Distribuições de Probabilidade discretas

Prof. Igor Nascimento

31 de maio de 2021

1 Distribuições discretas

Para facilitar o cálculo de probabilidades, existem modelos de probabilidade que nos auxiliam a mensurar a incerteza relacionados aos experimentos aleatórios.

Com isso, vamos continuar utilizando os conceitos de variáveis aleatórias, comumente representadas por letras maiúsculas X,Y e Z. As letras minúsculas x,y e z serão utilizadas para indicar um valor possível para essa variável.

É importante lembrar que os valores que as variáveis aleatórias assumem são sempre números. Isto é, a variável aleatória transforma o espaço amostral de um experimento aleatório em um espaço amostral números, podendo ser discreto ou contínuo.

As variáveis aleatórios discretas estão relacionadas à contagem de eventos. Por exemplo:

- número de resultados cara no lançamento de 10 moedas honestas.
- número de acidentes em cruzamento da avenida Frei Serafim.
- número de gols em uma rodada do Brasileiro
- número de pedidos de uma pizzaria

Perceba que associado a todas esses números há incerteza do valor que a variável aleatória e, para isso, os modelos de probabilidade nos ajudam a mensurar tal incerteza.

1.1 Distribuição Bernoulli

A distribuição Bernoulli é adequada para indicar a ocorrência de um evento:

- chove amanhã?
- cara no lançamento da moeda?
- o bebê de uma gestante é menina?
- ganhar a aposta esportiva

Veja que todos esses experimentos aleatórios são do tipo sucesso, o evento de interesse, ou fracasso, não ocorrer o evento de interesse. Mas lembre-se, uma variável aleatória é sempre um número então, caso seja sucesso representaremos por 1, caso seja fracasso, por 0. É fundamental compreender **qual é o sucesso do experimento aleatório**. A probabilidade do sucesso é representado por p.

Dessa forma, dizemos que uma variável aleatória X tem distribuição Bernoulli com probabilidade de sucesso p e representamos por:

$$X \sim Ber(p) \tag{1.1}$$

A **função de probabilidade é** $P(X = 1) = p$ e $P(X = 0) = 1 - p$. Essa é a distribuição de probabilidade discreta mais simples.

Podemos calcular o valor médio de $E(X)$:

$$E(X) = \sum_{i=1}^{2} x_i \times P(X = x_i) = 1 \times p + 0 \times (1 - p) = p \tag{1.2}$$

Também podemos calcular a variância $V(X)$:

$$V(X) = \sum_{i=1}^{2} (x_i - E(X))^2 \times P(X = x_i) = (1 - p)^2 \times p + (0 - p)^2 \times (1 - p) \tag{1.3}$$

$$V(X) = (1 - 2p + p^2)^2 \times p + p^2 \times (1 - p) = p \tag{1.4}$$

$$V(X) = (p - 2p^2 + p^3) + p^2 - p^3 = p - p^2 = p(1 - p) \tag{1.5}$$

1.2 Distribuição Binomial

A distribuição binomial é a soma de 2 ou mais bernoullis independentes e idênticas (mesmo p). Por exemplo, o lançamento de uma moeda honesta com resultado sucesso cara é um variável do tipo Bernoulli, mas o lançamento de 10 moedas e a contagem do número de sucessos (caras) é uma distribuição Binomial. Uma variável aleatória X que conta esses sucessos pode assumir os valores $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$.

De maneira geral, representamso uma distribuição bionomial por:

$$X \sim Bin(n, p) \tag{1.6}$$

Além da probabilidade de sucesso p, precisamos saber o número de tentativas n. No nosso caso, como a moeda é honesta $p = 0.5$ e $n = 10$.

A distribuição de probabilidade é:

$$P(X = x_i) = C^n_{x_i} p^{x_i} (1 - p)^{n - x_i} \tag{1.7}$$

Para compreendermos o quanto essa distribuição é útil, vamos calcular probabilidade de que em 10 lançamentos tenhamos 4 resultados sucesso. Considere a possibilidade de termos $(1, 1, 1, 1, 0, 0, 0, 0, 0, 0)$. Como os eventos são independentes, então a probabilidade desse evento é:

$$P(1, 1, 1, 1, 0, 0, 0, 0, 0, 0) = p \times p \times p \times p \times (1 - p) \times (1 - p) \times (1 - p) \times (1 - p) \times (1 - p) \times (1 - p) \tag{1.8}$$

$$P(1,1,1,1,0,0,0,0,0,0) = p^4 \times (1-p)^6 \tag{1.9}$$

Outra possibilidade é $(0,1,1,1,0,0,0,0,0,1)$.

$$P(0,1,1,1,0,0,0,0,0,1) = p \times p \times p \times (1-p) \times (1-p) \times (1-p) \times (1-p) \times (1-p) \times (1-p) \times p \tag{1.10}$$

$$P(0,1,1,1,0,0,0,0,0,1) = p^4 \times (1-p)^6 \tag{1.11}$$

Veja que é a mesma probabilidade. Todas as possibilidades de resultado com 4 sucesso, a probabilidade será a mesma! Por tanto, para obtermos $P(X=4)$ temos que identificar todas as possibilidades e somar a probabilidade de cada uma com 4 sucessos dentre as 10. Qual o total de possibilidades de que em 10, tenhamos 4? Eu preciso escolher? Sim, apenas 4 das 10 vagas precisam ser sucesso. A ordem importa? Não, basta que a soma seja 4. Então é por combinação! O total de possibilidades é:

$$C^{10}_{x_i=4} = \frac{10!}{4!6!} = \frac{10 \times 9 \times 8 \times 7}{4 \times 3 \times 2 \times 1} = 210 \tag{1.12}$$

Portanto, se $X \sim Bin(n=10,p)$:

$$P(X=4) = 210 \times p^4(1-p)^6 \tag{1.13}$$

Esse resultado é encontrado diretamente pela fórmula da distribuição de probabilidade da Binomial:

$$P(X=4) = C^n_4 p^4(1-p)^{n-4} = C^{10}_4 \times p^4(1-p)^{10-4} = 210 \times p^4(1-p)^6 \tag{1.14}$$

Dessa forma, quando temos 1 ou mais eventos do tipo Bernoullis com probabilidade de sucesso p idêndicos e independentes, para calcular o número de sucessos em n tentativas, utilizamos o modelo Binomial!

Por exemplo, qual a probabilidade de um casal que terá 5 filhos, nenhum seja menino? Primeiro, vamos identificar o evento Bernoulli. Cada nascimento é um evento Bernoulli, pode ser um menino ou menina. O "sucesso"nesse caso, o evento de interesse, é menino. A probabilidade de sucesso é $p = \frac{1}{2}$. Então, se X é uma variável aleatória que conta o número de filhos meninos dentre 5 nascimentos, temos que $X \sim Bin(n=5, p=0.5)$. Então, a probabilidade $P(X=0)$:

$$P(X=0) = C^5_0(0.5)^0(1-0.5)^5 = 1 \times 1 \times (0.5)^5 = 3.125\% \tag{1.15}$$

Portanto, é muito pouco provável de que não tenham pelo menos 1 filho homem.

Vamos considerar agora uma situação em que a probabilidade de sucesso não é igual a do fracasso. Considere que a probabilidade de que uma pessoa adulta contaminada pelo COVID-19 necessite de atendimento na Unidade de Tratamento Intesivo (UTI) é de 5%. Em uma determinada cidade há apenas 2 leitos de UTI e sabe-se que há 30 adultos simultaneamente infectados. Qual a probabilidade de que essa quantidade de leitos não seja suficiente? Nesse caso, se dentre as 30 pessoas contaminadas, 3 ou mais precisem de leitos, não haverá leitos suficientes. Dessa forma, seja X a variável aleatória que conta o número de infectados que precisam de UTI. Dessa forma, gostaríamos de calcular a probabilidade $P(X>2) = P(X=3)+P(X=4)+P(X=5)+...P(X=30)$. Utilizando as propriedades da probabilidade (probabilidade complementar), temos que $P(X \leq 2) + P(X > 2) = 1$, então $P(X > 2) = 1 - P(X \leq 2)$. Ou seja, precisamos calcular apenas $P(X \leq 2) = P(X=0) + P(X=1) + P(X=2)$.

Nesse caso, o evento Beroulli é se cada pessoa precisa ou não de UTI. O sucesso é precisar de UTI. A probabilidade de sucessp $p = 5\%$ e são feitas $n = 30$ tentativas (número de pessoas infectadas). Assim, temos que $X \sim Bin(n = 30, p = 5\%)$.

$$P(X = 0) = C_0^{30}(0.05)^0(1 - 0.05)^{30} = 0.95^{30} = 0.2146388 \tag{1.16}$$

$$P(X = 1) = C_1^{30}(0.05)^1(1 - 0.05)^{29} = 30 \times 0.05 \times 0.95^{29} = 0.3389033 \tag{1.17}$$

$$P(X = 2) = C_2^{30}(0.05)^2(1 - 0.05)^{28} = 435 \times 0.05^2 \times 0.95^{28} = 0.2586367 \tag{1.18}$$

Portanto,

$$P(X > 2) = 1-(0.95^{30}+30\times0.05\times0.95^{29}+435\times0.05^2\times0.95^{28}) = 1-(0.2146388+0.3389033+0.2586367) = 0.1878212 = 18.78\% \tag{1.19}$$

A probabilidade de que os 2 leitos não sejam suficientes é de 18.78%. Ou seja, é preciso reforçar a medidas para tentar reduzir a trasmissão de vírus. Percebe a importância desses modelos de probabildade para nosso cotidiano e para a tomada de decisão?

Se não há recursos para aumentar o número de leitos, e ele quer diminuir a probabilidade de superlotação então é necessário conter o número de infectados. Se as medidas de distanciamento social pudessem reduzir os infectados para 20, então nesse caso $X \sim Bin(n = 20, p = 5\%)$.

$$P(X = 0) = C_0^{20}(0.05)^0(1 - 0.05)^{20} = 0.95^{20} = 0.3584859 \tag{1.20}$$

$$P(X = 1) = C_1^{20}(0.05)^1(1 - 0.05)^{19} = 20 \times 0.05 \times 0.95^{19} = 0.3773536 \tag{1.21}$$

$$P(X = 2) = C_2^{20}(0.05)^2(1 - 0.05)^{18} = 190 \times 0.05^2 \times 0.95^{18} = 0.1886768 \tag{1.22}$$

Portanto,

$$P(X > 2) = 1-(0.95^{20}+20\times0.05\times0.95^{19}+190\times0.05^2\times0.95^{18}) = 1-(0.3584859+0.3773536+0.1886768) = 0.07548367 = 7.54\% \tag{1.23}$$

Uma empresa que fabrica motores afirma que em um lote com 25 haverá no máximo 1 motor com problema. Sabendo que a probabilidade de um motor falhar é de 1%, qual a probabilidade de que essa empresa não cumpra com o prometido? O evento Bernoulli é o motor falhar. O sucesso é a falha do motor. A probabilidade do sucesso é $p = 1\%$. São feitas 25 tentativas (motores no lote). Assim, temos que $X \sim Bin(n = 25, p = 1\%)$, sendo X o número de falhas.

A empresa não terá problemas com $X = 0$ ou $X = 1$. Então a probabilidade desejada é $1 - (P(X = 0) + P(X = 1))$:

$$P(X = 0) = C_0^{25}(0.01)^0(1 - 0.01)^{25} = 0.99^{25} = 0.7778214 \tag{1.24}$$

$$P(X=1) = C_1^{25}(0.01)^1(1-0.01)^{24} = 25 \times 0.01 \times 0.99^{24} = 0.1964195 \tag{1.25}$$

Dessa forma, a probabilidade de que ela tenha a **afirmação desmentida pelo processo aleatório de fabricação** é de $1 - 0.9742409 = 0.0257911 = 2.57\%$.

A média da distribuição Binomial é:

$$E(X) = n \times p \tag{1.26}$$

A variância da distribuição Binomial é:

$$V(X) = n \times p \times (1-p) \tag{1.27}$$

1.3 Distribuição Poisson

A distribuição de Poisson é indicada para contagem de eventos de maneira geral, dependendo apenas do valor médio λ de ocorrência. Seja $X \sim Pois(\lambda)$, a distribuição de probabilidade é:

$$P(X = x_i) = \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} \tag{1.28}$$

A média e a variância são iguais para o modelo Poisson $E(X) = V(X) = \lambda$.

Considere que a média de acidentes em um cruzamento do avenida Frei Serafim é de 3.2 por semana. Considerando X o número de acidentes em uma semana específica, qual a probabilidade de que tenha mais do que 2 acidentes?

Nesse caso, temos $P(X > 2) = 1 - P(X = 0) + P(X = 1) + P(X = 2)$

$$P(X=0) = \frac{e^{-3.2}(3.2)^0}{0!} = e^{-3.2} = 0.0407 \tag{1.29}$$

$$P(X=1) = \frac{e^{-3.2}(3.2)^1}{1!} = e^{-3.2} \times 3.2 = 0.1304 \tag{1.30}$$

$$P(X=2) = \frac{e^{-3.2}(3.2)^2}{2!} = \frac{e^{-3.2} \times 3.2^2}{2} = 0.2087 \tag{1.31}$$

Portanto, a probabilidade é de (1 - (0.0407+0.1304+0.2087)) = 62.00%

Em uma determinada cidade, o número de novos casos de contaminados pelo COVID-19 por dia é de 1.2 pessoas. Qual a probabilidade de que haja mais do que 1 contaminado por dia?

$$P(X=0) = \frac{e^{-1.2}(1.2)^0}{0!} = e^{-1.2} = 0.3011942 \tag{1.32}$$

Sabendo disso, qual a probabilidade de que em 7 dias de uma semana, não haja nenhum caso de COVID-19? Veja, cada dia é uma Bernoulli. O sucesso é não haver nenhum novo caso. A probabilidade do sucesso é $p = 0.3011942$. São $n = 7$ réplicas. Seja Y o número de dias com zero casos de contaminados, temos que $Y \sim Bin(n = 7, p = 0.3011942)$:

$$P(Y=7) = C_7^7(0.3011942)^7(1-0.3011942)^0 = (0.3011942)^7 = 0.0002248673 \tag{1.33}$$

Outra forma de calcular essa probabilidade é utilizando a própria distribuição Poisson. Se em 1 dia a média de $\lambda = 1.2$, qual a média em uma semana? $\lambda_7 = 7 \times 1.2 = 8.4$.

$$P(X=0) = \frac{e^{-8.4}(8.4)^0}{0!} = e^{-8.4} = 0.0002248673 \tag{1.34}$$

1.4 Distribuição Hipergeométrica

A distribuição Hipergeométrica é adequada quando temos um experimento aleatório com retiradas sem reposição de eventos do tipo "sucesso"e "não sucesso". São exemplo disso, retirada de cartas de um baralho e sorteios de loteria.

Considere uma população com N objetos, sendo r com uma característica de interesse (sucesso) e $N-r$ tem a característica (fracasso). Dentre os elementos dessa população, são selecionados aleatoriamente n objetos. A probabilidade de que dentre esses n objetos selecionados, k sejam do tipo "sucesso" é calculada por MORETTIN; BUSSAB, 2017

$$P(X=k) = \frac{C_k^r C_{n-k}^{N-r}}{C_n^N}$$

Em uma população com $N = 100$ de animais há $r = 10$ machos. Escolhendo $n = 5$ animais dessa população sem reposição, a probabilidade de não se obter animais machos é:

$$P(X=0) = \frac{C_0^{10} C_{5-0}^{100-10}}{C_{10}^{10}} = \frac{\frac{10!}{10!0!}\frac{90!}{5!85!}}{\frac{100!}{5!95!}} = 0,584$$

Então, a probabilidade que tenham pelo menos 1 macho, é a probabilidade complementar:

$$P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5) = P(X>0) = 1 - P(X=0)$$

Imagine agora que nessa mesma população, as 5 retiradas serão feitas **com reposição**, qual seria distribuição? Binomial, pois a probabilidade de selecionar um macho é sempre igual a 10/100, isto é, são retiradas independetes, isto é, são Bernoulli's independentes.

$$P(X=0) = C_0^5 \left(\frac{10}{100}\right)^0 \left(1 - \frac{10}{100}\right)^5 = \left(1 - \frac{10}{100}\right)^5$$

Então a diferença básica entre hipergeométrica e a binomial é que no caso da binomial, as retiradas são independentes.

Referências

MORETTIN, P. A.; BUSSAB, W. O. *Estatística básica*. São Paulo: Editora Saraiva, 2017.