MAT 461 – Tópicos de Matemática II
Aula 6: Resumo de Probabilidade

Edson de Faria

Departamento de Matemática
IME-USP

26 de Agosto, 2013

1. Variáveis Aleatórias e Distribuições Especiais
   - Variáveis Aleatórias Binomiais
   - Variáveis Aleatórias de Poisson

Distribuição binomial

- Como vimos, uma das mais importantes distribuições discretas de probabilidade é a distribuição binomial.
- Já ilustramos o uso da binomial na aula passada.
- Vamos iniciar esta aula com mais um exemplo de caráter prático.

Exemplo (1)

Uma companhia aérea oferece 1 *vôo diário NY-Boston a preços populares. Suas aeronaves têm capacidade máxima para* 150 *passageiros. Para se proteger de problemas de* **no-show**, *a companhia tem por estratégia de* **overbooking** *vender* 160 *passagens para este vôo. A experiência mostra que a probabilidade de um passageiro ser um* **no-show** *é de* 10%. *Assuma que os passageiros tomem suas decisões independentemente uns dos outros. Qual é a probabilidade de que a companhia deva indenizar alguns passageiros por overbooking nesse vôo?*

Solução:

- Podemos imaginar que cada passageiro toma a decisão de comparecer ou não para o embarque lançando uma moeda: se ocorrer cara, o passageiro não comparece; se ocorrer coroa, ele comparece.
- No caso, a moeda tem um viés: tem probabilidade 10% de resultar em cara, e 90% de resultar em coroa.

Distribuição binomial: exemplo 1

- Assim, como há 160 passageiros, temos a repetição de 160 experimentos de Bernoulli independentes, com probabilidade de sucesso $p = \frac{9}{10}$ (aqui, sucesso significa comparecimento do passageiro).
- Denotemos por X a variável aleatória que nos fornece o número de passageiros que comparecem para o vôo.
- Então X é binomial com parâmetros $n = 160$ e $p = \frac{9}{10}$.
- Nesta linguagem, o que o problema pede é $P[X > 150]$.
- Lembramos que, como X é binomial com os parâmetros especificados, temos:

$$P[X = k] \;=\; \binom{160}{k} \left\{ \frac{9}{10} \right\}^{k} \left\{ \frac{1}{10} \right\}^{160-k} \,.$$

- Portanto, temos:

$$P[X > 150] = \sum_{k>150} \binom{160}{k} \left\{\frac{9}{10}\right\}^k \left\{\frac{1}{10}\right\}^{160-k} .$$

- Esta soma é finita, a despeito das aparências: os únicos valores possíveis para X maiores do que 150 são $151, 152, \ldots, 160$.
- É fácil calcular o valor aproximado desta soma com o auxílio do computador, ou de uma calculadora. Veja a tabela na próxima página.
- Somando as probabilidades listadas naquela tabela, obtemos:

$$P[X > 150] \simeq 0.0359 .$$

- Em outras palavras: a probabilidade de que a companhia tenha que indenizar alguns passageiros é de pouco menos de 3.6%.

Tabela da distribuição binomial dos exemplos 1 e 2

k	$P[X = k]$
151	0.0185542
152	0.0098874
153	0.0046529
154	0.0019035
155	0.0006631
156	0.0001913
157	0.0000439
158	0.0000075
159	0.0000008
160	0.0000000(5)

Distribuição binomial: exemplo 2

Exemplo (2)

No exemplo 1, suponha que a companhia aérea pague uma indenização individual de 50 *dólares por passageiro vítima de overbooking. Qual o valor total das indenizações que a companhia espera ter que pagar por vôo?*

Solução:

- Denotemos por Y a variável aleatória que representa a indenização total paga pela companhia aérea num dado vôo.
- Então, temos

$$Y = \begin{cases} 0 & \text{se } X \leq 150 \\ 50(X - 150) & \text{se } X > 150 \end{cases}$$

Distribuição binomial: exemplo 2

- Assim, podemos escrever:

$$E[Y] = \sum_{k=151}^{160} 50(k-150)P[X=k]$$

- Utilizando novamente os valores da tabela dada, concluimos que:

$$E[Y] \simeq 3.24$$

- Ou seja, a companhia aérea espera ter que pagar por vôo, em média, apenas 3 dólares e 24 centavos a título de indenização.

Para os que têm a curiosidade de saber, aqui vai o código em Scilab para a função que fornece o cálculo da distribuição binomial (utilizada na resolução dos exemplos acima).

```
// Números combinatórios
function z = choose(n,k)
 z = 1
 for i = 1:k
 z = z*(n-i+1)/(k-i+1)
 end
endfunction

// Distribuição binomial
function y = binom(p,n,k)
 y = (p**k) * ((1-p)**(n-k)) * choose(n,k)
endfunction
```

Distribuição binomial: exemplo 3

- O exemplo clássico a seguir recebe o nome de "The newsboy problem". [1]
- Apresentaremos duas soluções deste exemplo: uma aqui utilizando a distribuição binomial, e outra mais adiante utilizando a distribuição de Poisson (a ser definida).

Exemplo (3)

Estamos de volta aos gloriosos anos cinqüenta. Um garoto vende jornais à porta da Grand Central Station, e tem em média 40 *clientes por dia, entre as cerca de* 2 000 *pessoas que passam por ele diariamente. Toda manhã, ele compra os jornais de um fornecedor por $ 0.50, e os vende por $ 0.75 (por unidade), mas não pode devolver os jornais não-vendidos ao fornecedor para obter um reembolso no final do dia. Quantos jornais ele deve comprar a fim de maximizar seus ganhos?*

[1] Referência: C. MacCluer, *Industrial Mathematics*, Prentice-Hall, 2000.

Distribuição binomial: exemplo 3

Solução:

- Cada uma das 2 000 pessoas que passa pelo garoto tem uma probabilidade $p = \frac{40}{2000} = 0.02$ de comprar um jornal.
- Assumimos que cada pessoa decide se compra ou não um jornal do garoto independentemente das outras.
- Em outras palavras, estamos diante de uma repetição de 2 000 experimentos de Bernoulli independentes, com a mesma probabilidade de sucesso $p = 0.02$.
- Seja X a variável aleatória que representa o número total de compradores em potencial ao final do dia. Ou seja, X representa o número de sucessos entre os 2 000 experimentos de Bernoulli em questão.
- Portanto, X é binomial, com parâmetros $n = 2\,000$ e $p = 0.02$.

Distribuição binomial: exemplo 3

- Em particular, sabemos que a distribuição de probabilidade de X é dada por:

$$P[X = k] = \binom{2\,000}{k} (0.02)^k (0.98)^{2000-k}$$

para $k = 0, 1, \ldots, 2000$

- A rigor, não estamos levando em conta ainda o fato de que o número de compradores não pode ser maior do que o número de jornais que o garoto tem para vender!
- Assim, seja m o números de jornais que o garoto compra do fornecedor pela manhã.
- E seja Y_m o lucro obtido pelo garoto ao final do dia.
- Qual é a relação entre Y_m e X?

Distribuição binomial: exemplo 3

- Esta relação é simples de calcular. Note que o lucro que o garoto obtém em cada jornal vendido é de 0.25 (isto é, 25 centavos).
- Portanto:

$$Y_m = \begin{cases} 0.75X - 0.50m & \text{se } X \leq m \\ 0.25m & \text{se } X > m \end{cases}$$

- Nosso objetivo agora é obter o valor esperado de Y_m, ou seja, o lucro esperado pelo garoto quando ele adota a estratégia de comprar m jornais do fornecedor.
- Temos:

$$E[Y_m] = \sum_{k=0}^{m} (0.75k - 0.50m)P[X = k] + 0.25mP[X > m] .$$

Distribuição binomial: exemplo 3

- Em seguida, queremos determinar qual é a estratégia que maximiza o lucro esperado do garoto, ou seja, queremos determinar para qual valor de m a esperança $E[Y_m]$ é máxima.
- Isto é um problema de máximo de uma função de uma variável. No caso, a variável é m. A função é $f(m) = E[Y_m]$, e como vimos no slide anterior, é uma função um tanto complicada de m.
- Assim, vamos resolver o problema de maneira computacional: vamos plotar a função $f(m)$ através de uma rotina em Scilab.
- O resultado gráfico é mostrado no próximo slide. Vemos que o valor de m que maxima o lucro esperado é $m = 37$. O lucro esperado correspondente a este valor de m é $\simeq$ 8.31 dólares.

Distribuição binomial: exemplo 3

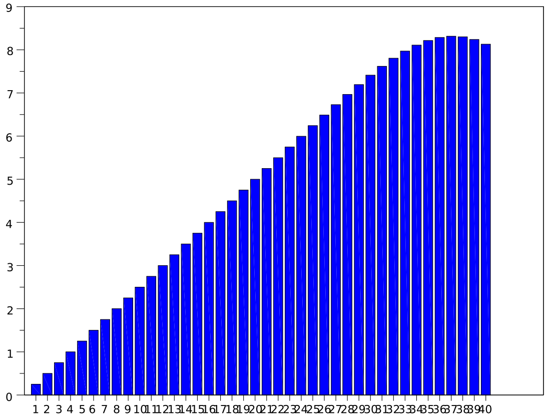

Figura 1: O “newsboy problem” com $n = 2000$ e $p = 0.02$.

Algoritmo em Scilab para o "newsboy problem"

Para aqueles que gostariam de saber como foi gerada a figura 1, aqui vai o código em Scilab. Este algoritmo utiliza a função `binom` introduzida anteriormente.

```
 // Algoritmo para o "newsboy problem"
function EY = newsboy(n,k)
p = k/n
for m = 1:k
e = -0.5 * m * binom(p,n,0)
for j = 1:m
e = e + (0.75*j - 0.5*m)*binom(p,n,j)
end;
for j = m+1:n
e = e + 0.25*m*binom(p,n,j)
end;
EY(1,m) = e;
end
plot(EY)
endfunction
```

Variáveis Aleatórias de Poisson: definição

- A distribuição de Poisson foi introduzida (de maneira indireta) em 1837 pelo matemático francês Siméon-Denis Poisson.
- Esta distribuição ocorre mais frequentemente como um limite da distribuição binomial, quando o número de repetições independentes dos experimentos de Bernoulli é muito grande e a probabilidade de sucesso é muito pequena.
- A distribuição de Poisson é usada na prática para modelar situações em que temos uma variável aleatória que assume apenas valores inteiros não-negativos, tipicamente indicando a contagem de ocorrências de algum evento relativamente raro (e.g. o número de "cliques" de um contador Geiger num intervalo de tempo fixado; o número de erros tipográficos por página de um livro).

Vejamos a definição formal.

Definição

Dizemos que uma função $X : \Omega \to \mathbb{N}$ *é uma variável aleatória de Poisson com parâmetro* $\lambda > 0$ *sobre um espaço de probabilidade* $(\Omega, \mathcal{A}, P)$ *se para todo* $k = 0, 1, 2, \ldots$ *temos*

$$P[X = k] \;=\; e^{-\lambda} \frac{\lambda^k}{k!} \;.$$

- Aqui, $e = 2.71828\ldots$ é a base dos logaritmos naturais.
- Observe que X assume infinitos valores (todos os inteiros não-negativos) com probabilidade positiva.
- A distribuição de probabilidades correspondente é chamada de distribuição de Poisson.

Variáveis Aleatórias de Poisson: definição

- Cumpre verificar que a definição dada é consistente. Em particular, precisamos verificar que a soma de todas as probabilidades da distribuição de Poisson é de fato igual a um, ou seja, que

$$\sum_{k=0}^{\infty} P[X = k] \;=\; 1 \,.$$

- Para isto, lembramos o seguinte fato do Cálculo, que é a fórmula de Taylor para a função exponencial (e^x):

$$e^x \;=\; 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^k}{k!} + \cdots$$

Variáveis Aleatórias de Poisson: definição

- Aplicando essa fórmula com $x = \lambda$, temos:

$$e^{\lambda} \;=\; 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots + \frac{\lambda^k}{k!} + \cdots$$

e multiplicando ambos os lados dessa igualdade por $e^{-\lambda}$, concluimos que

$$\begin{aligned} 1 \;&=\; e^{-\lambda} + e^{-\lambda}\lambda + e^{-\lambda}\frac{\lambda^2}{2!} + e^{-\lambda}\frac{\lambda^3}{3!} + \cdots + e^{-\lambda}\frac{\lambda^k}{k!} + \cdots \\ &= \sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} \;=\; \sum_{k=0}^{\infty} P[X = k] \;, \end{aligned}$$

como afirmado.

- Em seguida, perguntamos:

1. Qual é a esperança de X?
2. Qual é a variância de X?

Variáveis Aleatórias de Poisson: propriedades

- Para respondermos a essas perguntas, recorremos novamente à fórmula de Taylor para a função exponencial.
- Temos, por definição de esperança:

$$E[X] = \sum_{k=0}^{\infty} k\, P[X = k] = \sum_{k=1}^{\infty} k\, e^{-\lambda} \frac{\lambda^k}{k!}$$

$$= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$

- Mas:

$$\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \sum_{j=0}^{\infty} \frac{\lambda^j}{j!}$$

$$= 1 + \lambda + \frac{\lambda^2}{2!} + \cdots = e^{\lambda}$$

- Portanto, obtemos o valor da esperança de X, a saber:

$$E[X] = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda \,.$$

- De maneira inteiramente análoga, pode-se mostrar facilmente que:

$$E[X^2] = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = \lambda(\lambda+1)$$

- Disto resulta que a variância de X é dada por:

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2 = \lambda(\lambda+1) - \lambda^2 = \lambda \,.$$

- Resumindo, temos o seguinte resultado.

Proposição

A esperança e a variância de uma variável aleatória de Poisson com parâmetro λ são ambas iguais a λ.

Distribuição de Poisson: relação com a binomial

- Vamos agora esclarecer a relação entre a distribuição de Poisson e a distribuição binomial.
- Suponha que para cada n natural tenhamos uma variável binomial Y_n com parâmetros (n, p_n) tais que $np_n = \lambda$, para um certo $\lambda > 0$ fixo.
- Então podemos escrever (desde que $n > k$) que:

$$P[Y_n = k] \; = \; \binom{n}{k} p_n^k (1 - p_n)^{n-k} \; .$$

- Substituindo p_n por λ/n na expressão acima e desenvolvendo, obtemos:

$$P[Y_n = k] \; = \; \frac{n(n-1)\cdots(n-k+1)}{n^k} \cdot \frac{\left(1 - \dfrac{\lambda}{n}\right)^n}{\left(1 - \dfrac{\lambda}{n}\right)^k} \cdot \frac{\lambda^k}{k!}$$

Distribuição de Poisson: relação com a binomial

- Mantendo k fixo e fazendo $n \to \infty$, é possível mostrar que:

$$\lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} \;=\; 1$$

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n \;=\; e^{-\lambda}$$

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^k \;=\; 1$$

- Concluimos assim que:

$$\lim_{n\to\infty} P[Y_n = k] \;=\; e^{-\lambda}\frac{\lambda^k}{k!}$$

- Vemos assim que, no limite quando $n \to \infty$, as distribuições binomiais das variáveis Y_n se aproximam de uma distribuição de Poisson.
- Vale a pena enunciarmos este fato da seguinte maneira.

Proposição

Na repetição de um grande número n de experimentos de Bernoulli independentes com probabilidade de sucesso p muito pequena, o número total de sucessos (Y_n) possui uma distribuição aproximadamente igual a uma distribuição de Poisson com parâmetro $\lambda = np$.

Este fato possui grandes aplicações práticas em problemas de modelagem estatística, como veremos em alguns exemplos.

Comparando as duas distribuições

- Para ilustrarmos o conteúdo da proposição acima, consideremos o caso de uma variável aleatória binomial Y com parâmetros $n = 50$ e $p = 0.2$ (assim, Y representa o número de sucessos em 50 repetições de um experimento de Bernoulli com probabilidade de sucesso 0.2.
- Aqui, temos $\lambda = np = 50 \cdot 0.2 = 10$.
- Assim, comparamos a variável aleatória binomial Y com uma variável aleatória de Poisson X com parâmetro $\lambda = 10$.
- Os gráficos das duas distribuições, de Y e de X, encontram-se nos slides seguintes.
- Cumpre notar que o ajuste dos gráficos seria certamente melhor se tivéssemos tomado n maior e p menor (mas ainda com $np = 10$).

Distribuição de Poisson com $\lambda = 10$

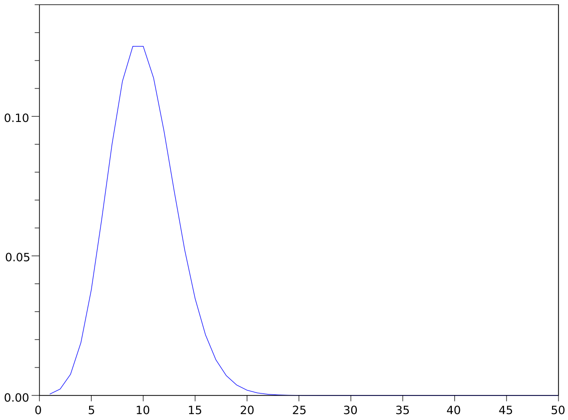

Figura 2: Distribuição de Poisson com $\lambda = 10$

Distribuição binomial com $n = 50$ e $p = 0.2$

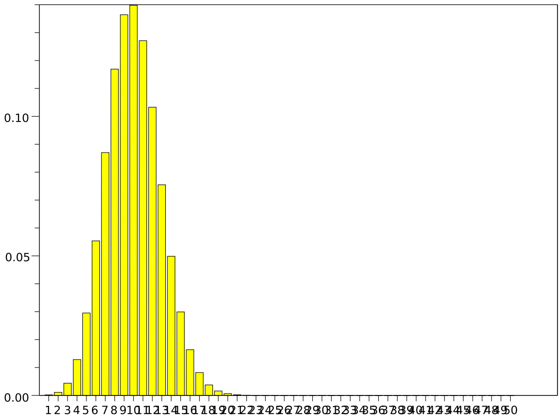

Figura 3: Distribuição binomial com $n = 50$ e $p = 0.2$

Superposição das duas distribuições

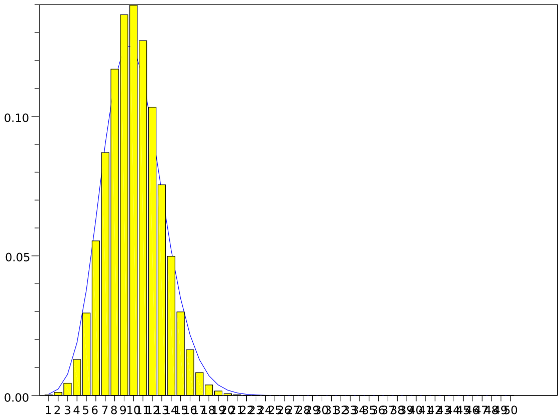

Figura 4: Poisson e binomial superpostas

Distribuição de Poisson: algoritmo em Scilab

Novamente, para os que têm curiosidade, aqui vai o algoritmo em Scilab utilizado para plotar a distribuição de Poisson das figuras anteriores.

```
 // Rotina para gerar e plotar a distribuição de Poisson
function P = poisson(l,m,i)
for k = 1:m
 P(1,k) = exp(-l)
  for j = 1:k
  P(1,k) = P(1,k)*l/j
  end
end
if  ( i == 0) then
 bar(P,'cyan')
elseif ( i == 1 ) then
 bar(P,'grey')
else
 plot(P)
end
endfunction
```

Distribuição de Poisson: exemplos

Vejamos agora alguns exemplos simples do uso da distribuição de Poisson.

Exemplo (4)

Suponha que o número de erros tipográficos por página de um livro de Estatística siga uma distribuição de Poisson com parâmetro $\lambda = 1/4$. *Calcule a probabilidade de que haja pelo menos um erro tipográfico na página* 100 *do livro.*

Solução:

- Seja X a variável aleatória que fornece o número de erros tipográficos na página 100.
- Por hipótese, a variável aleatória X é Poisson, com $\lambda = 1/4$.
- Logo:

$$P[X \geq 1] \;=\; 1 - P[X = 0] \;=\; 1 - e^{-1/4} \simeq \; 0.221 \;.$$

ou seja, a probabilidade desejada é de 22.1%.

Distribuição de Poisson: exemplos

Exemplo (5)

Uma loja atende em média 60 *clientes por hora. Qual é a probabilidade de que apareçam na loja mais do que* 20 *clientes num dado intervalo de* 15 *minutos (por exemplo, entre* $3:00$ *e* $3:15$*)?*

Solução:

- Seja X a variável aleatória que fornece o número de clientes que aparecem na loja entre $3:00$ e $3:15$.
- Estamos supondo aqui que X é uma variável aleatória de Poisson.
- Como o número médio de clientes que entram na loja por hora é 60, o número médio de clientes que visitam a loja num dado intervalo de 15 minutos ($= \frac{1}{4}$ de hora) é $\frac{1}{4} \times 60 = 15$.
- Logo, X é Poisson com parâmetro $\lambda = 15$.

- Em particular, temos:

$$P[X = k] \;=\; e^{-15}\,\frac{15^k}{k!} \quad , k = 1, 2, \ldots$$

- O problema pede $P[X > 20]$, a probabilidade de que mais do que 20 clientes visitem a loja entre as $3:00$ e as $3:15$.
- A rigor, teríamos que calcular a soma infinita:

$$P[X > 20] \;=\; e^{-15}\sum_{k=21}^{\infty}\frac{15^k}{k!}$$

- Mas é mais fácil perceber que:

$$P[X > 20] \;=\; 1 - P[X \leq 20] \;=\; 1 - e^{-15}\sum_{k=0}^{20}\frac{15^k}{k!}$$

Distribuição de Poisson: exemplos

- Esta última soma, que é finita, é facilmente obtida por meio de uma calculadora, ou através de um pequeno algoritmo em Scilab (ou Matlab, ou Maple, ou...).
- O resultado final: $P[X > 20] \simeq 0.083$, ou seja, 8.3%.
- Vejamos agora como a distribuição de Poisson pode ser utilizada para resolver o "newsboy problem" visto anteriormente.

Distribuição de Poisson: Revisitando o "newsboy problem"

- Como o garoto vendedor de jornais tem em média 40 potenciais compradores por dia, podemos modelar o problema original utilizando, ao invés de uma variável binomial, uma variável aleatória de Poisson com parâmetro $\lambda = 40$.
- Esta variável representa o número total de compradores ao final do dia.
- Neste caso, o valor esperado do lucro obtido pelo garoto com a venda dos jornais é:

$$E[Y_m] = \sum_{k=0}^{m} (0.75k - 0.5m) \frac{e^{-40} 40^k}{k!} + \sum_{k=m+1}^{\infty} 0.25m \frac{e^{-40} 40^k}{k!}$$

- E, como antes, o que queremos é maximizar esta expressão como função de m.

Distribuição de Poisson: Revisitando o "newsboy problem"

- Uma cálculo simples (exercício) mostra que:

$$E[Y_{m+1}] - E[Y_m] \; = \; 0.25 - 0.75 \sum_{k=0}^{m} \frac{e^{-40} 40^k}{k!}$$

- Note que esta diferença é claramente uma função decrescente de m.
- O valor de m que corresponde ao máximo desejado é justamente o momento em que a diferença acima fica negativa.
- Este valor de m pode ser facilmente calculado, novamente utilizando-se uma rotina em Scilab (ou em sua linguagem de programação favorita).
- O valor que se obtém dessa forma é $m = 37$, como anteriormente.