Probabilidade e Estatística

Aula 5

Probabilidade: Distribuições de Discretas – Parte 2

Leitura obrigatória:

Devore, seções 3.4, 3.5 (hipergeométrica), 3.6

Objetivos

Nesta parte 01 aprendemos a representar, de forma geral, o modelo de probabilidade de uma variável aleatória discreta.

Na parte 02, estudamos alguns casos particulares de variáveis aleatórias discretas e suas distribuições, que chamaremos de famílias de distribuições discretas. Os casos particulares são:

- binomial, Poisson e hipergeométrica

Distribuições de Probabilidade

Distribuições de Probabilidade

- Distribuições de Probabilidade **Discretas**
  - Binomial
  - Hipergeométrica
  - Poisson
- Distribuições de Probabilidade **Contínuas**
  - Uniforme
  - Normal
  - Exponencial

Famílias e Parâmetros

Exemplo de família de varíavel aleatória Bernoulli: Um experimento pode ter dois resultados, S = {S,F} em que S é um sucesso e F um fracasso.

- A variável aleatória de Bernoulli é definida como:
  - $X(S) = 1 \; e \; X(F) = 0$
- A fmp de X pode ser:
  - $p(1) \; = \; 0.3 \; e \; p(0) \; = \; 0.7$
  - $p(1) \; = \; 0.9 \; e \; p(0) \; = \; 0.1$
  - …
  - De uma forma geral:

$\boldsymbol{x}$	0	1
$\boldsymbol{p(x;\alpha)}$	$(1-\alpha)$	α

Famílias e Parâmetros

- A função massa de probabilidade (fmp) de uma v.a. X de Bernoulli depende de uma quantidade que assume diferentes valores (α no exemplo anterior).

x	0	1
p(x;α)	(1-α)	α

  - Chamamos esta quantidade de **parâmetro.**
  - O conjunto de todas as fmp que podemos obter ao variarmos os valores dos parâmetros é chamada de **família** de distribuições. No exemplo acima, $p(x;\ \alpha)$ é a família de distribuições de Bernoulli.

Definição!

Definição!

Binomial

- A família de distribuições **binomial** é bastante usada na prática, quando queremos contar o número de sucessos em uma amostra de tamanho fixo.

- Aplicações:
  - Uma fábrica classifica produtos como defeituosos ou aceitáveis em um lote de 100 produtos. Qual a probabilidade de existirem menos de 10 itens defeituosos neste lote?
  - Uma firma em leilões por 10 contratos ganha ou não ganha o contrato referente a cada leilão. Qual é a probabilidade de a firma ganhar ao menos 3 contratos?
  - 6 candidatos entrevistados para vaga aceitam ou rejeitam oferta de emprego. Qual é a probabilidade de ao menos 1 candidato aceitar a oferta?

Binomial

Condições para podermos usar a binomial:

1. Experimento: **sequência de n** experimentos menores denominados **tentativas**, onde **n é estabelecido antes do experimento**.
2. **Cada tentativa** pode resultar em um de **dois resultados possíveis**, chamados de sucesso (S) ou fracasso (F).
3. **As tentativas são independentes**, de forma que o resultado de qualquer tentativa particular não influência o resultado de qualquer outra tentativa.
4. **A probabilidade de sucesso é constante** de uma tentativa para outra. Denominamos essa probabilidade **p = P(S)**.
5. Estamos interessados no número de Sucessos nas **n** tentativas!

Como calcular a função massa de probabilidade de uma binomial?? Isto é, qual é a probabilidade de exatamente x sucessos em n tentativas?

Binomial: técnicas de contagem

Exemplo da aula de exercícios: Selecionamos 5 espécimes aleatoriamente em um laboratório. Cada espécime tem 10% de probabilidade de estar contaminado. A contaminação entre os espécimes são idenpendentes.

Queremos saber a probabilidade de "exatamente 2 espécimes contaminados em 5".

- Seja S =espécime contaminado.
- X =número de S's (contaminados) em n (5) espécimes na amostra.
- A probabilidade de sucesso é constante: P(S) = p =0.1
- O total de tentativas é: n=5. E cada tentativa é independente.

Binomial: técnicas de contagem

Exemplo da aula de exercícios:

- Queremos a probabilidade de $X = 2$:
  - Exemplos de possibilidades com 2 contaminadas em 5: CCNNN, CNCNN, NNCNC,…
  - Pela independência e probabilidade de contaminação constante: cada possibilidade tem probabilidade: $0.1^2(0.9)^3$
  - Qual é o nº total de possibilidades?
    - Temos um conjunto de 5 espécimes para escolher, sem reposição, 2 como contaminadas: {1,2,3,4,5}. A ordem não é importante, pois tanto faz selecionar a 1ª e a 3ª para serem contaminadas como a 3ª e a 1ª.
    - Então queremos determinar o nº de combinações de 5, selecionadas 2 a 2:
    - Lembrando que:

$$n^{\text{º}}\ de\ combinações\ de\ n\ k\ a\ k = \binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$

  - Então: $P(X = 2) = \binom{5}{2} 0.1^2(0.9)^3 = \frac{5!}{3!2!} 0.1^2(0.9)^3$

Binomial: fmp

Para X = n° de S's em n tentativas $\sim Bin(n, p)$:

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

para $x = 0,1,2, \dots, n$.

Em que:

$\boldsymbol{p(x)}$ = **probabilidade de x sucessos em n tentativas.**

n = número de tentativas (tentativas independentes)

x = número de "sucessos" observados

p =probabilidade de "sucesso" em cada tentativa (constante)

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$: número de combinações de n elementos k a k.

Lembrando que 0!=1

Binomial: fmp

Para X = n° de S's em n tentativas $\sim Bin(n,p)$:

$$p(x) = P(X = x) = \boxed{\frac{n!}{x!\,(n-x)!}}\,\boxed{p^x(1-p)^{n-x}}$$

para $x = 0,1,2,\ldots,n$.

Nº de formas de obtermos x sucessos em amostra de tamanho n.
Ex: $n = 3, x = 2$.
$S \cap S \cap F$ ou $S \cap F \cap S$ ou $F \cap S \cap S$
Temos ($\frac{3!}{2!1!} = 3$) possibilidades de colocar 2 sucessos em uma amostra de tamanho 3.

Probabilidade de cada uma das formas ocorrer. Usa independência e p constante.
Ex: n=3, x=2.

$$P(S \cap S \cap F) = P(S \cap F \cap S) =$$
$$= P(F \cap S \cap S) = p^2(1-p)^1$$

Binomial

Exercício: Se a probabilidade de comprarmos um computador com defeito é de 0.02, qual é a probabilidade de comprarmos 2 computadores com defeito em um lote de 10 computadores? Suponha que os computadores foram produzidos de maneira independente.

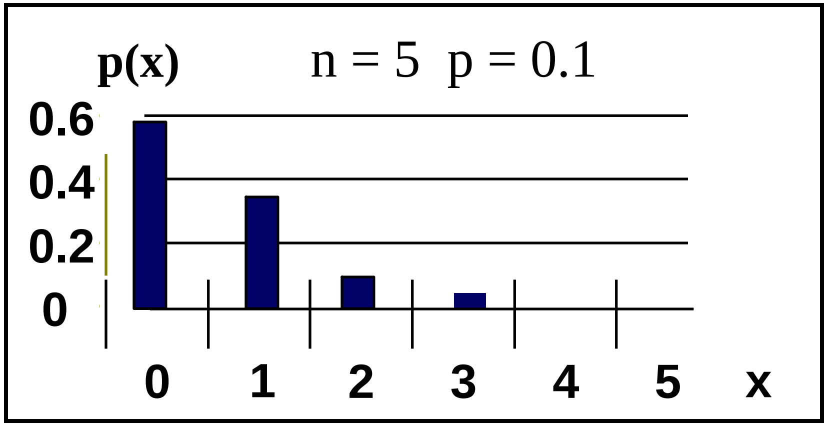
Solução: Seja X o nº de computadores com defeito no lote de 10 computadores. Assim, $p = 0.02$, $n = 10$, e queremos determinar $P(X = 2) = p(2)$ com $X \sim bin(n = 10, p = 0.02)$.

$$p(2) = \frac{10!}{2!\,8!} 0.02^2 0.98^8 = 0.0153$$

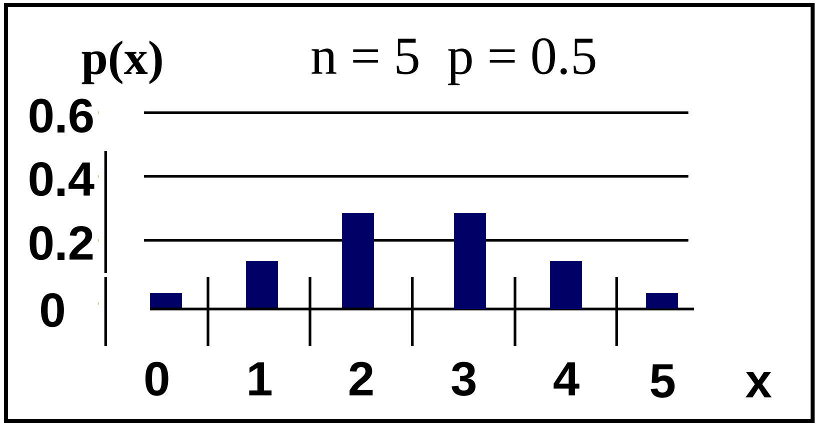
Formato da Distribuição Binomial

O formato da distribuição binomial depende dos valores dos parâmetros n e p:

- $n = 5\ e\ p = 0.1$



- $n = 5\ e\ p = 0.5$



Binomial: propriedades

Propriedade: Usando as definições de média e variância de uma variável aleatória, podemos obter para $X \sim Bin(n, p)$:

$$\boldsymbol{\mu = E(X) = \sum_{x=0}^{n} x \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x} = np}$$

e

$$\boldsymbol{\sigma^2 = E\left[(X-\mu)^2\right] = \sum_{x=0}^{n} (x-np)^2 p(x) = np(1-p)}$$

Em que $\boldsymbol{n}$ = número de tentativas (tamanho da amostra)

$\boldsymbol{p}$ = probabilidade de sucesso em cada tentativa

$(\boldsymbol{1-p})$ = probabilidade de fracasso em cada tentativa

Binomial

Exercício: Se a probabilidade de comprarmos um computador com defeito é 0.02, então:

a) Qual é o número esperado de computadores com defeito em um lote de 10 computadores?

b) Qual é o desvio-padrão do número de computadores com defeito em um lote com 10 computadores?

Solução: Para $X \sim bin(n = 10, p = 0.02)$, temos:

$$E(X) = np = 10 * 0.02 = 0.2$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{10 * 0.02 * 0.98} = 0.44$$

Binomial

Exercício: As linhas telefônicas em um sistema de reservas de uma companhia aérea estão ocupadas 40% do tempo. Suponha que os eventos em que as linhas estejam ocupadas em sucessivas chamadas sejam independentes. Considere que 10 chamadas aconteçam.

a) Qual a probabilidade de, para exatamente 3 chamadas, as linhas estarem ocupadas?
b) Qual a probabilidade de as linhas estarem ocupadas em no mínimo uma das chamadas?
c) Qual é o número esperado de chamadas em que as linhas estejam ocupadas?

Hipergeométrica

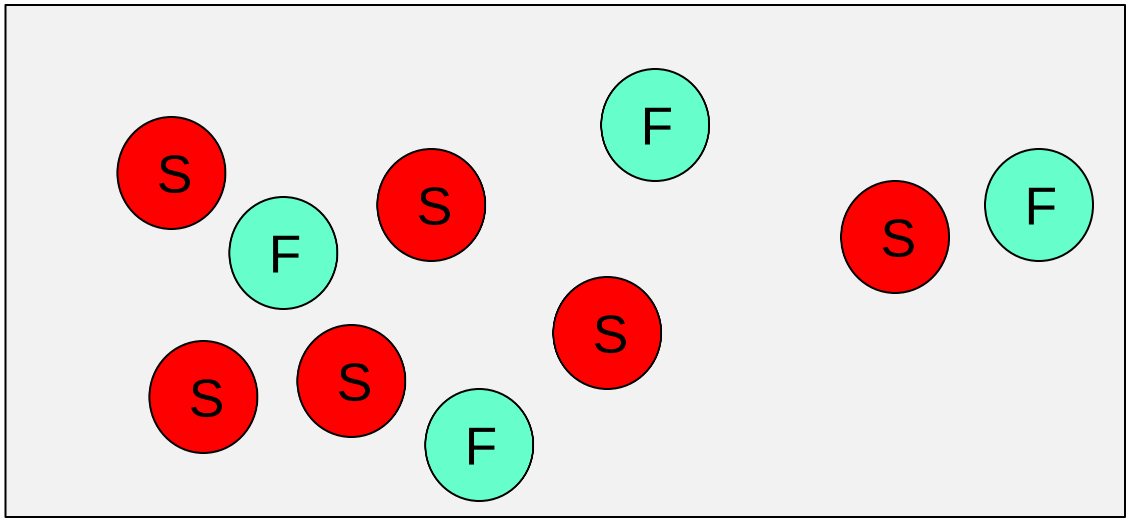
Experimento:

- "$\boldsymbol{n}$" itens selecionados ***sem reposição*** de uma população de tamanho $\boldsymbol{N}$, ou seja, pega-se uma amostra de tamanho $\boldsymbol{n}$ desta população.
- Cada indivíduo da população é classificado como "Sucesso " ou Fracasso".
- A população possui $\boldsymbol{M}$ "Sucessos".

X = número de "Sucessos" na amostra de tamanho n

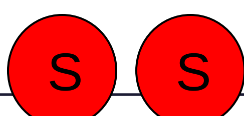
Como calcular a função massa de probabilidade de uma Hipergeométrica?? Isto é, qual é a probabilidade de x sucessos em amostra de tamanho n selecionada sem reposição?

Hipergeométrica

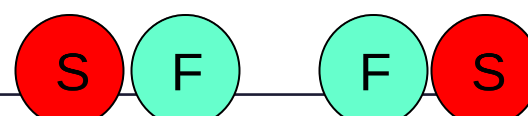


Neste exemplo, temos: 6 Sucessos na população, 4 fracassos na população. Portanto, $N = 10$. Seleciona-se uma amostra de tamanho 2:

Qual é a probabilidade de sortearmos 2 S em $n = 2$? p(2)=???



Qual é a probabilidade de sortearmos 1 S em $n = 2$? p(1)=???



Hipergeométrica

Qual é a diferença entre uma binomial e uma hipergeométrica?

Ambas contam o nº de sucessos em uma amostra….

- Como vimos, para a binomial:
  - probabilidade de sucesso em cada tentativa, $\boldsymbol{p}$, é independente dos itens selecionados nas outras tentativas e é constante. Isto pode ser obtido por:
    - amostra *com reposição* a partir de população *finita*.
    - amostra *sem reposição* a partir de população *infinita (se* $n \leq 0.1N$).
- No experimento da hipergeométrica:
  - amostra *sem reposição* a partir de população *finita*.

ou seja, a probabilidade de um item da amostra depende dos itens que já foram sorteados e não é constante!

Hipergeométrica: fmp

Para X $=$ n° de S's em amostra de tamanho n sem reposição $\sim \boldsymbol{h}(\boldsymbol{n}, \boldsymbol{M}, \boldsymbol{N})$:

$$\boldsymbol{p}(\boldsymbol{x}) = \frac{\binom{\boldsymbol{M}}{\boldsymbol{x}}\binom{\boldsymbol{N}-\boldsymbol{M}}{\boldsymbol{n}-\boldsymbol{x}}}{\binom{\boldsymbol{N}}{\boldsymbol{n}}}$$

para um inteiro x tal que: $\max\{0, n - N + M\} \leq x \leq \min\{n, M\}$

Em que:

$\boldsymbol{N}$ $=$ tamanho da população

$\boldsymbol{M}$ $=$ nº de sucessos na população

$\boldsymbol{N} - \boldsymbol{M}$ $=$ nº de fracassos na população

$\boldsymbol{n}$ $=$ tamanho da amostra

$\boldsymbol{x}$ $=$ nº de sucessos na amostra

$\boldsymbol{n} - \boldsymbol{x}$ $=$ nº de fracassos na amostra

Hipergeométrica: fmp

Para X = n° de S's em amostra de tamanho n sem reposição $\sim \boldsymbol{h(n, M, N)}$:

$$\boldsymbol{p(x)} = \frac{\boxed{\binom{\boldsymbol{M}}{\boldsymbol{x}}}\boxed{\binom{\boldsymbol{N-M}}{\boldsymbol{n-x}}}}{\boxed{\binom{\boldsymbol{N}}{\boldsymbol{n}}}}$$

para um inteiro x tal que: $\max\{0, n - N + M\} \leq x \leq \min\{n, M\}$

Nº de combinações para obter (x) sucessos em amostra de tamanho n, partindo de população com M sucessos.

Nº de combinações para obter $(n - x)$ fracassos em amostra de tamanho n, partindo de população com $N - M$ fracassos.

Nº total de formas para obter amostra de tamanho n, partindo de população de tamanho N

Hipergeométrica: fmp

Para $X = $ n° de S's em amostra de tamanho n sem reposição $\sim \boldsymbol{h(n, M, N)}$:

$$\boldsymbol{p(x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}}$$

para um inteiro x tal que: $\max\{0, n-(N-M)\} \leq x \leq \min\{n, M\}$

Se existirem **poucos fracassos** na população, este nº limita o mínimo de sucessos que podemos obter na amostra. Ex: N=100, n=20, M=95 $\Rightarrow N - M = 5$. O menor valor possível para x é 15 (todos os 5 fracassos na amostra) e não 0.

Se existirem **poucos sucessos** na população, este nº limita o máximo de sucessos que podemos obter na amostra. Ex: $N = 100, n = 20, M = 3$. O maior valor possível para x é 3 e não 20.

Hipergeométrica: propriedades

Propriedade: Usando a definição de média e variância de uma v.a. e a fmp da v.a. hipergeométrica, temos:

- A média de uma v.a. $X \sim hipergeométrica($

$$\boldsymbol{\mu = E(X) = \boxed{\frac{nM}{N}}}$$

$\frac{M}{N}$ é a proporção de sucessos na população. É equivalente ao p da binomial.

e o desvio-padrão é dado por:

$$\boldsymbol{\sigma = \sqrt{\boxed{n\frac{M}{N}\frac{(N-M)}{N}}} * \boxed{\sqrt{\frac{N-n}{N-1}}}}$$

Fator de correção de pequenas amostras. Vejam que se $N \gg n$, este termo tende a 0 e o desvio tende ao da binomial.

$\frac{M}{N}$ e $\frac{N-M}{N}$ são a proporção de sucessos e fracassos na população, respectivamente. Este termo é igual ao da binomial: $np(1-p)$

Hipergeométrica: exercício

Exercício: Em um departamento existem 10 computadores diferentes. Destes, 4 tem programas ilegais instalados.

A equipe de informática decide inspecionar 3 computadores aleatoriamente.

Qual a probabilidade de que 2 dos 3 computadores inspecionados tenham programas ilegais instalados?

Hipergeométrica: exercício

Exercício: Solução. Amostragem sem reposição.

Seja X=nº de computadores com programas ilegais em amostra de 3 computadores selecionados.

$X \sim hipergeométrica(N = 10, n = 3, M = 4).$

Queremos determinar $P(X = 2)$:

$$p(2) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = 0.3.$$

A probabilidade de que 2 de 3 computadores tenham programas ilegais instalados é de 0.3 ou seja, 30%.

Poisson

Definição!

- Uma **área de oportunidade** é uma *unidade contínua* (um intervalo de tempo, volume ou área) na qual podem ocorrer mais de um *evento discreto*.
  - ex: número de carros que passam em sinal em uma determinada hora.
  - ex: O número de arranhões na pintura do carro.
  - ex: O número de mordidas de mosquito em uma pessoa.
  - ex: O número de vezes que o computador trava em um dia.
  - ex: número de pepitas de chocolate em cookie

Poisson

Aplique a distribuição de Poisson quando:

1. Deseja-se saber a probabilidade do número de vezes que um evento pode ocorrer em uma área de oportunidade.
2. Probabilidade de um evento ocorrer em uma área de oportunidade é a *mesma para áreas de mesmo tamanho*.
3. O número de eventos que ocorre em uma área de oportunidade é *independente* do número de eventos que ocorrem em outras áreas de oportunidade.
4. A *probabilidade* de dois ou mais eventos acontecerem em uma área de oportunidade se *aproxima de zero a medida que a área fica menor.*
5. O número médio de eventos por área é dado por λ (lambda)

Relação entre Poisson e Binomial

Exercício: Observa-se a esquina de uma rua pouco movimentada. Sabemos que em média passam 2.7 carros/hora. Qual a probabilidade de passarem 4 carros em uma hora?

Relação entre Poisson e Binomial

- Exercício: Solução aproximada por binomial:
  - 1 hora = 60 minutos.
  - Seja S= "carro passa em um dado minuto". $P(S) = 2.7/60 = 0.045$.
  - Suponha que a probabilidade de passar um carro em um dado minuto é independente dos demais minutos.
  - Seja X =n° de minutos em 60 que observamos um carro passar.

$$P(X = 4) = \frac{60!}{56!\,4!} 0.045^4 (1 - 0.045)^{56} = 0.1517518$$

Mas, e se mais de um carro passar em um certo minuto?

Relação entre Poisson e Binomial

- Exercício: Solução aproximada por binomial:
  - 1 hora = 3600 segundos.
  - Seja S: carro passa em um dado segundo. $P(S) = 2.7/3600 = 0.00075$.
  - Suponha que a probabilidade de passar um carro em um dado minuto é independente dos demais minutos.
  - Seja X =n° de segundos em 3600 que observamos um carro passar.

$$P(X = 4) = \frac{3600!}{3596!\,4!} 0.00075^4 (1 - 0.00075)^{3596} = 0.1488635$$

Mas e se mais de um carro passa em um certo segundo?

Relação entre Poisson e Binomial

- Exercício: Solução aproximada por binomial:
  - 1 hora $= n$ unidades muito pequenas.
  - Seja S: carro passa em um dado segundo. $P(S) = 2.7/n$.
  - Seja X =n° de unidades em n que observamos um carro passar.

Qual a probabilidade de obsevarmos 4 unidades pequenas de tempo com um carro passando em 1 hora?

$$\lim_{n\to\infty} P(X=4) = \lim_{n\to\infty} \frac{n!}{(n-4)!\,4!}\left(\frac{\lambda}{n}\right)^4\left(1-\frac{\lambda}{n}\right)^{n-4}$$

Relação entre Poisson e Binomial

De uma forma geral, para qualquer nº de ocorrências do evento em uma certa área de oportunidade:

$$\lim_{n\to\infty} P(X=x) = \lim_{n\to\infty} \frac{n!}{(n-x)!\,x!}\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

Pode-se mostrar que:

$$\lim_{n\to\infty} P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

Poisson: fmp

Seja X: n° de ocorrências por área de oportunidade ~ $Poisson(\lambda)$:

$$\boldsymbol{p(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}}$$

para $x = 0,1,2,3, \ldots$

Em que:

$p(x)$ = probabilidade de x ocorrências do evento na área de oportunidade.

λ (parâmetro da distribuição) = número médio de eventos por área de oportunidade.

CUIDADO: converter unidade de parâmetro λ para a mesma unidade da área de oportunidade!!!!

Poisson: propriedades

Propriedades: Podemos obter a média e a variância de uma v.a. X ~ Poisson(λ) através de sua aproximação pela $Y \sim bin(n, p)$.

- Fazendo n→∞ e p→0 de tal forma que $E(Y) = np \rightarrow \lambda > 0$.

Assim:

- O valor esperado da Poisson, μ_X, é o limite do valor esperado da binomial:

$$\mu_X = \lim_{n\to\infty} E(Y) = \lim_{n\to\infty} np = \lambda$$

- A variância da Poisson, $V(X)$, é o limite da variância da binomial:

$$\mathrm{V(X)} = \lim_{n\to\infty} V(Y) = \lim_{n\to\infty} np(1-p) = \lambda * 1 = \lambda$$

$$\therefore \boldsymbol{E(X) = V(X) = \lambda}$$

Relação entre Poisson e Binomial

Exercício: Observa-se a esquina de uma rua pouco movimentada. Sabemos que em média passam 2.7 carros/hora. Qual a probabilidade de passarem 4 carros em uma hora?

Seja X= nº de carros que passam em 1 hora, então X tem distribuição de Poisson com $E(X) = \lambda = 2.7$ carros/hora (assumindo que todas as condições para a distribuição de Poisson são válidas).

$$P(X = 4|\lambda = 2.7) = \frac{e^{-2.7}2.7^4}{4!} = 0.1488157$$

Compare este valor com as aproximações da binomial obtidas anteriormente. Veja que o cálculo da probabilidade usando aproximação de X~binom(n=3600, p=0.00075) só difere na quinta casa decimal.

Poisson: exercício

Exercício: Suponha que o nº de carros que entra em um estacionamento em um certo minuto tem distribuição de Poisson e, na média, 5 carros entrem um no estacionamento por minuto.

Qual a probabilidade de 7 carros entrarem no estacionamento em um dado minuto?

Poisson: exercício

Exercício: Solução.

- Seja X= nº de carros que entram no estacionamento por minuto. Sabemos que $X \sim Poisson\left(\lambda = 5\,\frac{carros}{minuto}\right)$.
- Queremos calcular $P(X = 7)$:

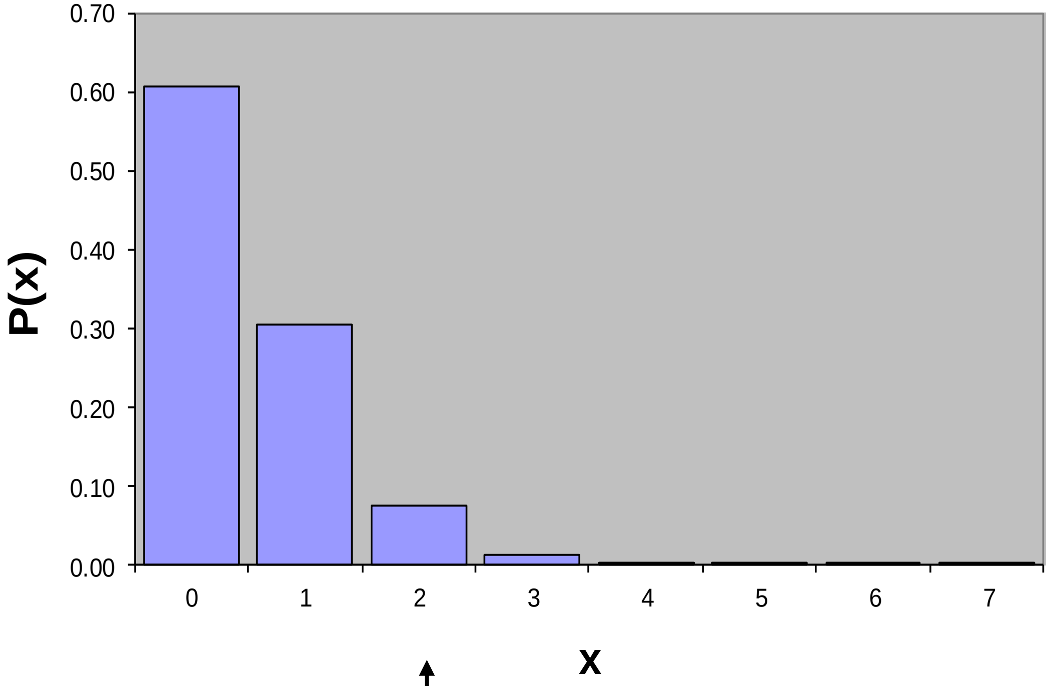
$$p(7) = \frac{e^{-\lambda}\lambda^{x}}{x!} = \frac{e^{-5}5^{7}}{7!} = 0.104$$

- Então, existe uma probabilidade igual 10.4% de 7 carros entrarem no estacionamento no próximo minuto.

Poisson: formato

λ = 0.50

x	**P(x)**
0	0.6065
1	0.3033
2	0.0758
3	0.0126
4	0.0016
5	0.0002
6	0.0000
7	0.0000

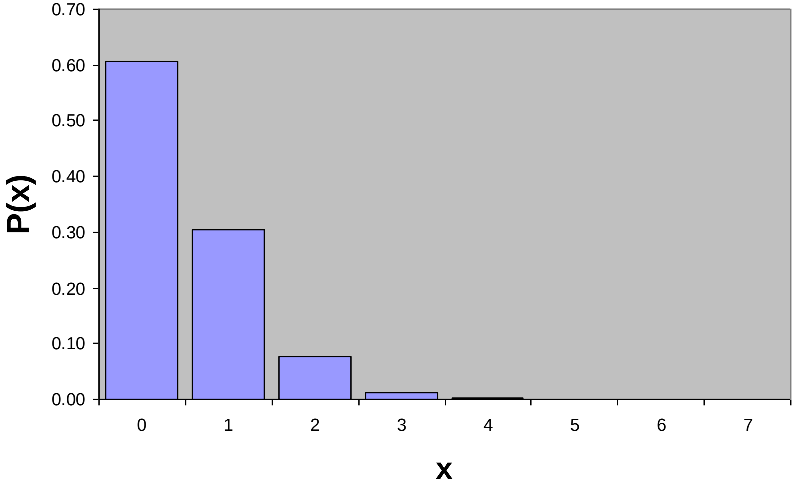


P(x = 2) = 0.0758

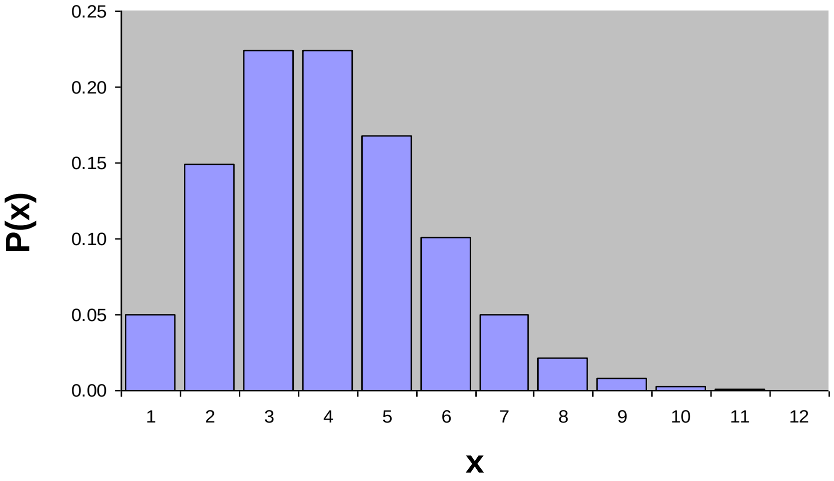
Poisson: formato

- O formato da distribuição de Poisson depende do parâmetro λ :

$\lambda = 0.50$



$\lambda = 3.00$



Poisson: exercício

Exercício: O número de falhas nas máquinas da indústria têxtil segue a distribuição de Poisson com uma média de 0.1 por m².

a) Qual é a probabilidade de que haja duas falhas em 1 m² de tecido?
b) Qual é a probabilidade de que haja uma falha em 10 m² de tecido?
c) Qual é a probabilidade de que não haja falha em 20 m² de tecido?
d) Qual é a probabilidade de que haja no minimo duas falhas em 10 m² de tecido?

Funções no Excel

- **Binomial**
  - fmp: =DISTR.BINOM(x; n; p; Falso)
  - fda: =DISTR.BINOM(x; n; p; Verdadeiro)

- **Hipergeométrica**
  - fmp: =DIST.HIPERGEOM.N(x;n;M;N;Falso)
  - fda: =DIST.HIPERGEOM.N(x;n;M;N;Verdadeiro)

- **Poisson**
  - fmp: =DISTR.POISSON(x; λ; Falso)
  - fda: =DISTR.POISSON(x; λ; Verdadeiro)

Comandos em R

- **Binomial**
  - fmp: dbinom(x, n, p)
  - fda: pbinom(x, n, p, lower.tail = TRUE)
  - inversa da fda: qbinom(prob, n, p, lower.tail = TRUE)

- **Hipergeométrica**
  - fmp: dhyper(x, M, N-M, n)
  - fda: phyper(x, M, N-M, n, lower.tail = TRUE)
  - inversa da fda: qhyper(prob, M, N-M, n, lower.tail = TRUE)

- **Poisson**
  - fmp: dpois(x, lambda)
  - fda: ppois(x, lambda, lower.tail = TRUE)
  - inversa da fda: qpois(prob, lambda, lower.tail = TRUE)

Resumo

Nesta aula vimos as seguintes distribuições discretas:

- Binomial
- Hipergeométrica
- Poisson

- Para cada uma destras distribuições, aprendemos a identificar quando podemos usá-las; e
- Uma vez identificada a distribuição, como calcular:
  - probabilidade de eventos.
  - médias e variâncias das distribuições.