AULA 07
Distribuições Discretas de Probabilidade

Ernesto F. L. Amaral

31 de agosto de 2010
Metodologia de Pesquisa (DCP 854B)

Fonte:

Triola, Mario F. 2008. "Introdução à estatística". 10 ª ed. Rio de Janeiro: LTC. Capítulo 5 (pp.158-191).

ESQUEMA DA AULA

– Variáveis aleatórias

– Distribuição de probabilidade binomial

– Média, variância e desvio padrão para a distribuição binomial

– Distribuição de probabilidade de Poisson

VISÃO GERAL

– Neste capítulo, métodos de estatística descritiva (caps. 2 e 3) são combinados com métodos de probabilidade (cap. 4).

– São desenvolvidas as distribuições de probabilidade, que descrevem o que provavelmente acontecerá, ao invés do que realmente aconteceu.

– São discutidas distribuições discretas de probabilidade (binomial e Poisson).

– Com o conhecimento da população de resultados, é possível achar suas características (média, desvio padrão...).

VARIÁVEIS ALEATÓRIAS

VARIÁVEIS ALEATÓRIAS

– **Variável aleatória** é uma variável (representada por *x*) que tem um único valor numérico, determinado pelo acaso, para cada resultado de um experimento.

– **Distribuição de probabilidade** é uma descrição que dá a probabilidade para cada valor da variável aleatória. Ela é representada na forma de um gráfico, tabela ou fórmula.

– **Variável aleatória discreta** tem número finito de valores (número inteiro) ou quantidade enumerável de valores (é possível realizar um processo de contagem, mesmo que existam infinitos valores).

– **Variável aleatória contínua** tem infinitos valores que podem ser associados com medidas em uma escala contínua, de modo que não há pulos ou interrupções.

GRÁFICOS

– O **histograma de probabilidade** é o principal gráfico para representar uma distribuição de probabilidade.

– Este histograma mostra probabilidades na escala vertical.

– O histograma de frequência mostra frequências relativas com base em dados amostrais reais.

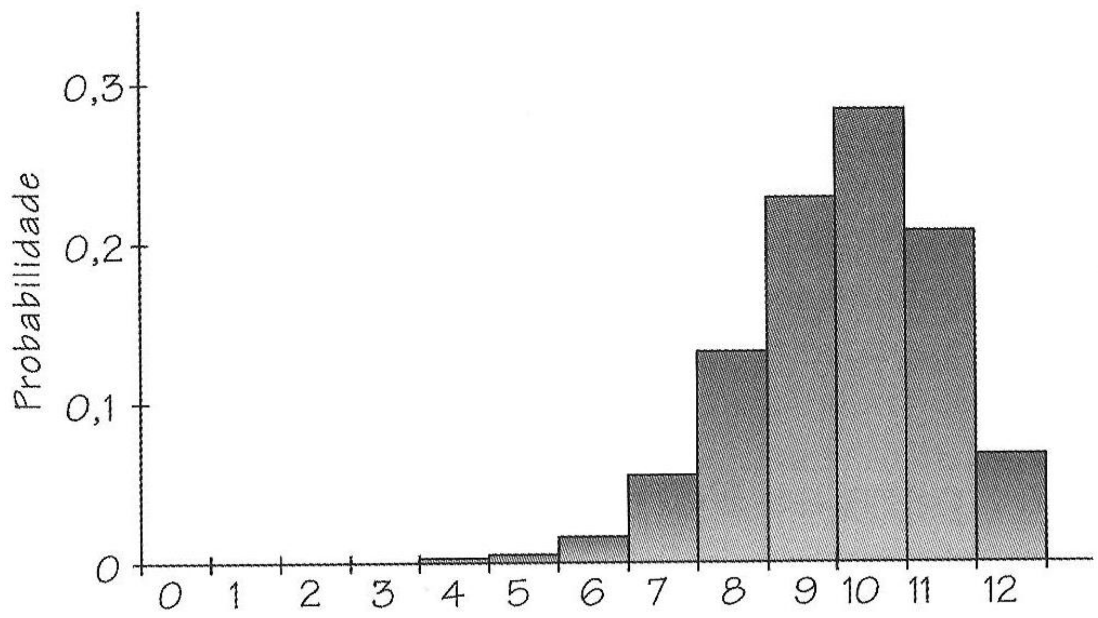
– **Requisitos** para uma distribuição de probabilidade:

  – A soma de todas as probabilidades deve ser um [$\sum P(x)=1$], em que x assume todos valores possíveis.

  – Cada valor de probabilidade deve estar entre zero e um, inclusive [$0 \leq P(x) \leq 1$], para todo valor individual de x.

PROBABILIDADE DE NÚMEROS DE MEXICANOS EM JÚRI

***x* (mexicanos-americanos)**	***P*(*x*)**
0	0+
1	0+
2	0+
3	0+
4	0,001
5	0,003
6	0,016
7	0,053
8	0,133
9	0,236
10	0,283
11	0,206
12	0,069



Histograma de Probabilidade para o Número de Jurados Mexicanos-Americanos entre 12

MÉDIA, VARIÂNCIA E DESVIO PADRÃO

– Além do histograma de probabilidade que nos informa sobre natureza e forma da distribuição, podemos achar média, variância e desvio padrão dos dados.

– **Média** para uma distribuição de probabilidade:

$$\mu = \sum [x * P(x)]$$

– **Variância** para uma distribuição de probabilidade:

$$\sigma^2 = \sum [(x - \mu)^2 * P(x)]$$

$$\sigma^2 = \sum [x^2 * P(x)] - \mu^2$$

– **Desvio padrão** para uma distribuição de probabilidade:

$$\sigma = \sqrt{\sum [x^2 * P(x)] - \mu^2}$$

FUNDAMENTOS DAS FÓRMULAS

– A **fórmula da média** faz o mesmo que a fórmula para a média de uma tabela de frequência, em que *f* representa a frequência da classe e *N* representa o tamanho da população:

$$\mu = \frac{\sum(f * x)}{N} = \sum\left(\frac{f * x}{N}\right) = \sum\left(x * \frac{f}{N}\right) = \sum[x * P(x)]$$

– Na fração *f/N*, o valor de *f* é a frequência com que ocorre o valor *x* e *N* é o tamanho da população, de modo que *f/N* é a probabilidade do valor de *x*.

– **Regra de arredondamento:** arredonde os resultados usando uma casa decimal a mais do que o número de casas decimais usadas para a variável aleatória *x*.

RESULTADO NÃO-USUAL COM REGRA EMPÍRICA

– A **regra empírica da amplitude** pode também ser útil na interpretação do valor de um desvio padrão.

– De acordo com esta regra, a maioria dos valores (resultados usuais) deve ficar a dois desvios padrões da média:

- – Valor usual máximo = $\mu + 2\sigma$
- – Valor usual mínimo = $\mu - 2\sigma$

EXEMPLO

Tabela 5-3 Calculando μ, σ e σ^2 para uma Distribuição de Probabilidade

x	$P(x)$	$x \cdot P(x)$	x^2	$x^2 \cdot P(x)$
0	0+	0,000	0	0,000
1	0+	0,000	1	0,000
2	0+	0,000	4	0,000
3	0+	0,000	9	0,000
4	0,001	0,004	16	0,016
5	0,003	0,015	25	0,075
6	0,016	0,096	36	0,576
7	0,053	0,371	49	2,597
8	0,133	1,064	64	8,512
9	0,236	2,124	81	19,116
10	0,283	2,830	100	28,300
11	0,206	2,266	121	24,926
12	0,069	0,828	144	9,936
Total		9,598		94,054

– **Média** = $\mu = \sum [x * P(x)] = 9{,}598 \approx 9{,}6$

– **Variância** = $\sigma^2 = \sum [x^2 * P(x)] - \mu^2 = 94{,}054 - 9{,}598^2 \approx 1{,}9$

– **Desvio padrão** = $\sigma = \sqrt{\sigma^2} = \sqrt{1{,}9} = 1{,}4$

– **Valor usual máximo** = $\mu + 2\sigma = 9{,}6 + 2(1{,}4) = 12{,}4$

– **Valor usual mínimo** = $\mu - 2\sigma = 9{,}6 - 2(1{,}4) = 6{,}8$

RESULTADO NÃO-USUAL ATRAVÉS DE PROBABILIDADE

– **Regra do evento raro:**
  – Se, sob uma dada hipótese (tal como a hipótese de que uma moeda seja honesta), a probabilidade de um evento particular observado (tal como 992 caras em 1000 lançamentos de uma moeda) é extremante pequena.
  – Concluímos que a hipótese provavelmente não é correta.

– Use probabilidades para determinar resultados não-usuais:
  – **Número de sucessos não usualmente alto:** x sucessos em n tentativas é um número não usualmente alto de sucessos se $P(x \text{ ou mais}) \leq 0{,}05$.
  – **Número de sucessos não usualmente baixo:** x sucessos em n tentativas é um número não usualmente baixo de sucessos se $P(x \text{ ou menos}) \leq 0{,}05$.

VALOR ESPERADO

– O **valor esperado** (esperança ou esperança matemática) de uma variável aleatória discreta:

– É o resultado médio teórico para um número infinito de tentativas.

– É o valor médio que esperaríamos se as tentativas pudessem continuar indefinidamente.

– É designado por E e representa o valor médio dos resultados.

– É obtido pelo cálculo de: $E = \sum [x * P(x)]$

– A **média** de uma variável aleatória discreta (μ) é o mesmo que o **valor esperado** (E).

DISTRIBUIÇÃO DE PROBABILIDADE BINOMIAL

DISTRIBUIÇÃO DE PROBABILIDADE BINOMIAL

– A distribuição de probabilidade binomial envolve proporções usadas com métodos de inferência estatística.

– Uma distribuição de probabilidade binomial resulta de um experimento que satisfaz os seguintes requisitos:

- – Experimento tem um **número fixo de tentativas**.
- – Tentativas têm que ser **independentes** (resultado de qualquer tentativa individual não afeta probabilidades de outras tentativas).
- – Cada tentativa deve ter todos resultados classificados em **duas categorias** (chamadas de sucesso e fracasso).
- – Probabilidade de **sucesso permanece constante** em todas tentativas.

NOTAÇÃO PARA DISTRIBUIÇÃO BINOMIAL

– ***S* e *F* (sucesso e fracasso):** representam as duas categorias possíveis de todos resultados.

– ***P*(*S*) = *p*:** probabilidade de um sucesso.

– ***P*(*F*) = 1 – *p* = *q*:** probabilidade de um fracasso.

– ***n*:** número fixo de tentativas.

– ***x*:** número específico de sucessos em *n* tentativas, de modo que *x* pode ser qualquer número inteiro entre 0 e *n*, inclusive.

– ***p*:** probabilidade de sucesso em uma das *n* tentativas.

– ***q*:** probabilidade de fracasso em uma das *n* tentativas.

– ***P*(*x*):** probabilidade de se obterem exatamente *x* sucessos em *n* tentativas.

INFORMAÇÕES IMPORTANTES

– Certifique-se de que *p* e *x* se refiram à mesma categoria designada como um sucesso.

– Na amostragem sem reposição, considere os eventos como independentes se o tamanho da amostra (*n*) não for maior do que 5% do tamanho da população (*N*).

– Há diferentes métodos para achar as probabilidades correspondentes à variável aleatória *x* em uma distribuição binomial:

  – Fórmula da probabilidade binomial.
  – Uso da tabela de probabilidades binomiais (pág. 615).
  – Uso de pacote estatístico.

FÓRMULA DA PROBABILIDADE BINOMIAL

– Em uma distribuição binomial, as probabilidades podem ser calculadas através da fórmula da probabilidade binomial:

$$P(x) = \frac{n!}{(n-x)!\,x!} * p^x * q^{n-x}, para\ x = 0, 1, 2, \dots, n$$

– *n* = número de tentativas.

– *x* = número de sucessos em *n* tentativas.

– *p* = probabilidade de sucesso em qualquer tentativa.

– *q* = probabilidade de fracasso em qualquer tentativa.

EXEMPLO DA FÓRMULA DA PROBABILIDADE BINOMIAL

– Qual a probabilidade de se obterem 7 jurados mexicanos-americanos quando 12 jurados são selecionados aleatoriamente de uma população que é 80% mexicana-americana?

– P(7) = ?

– n = 12

– x = 7

– p = 0,8

– q = 0,2

$$P(x) = \frac{n!}{(n-x)!\,x!} * p^x * q^{n-x}$$

$$P(7) = \frac{12!}{(12-7)!\,7!} * 0{,}8^7 * 0{,}2^{12-7}$$

$$P(7) = 0{,}0532$$

TABELA DE PROBABILIDADES BINOMIAIS (pág. 615)

– É preciso ter:
  – O número de tentativas (*n*).
  – A probabilidade de sucesso em qualquer tentativa (*p*).
  – O número de sucessos em *n* tentativas (*x*).

– Em seguida, é possível encontrar *P*(*x*) na tabela de probabilidades binomiais.

PACOTE ESTATÍSTICO

– Se você deseja encontrar P(x ou menos), sendo:

  – Número de tentativas: $n = 12$.

  – Número de sucessos em n tentativas: $x = 7$.

  – Probabilidade de sucesso: $p = 0{,}8$.

– No Stata, assim como no SPSS e SAS, o programa fornece probabilidades cumulativas de x ou menos sucessos com a função “*binomial*(n,x,p)”.

– P(7 ou menos) = 0,0725555

  di binomial(12,7,0.8)

FUNDAMENTOS PARA FÓRMULA DA BINOMIAL

– A fórmula da probabilidade binomial é a base para todos os métodos apresentados anteriormente.

– Esta fórmula é uma combinação da regra da multiplicação da probabilidade e a regra da contagem para o número de permutações de *n* itens quando *x* deles são idênticos ente si e os outros *n*–*x* são idênticos entre si.

$$P(x) = \overbrace{\frac{n!}{(n-x)!\,x!}}^{\text{Nº de resultados com exatamente } x \text{ sucessos entre } n \text{ tentativas}} \underbrace{* \, p^{x} * q^{n-x}}_{\text{Probabilidade de } x \text{ sucessos entre } n \text{ tentativas para qualquer ordem particular}}$$

MÉDIA, VARIÂNCIA E DESVIO PADRÃO PARA A DISTRIBUIÇÃO BINOMIAL

MÉDIA, VARIÂNCIA E DESVIO PADRÃO NA BINOMIAL

– Dada uma distribuição de probabilidade binomial, é possível encontrar a média, variância e desvio padrão para serem interpretados e compreendidos.

Medidas	**Qualquer distribuição de probabilidade discreta**	**Distribuição binomial**
Média	$\mu = \sum [x * P(x)]$	$\mu = np$
Variância	$\sigma^2 = \sum [x^2 * P(x)] - \mu^2$	$\sigma^2 = npq$
Desvio padrão	$\sigma = \sqrt{\sum [x^2 * P(x)] - \mu^2}$	$\sigma = \sqrt{npq}$

REGRA EMPÍRICA DA AMPLITUDE

– A regra empírica da amplitude pode ser útil ao considerar como não-usuais valores que caiam fora dos limites obtidos do seguinte:

– Valor usual máximo = $\mu + 2\sigma$

– Valor usual mínimo = $\mu - 2\sigma$

EXEMPLO DA FÓRMULA SIMPLIFICADA

– Encontre a média e desvio padrão para o número de mexicanos-americanos em um júri de 12 pessoas, selecionadas de população que é 80% mexicana-americana.

- μ = ?
- σ = ?
- n = 12
- p = 0,8
- q = 0,2

- $\mu = np = 12*0,8 = 9,6$
- $\sigma = \sqrt{npq} = \sqrt{12*0,8*0,2} \approx 1,4$
- Valor usual máximo = $\mu + 2\sigma = 9,6 + (2*1,4) = 12,4$
- Valor usual mínimo = $\mu – 2\sigma = 9,6 – (2*1,4) = 6,8$

DISTRIBUIÇÃO DE PROBABILIDADE DE POISSON

DISTRIBUIÇÃO DE PROBABILIDADE DE POISSON

– A distribuição de probabilidade de Poisson é importante porque é usada para se descrever o comportamento de eventos raros (com pequenas probabilidades).

– Esta é uma distribuição de probabilidade discreta que se aplica a ocorrências de eventos ao longo de intervalos especificados.

– A variável aleatória *x* é o número de ocorrências do evento no intervalo.

– O intervalo pode ser de tempo, distância, área, volume ou alguma unidade similar.

– Probabilidade de ocorrência do evento *x* em um intervalo é:

$$P(x) = \frac{\mu^{x} * e^{-\mu}}{x!}, onde\ e \approx 2{,}71828$$

REQUISITOS PARA DISTRIBUIÇÃO DE POISSON

– A variável aleatória *x* é o número de ocorrências de um evento ao longo de algum intervalo.

– As ocorrências devem ser aleatórias.

– As ocorrências devem ser independentes umas das outras.

– As ocorrências devem ser uniformemente distribuídas sobre o intervalo em uso

PARÂMETROS DA DISTRIBUIÇÃO DE POISSON

– Probabilidade de ocorrência do evento *x*:

$$P(x) = \frac{\mu^{x} * e^{-\mu}}{x!}, onde\ e \approx 2{,}71828$$

– Média na distribuição de Poisson:

$$\mu$$

– Desvio padrão na distribuição de Poisson:

$$\sigma = \sqrt{\mu}$$

POISSON $\neq$ BINOMIAL

– Uma distribuição de Poisson difere de uma distribuição binomial em alguns aspectos fundamentais.

– A distribuição binomial é afetada pelo tamanho *n* da amostra e pela probabilidade *p*, enquanto a distribuição de Poisson é afetada apenas pela média μ.

– Na distribuição binomial, os valores possíveis da variável aleatória *x* são 0, 1, ..., *n*. Porém, numa distribuição de Poisson, os valores possíveis de *x* são 0, 1, 2, ..., sem qualquer limite superior.

EXEMPLO DE POISSON

– Um total de 535 bombas caiu na área combinada de 576 regiões, em que cada área tem 0,25km^2. Se uma região é selecionada aleatoriamente, qual a probabilidade dela ser atingida duas vezes? Quantas das 576 regiões espera-se que sejam atingidas exatamente duas vezes?

– Nº médio de impactos = nº impactos / nº regiões = μ

$$\mu = 535 / 576 = 0{,}929$$

– Probabilidade de dois impactos em uma região:

$$P(x) = (\mu^x * e^{-\mu}) / x! = P(2) = (0{,}929^2 * 2{,}71828^{-0{,}929})/2! = 0{,}170$$

– Número das 576 regiões atingidas duas vezes é:

$$576 * 0{,}170 = 97{,}9$$

POISSON E PREVISÃO DE RESULTADOS

– A distribuição de Poisson pode ser muito boa para prever resultados que realmente ocorreram.

– Impactos de bombas V-1 em 576 regiões do sul de Londres:

Nº de impactos de bombas	Probabilidade	Nº esperado de regiões	Nº real de regiões
0	0,395	227,5	229
1	0,367	211,4	211
2	0,170	97,9	93
3	0,053	30,5	35
4	0,012	6,9	7
5	0,002	1,2	1

POISSON COMO APROXIMAÇÃO PARA BINOMIAL

– Algumas vezes, a distribuição de Poisson é usada para aproximar a distribuição binomial quando n é muito grande e a probabilidade p é pequena.

– Os requisitos para o uso da distribuição de Poisson como aproximação da distribuição binomial são:

$$n \geq 100$$

$$np \leq 10$$

– Se ambas as condições são satisfeitas, precisamos de um valor para a média μ, que pode ser calculada desta forma:

$$\mu = np$$

EXEMPLO DE POISSON COMO BINOMIAL

– Em uma loteria, suponha que você paga 1 dólar para escolher uma sequência de quatro dígitos, por exemplo, 2283. Se você joga esse jogo uma vez todos os dias, ache a probabilidade de ganhar exatamente uma vez em 365 dias.

- $P(1) = ?$
- Intervalo de tempo = $n = 365$
- Probabilidade de sucesso = $p = 1/10.000$
- Média = $np = 365 * (1/10.000) = 0,0365$
- Isso satisfaz as condições de $n \geq 100$ e $np \leq 10$

- $P(x) = (\mu^x * e^{-\mu}) / x!$
- $P(1) = (0,0365^1 * 2,71828^{-0,0365})/1!$
- $P(1) = 0,0352$

APLICAÇÃO DO CONHECIMENTO

PARA PENSAR

– Suponha que você queira calcular a probabilidade de encontrar estudantes do DCP ao caminhar pelo campus. Digamos que você conheça 50 estudantes no DCP. Há por volta de 30.000 estudantes na UFMG, embora alguns não caminhem com frequência pelo campus. Então, a probabilidade de encontrar, entre algum estudante aleatório, um destes estudantes do DCP é de aproximadamente 0,002. Nos dias que você poderia encontrar um destes alunos, você fica fora da FAFICH por volta de 30 minutos. Num dia típico, você poderia passar por 5 estudantes por minuto. Nestes dias, qual a probabilidade de que x destes estudantes seja um conhecido do DCP, onde x = 0, 1, 2, 3, 4, 5 ou maior que 5? Dado isto, você pensa que seria não-usual haver dias em que você encontre 1, 2, 3, 4, 5 ou >5 destes alunos?