a 25 e meu colega tem outros 5 bilhetes, com os números $1, 11, 29, 68$ e 93. Quem tem maior possibilidade de ser sorteado?

Laboratório de Estatística e Geoinformação

Modelo Bernoulli

- Dizemos que uma variável X segue o modelo Bernoulli se atribui 0 ou 1 à ocorrência de fracasso ou sucesso, respectivamente. Sendo, p a probabilidade de sucesso, $0 \leq p \leq 1$, sua função de probabilidade é dada por
$$P(X = x) = p^x(1 - p)^{1-x}, \quad x = 0{,}1.$$
- Exemplo 3.8 - Sabe-se que a eficiência de uma vacina é de 80%. Um grupo de três indivíduos é sorteado, dentre a população vacinada, e submetido a testes pra averiguar se a imunização foi efetiva, evento representado por I. Determine o comportamento da va número de indivíduos imunizados neste grupo.

Modelo Binomial

- Considere a repetição de n ensaios Bernoulli independentes e todos com a mesma probabilidade de sucesso p. A va que conta o número total de sucessos é denominada Binomial com parâmetros n e p e sua função de probabilidade é dada por

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

onde

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

- Notação $X \sim b(n,p)$.
- Calculando probabilidades baseadas na Binomial.

```
> dbinom(3, prob = 0.4, size = 12)
[1] 0.141894
> round(dbinom(0:12, prob = 0.4, size = 12),3)
 [1] 0.002 0.017 0.064 0.142 0.213 0.227 0.177 0.101 0.042 0.012 0.002 0.000 0.000
```

Laboratório de Estatística e Geoinformação

Exemplo 3.9

- O escore de um teste internacional de proficiência na língua ingles varia de 0 a 700 pontos, com mais pontos indicando um melhor desempenho. Informações, coletadas durante vários anos, permite estabelecer o seguinte modelo para o desempenho no teste:

Pontos	[0,200)	[200,300)	[300,400)	[400,500)	[500,600)	[600,700
p_i	0.06	0.15	0.16	0.25	0.28	0.10

Várias universidades americanas, exigem um escore mínimo de 600 pontos para aceitar candidatos de paísed de língua não inglesa. De um grande grupo de estudantes brasileiros que prestaram o último exame, escolhemos ao acaso 20 deles. Qual seria a probabilidade de no máximo 3 atenderem ao requisito mínimo?

Laboratório de Estatística e Geoinformação

Exemplo 3.9

- Um veterinário está estudando o índice de natalidade em porcos à inseminação artificial. Para tal, coletou informações sobre a variável número de filhotes nascidos vivos em cada uma das 100 inseminações realizadas com o mesmo reprodutor. A tabela a seguir apresenta os resultados.

Número	1	2	3	4	5	6	7	8	9
Freq. Obs.	1	6	7	23	26	21	12	3	1

  O veterinário informa que 11 ou mais filhotes nascidos vivos é uma ocorrência muito rara e pode ser desprezada em termos do modelo. Ele sugeriu que a variável N: número de filhotes nascidos vivos, poderia ser ajustada pelo modelo Binomial com parâmetros $n = 10$ e $p = 0.5$. O que você acha da sugestão do veterinário?

Laboratório de Estatística e Geoinformação

Exemplo 3.9

- Usando o R temos.

```
> obs <- c(0,1,6,7,23,26,21,12,3,1,0)
> esp <- 100*dbinom(0:10, prob = 0.5, size = 10)
> cbind("Obs" = obs, "Esp" = round(esp, 1) )

      Obs  Esp
 [1,]   0  0.1
 [2,]   1  1.0
 [3,]   6  4.4
 [4,]   7 11.7
 [5,]  23 20.5
 [6,]  26 24.6
 [7,]  21 20.5
 [8,]  12 11.7
 [9,]   3  4.4
[10,]   1  1.0
[11,]   0  0.1
```

Modelo Geométrico

- Número de ensaios Bernoulli até o primeiro sucesso.
- X tem distribuição Geométrica de parâmetro p, se sua função de probabilidade tem a forma

$$P(X = k) = p(1-p)^k, \quad 0 \leq p \leq 1, \quad \text{e} \quad k = 0{,}1, \ldots .$$

- Exemplo 3.11: Uma linha de produção está sendo analisada para efeito de controle da qualidade das peças produzidas. Tendo em vista o alto padrão requerido, a produção é interrompida para regulagem toda vez que uma peça defeituosa é observada. Se 0.01 é a probabilidade da peça ser defeituosa, estudo o comportamento da variável Q, quantidade de peças boas produzidas antes da primeira defeituosa.

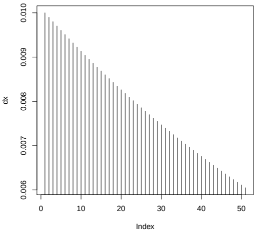
Laboratório de Estatística e Geoinformação

Exemplo 3.11

- $P(Q = k) = 0.01 * 0.99^k$
- Usando o R temos

```
> dgeom(x = 1, prob = 0.01)

[1] 0.0099
```

```
> x = c(0:50)
> dx = dgeom(x = x, prob = 0.01)
> plot(dx, type = "h")
```



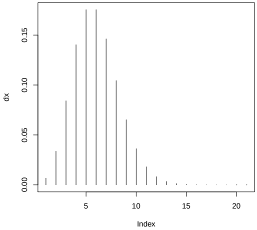
Modelo Poisson

- X tem distribuição Poisson com parâmetro $\lambda > 0$, se sua função de probabilidade é dada por

$$P(X = k) = \frac{\exp -\lambda\lambda^k}{k!}, \quad k = 0, 1, 2 \ldots.$$

- Notação $X \sim Po(\lambda)$.

```
> x = c(0:20)
> dx = dpois(x = x, lambda = 5)
> plot(dx, type = "h")
```



Exemplo 3.12

- A emissão de partículas radioativas tem sid modelada através de uma distribuição de Poisson, com o valor do parâmetro dependendo da fonte utilizada. Suponha que o número de partículas alfa, emitidas por minuto, seja uma variável aleatória seguindo o modelo Poisson com parâmetro 5, isto é, a taxa média de ocorrência é de 5 emissões a cada minuto. Calcule a probabilidade de haver mais de 2 emissões em um minuto.

$$
\begin{aligned}
P(A > 2) &= \sum_{a=3}^{\infty} P(A = a) = 1 - \sum_{a=0}^{2} P(A = a) && (1) \\
&= 1 - \sum_{a=0}^{2} \frac{\exp\{-5\}5^a}{a!} = 0.875 && (2)
\end{aligned}
$$

- Em R temos

```
> 1 - ppois(2, lambda = 5)
[1] 0.875348
```

Exemplo 3.13

- Engenheiros da companhia telefônica estudam se o modelo de Poisson com taxa de ocorrência de 4.5 chamadas por hora pode ser ajustado ao número N de chamadas interestaduais que chegam por hora, a uma central telefônica, durante o período noturno. Os dados coletados referentes a 650 períodos de uma hora, estão apresentados abaixo. Analise se esta suposição é razoável.

```
> obs <- c(9,38,71,115,125,106,79,50,57)
> esp <- c(dpois(0:7, lambda = 4.5), 1 - ppois(7, lambda = 4.5))
> cbind("Obs" = obs, "Esp" = 650*esp)
```

```
      Obs        Esp
 [1,]   9   7.220848
 [2,]  38  32.493815
 [3,]  71  73.111083
 [4,] 115 109.666625
 [5,] 125 123.374953
 [6,] 106 111.037458
 [7,]  79  83.278094
 [8,]  50  53.535917
 [9,]  57  56.281207
```

Modelo hipergeométrico

- Considere um conjunto de n objetos dos quais m são do tipo I e $n - m$ são do tipo II. Para um sorteio de r objetos $r < n$, feito ao acaso e sem reposição, defina X como o número de objetos de tipo I selecionados. Diremos que a va X segue o modelo Hipergeométrico e sua função de probabilidade é dada por

$$P(X = k) = \frac{\binom{m}{k}\binom{n-m}{r-k}}{\binom{n}{r}}, \quad k = \max(0, r-(n-m)), \ldots, \min(r,m).$$

Laboratório de Estatística e Geoinformação

Exemplo 3.14

- Uma fábrica produz peças que são embaladas em caixas com 25 unidades. Para aceitar o lote enviado por essa fábrica, o controle de qualidade de uma empresa procede da seguinte forma: Sorteia uma caixa do lote e, em seguida sorteia cinco peças, sem reposição, dessa mesma caixa. Se constatar no máximo duas defeituosas, aceita o lote fornecido pela fábrica. Se a caixa sorteada tivesse 4 peças defeituosas, qual seria a probabilidade de rejeitar o lote?

Exercícios recomendados

- Seção 3.1 - 3,5 e 6.
- Seção 3.2 - 2,3,5 e 6.
- Seção 3.3 - 1,2,3,4,5 e 6.