Distribuições discretas: Geométrica, Binomial Negativa e Hipergeométrica

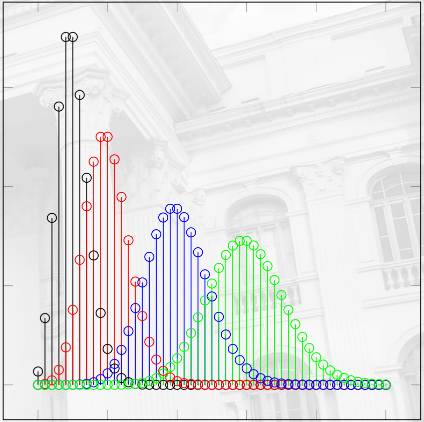
Prof. Walmes M. Zeviani

Departamento de Estatística
Universidade Federal do Paraná

Conteúdo

Neste vídeo

- Modelos probabilísticos discretos adicionais.
  - Geométrica.
  - Binomial Negativa.
  - Hipergeométrica.
- Fundamentação e propriedades.
- Exemplos de aplicação.



Distribuição Geométrica

Características de uma v.a. com distribuição Geométrica

Um experimento aleatório consiste em fazer **tentativas de Bernoulli**, de modo que

1. As tentativas sejam **independentes**.
2. Cada tentativa apresente apenas um de dois resultados possíveis (0: *fracasso* ou 1: *sucesso*).
3. A probabilidade de um sucesso em cada tentativa, $0 < p < 1$, é **constante**.
4. Seja Y a variável aleatória que conta o **número de tentativas** feitas até o **primeiro sucesso**.
5. Definida nestas condições, Y tem distribuição Geométrica com parâmetro p.

Figura 1. Cena de um jogo de bets. Extraído de gazetadopovo.com.br.

Distribuição Geométrica

A variável aleatória Y tem distribuição geométrica de parâmetro $0 < p < 1$ se sua função de probabilidades é dada por alguma das duas parametrizações a seguir.

1) Parametrização do **número de tentativas**

$$p(y) = (1-p)^{y-1} \cdot p,$$
$$y \in \{1, 2, \ldots\}.$$

Essa parametrização apresenta:

- $\mu = \text{E}(Y) = \frac{1}{p}$.
- $\sigma^2 = \text{V}(Y) = \frac{1-p}{p^2}$.

2) Parametrização do **número de fracassos**

$$p(y) = (1-p)^{y} \cdot p,$$
$$y \in \{0, 1, \ldots\}.$$

Essa parametrização apresenta:

- $\mu = \text{E}(Y) = \frac{1-p}{p}$.
- $\sigma^2 = \text{V}(Y) = \frac{1-p}{p^2}$.

Denotamos por $Y \sim \text{Geom}(p)$.

Gráfico da distribuição Geométrica

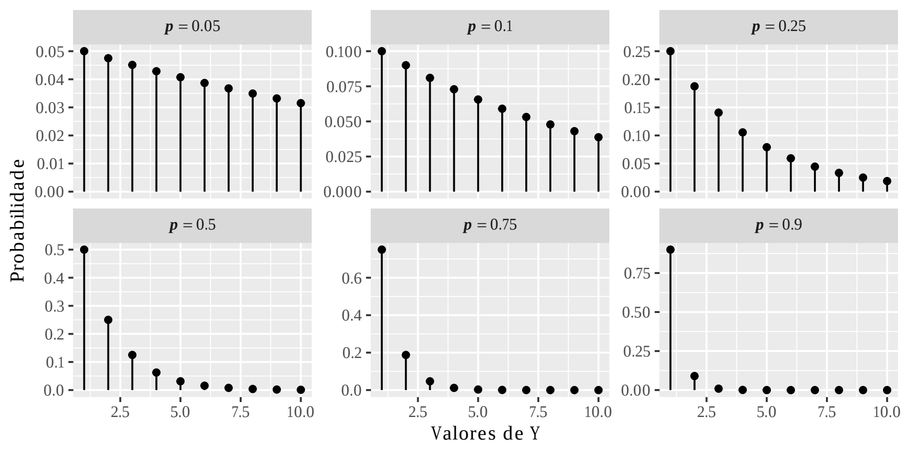


Figura 2. Gráficos para a distribuição Geométrica na parametrização de número de tentativas.

Propriedade da falta de memória

- Uma das propriedades da distribuição geométrica é a **falta de memória**, segundo a qual

$$\mathrm{P}(Y > y + a | Y > a) = \mathrm{P}(Y > y), \text{ para qualquer } a > 0.$$

- Assim, sob distribuição geométrica, a probabilidade de sucesso para a próxima tentativa após ter feito $a = 10$ tentativas é a mesma se já tivesse feito $a = 100$ tentativas ou após a primeira tentativa ($a = 1$).

Figura 3. Máquinas de aposta em um cassino. Foto de https://www.needpix.com/.

Exemplo: algoritmo para quebrar CAPTCHA

Uma estatística precisa consultar informações públicas disponibilizadas em um site governamental que é protegido com um sistema de CAPTCHA. Para isso, ela programou um algoritmo baseado em OCR (*optical character recognition*) que resolve corretamente as CAPTCHAS com $p = 0.5$.

1. Qual a probabilidade de quebrar a CAPTCHA na segunda tentativa?
2. Se o site tiver uma regra de bloquear o acesso após 7 ou mais tentativas erradas, para evitar ação de robô, qual a chance dela ser bloqueada?

Figura 4. Exemplo de CAPTCHA.

Solução

1. Usando a função de probabilidade, tem-se

$$p(2) = (1 - 0.5)^{2-1} \cdot 0.5 = 0.250.$$

2. Aplicando a fórmula, obtem-se pela regra do complementar

$$\begin{aligned} \mathrm{P}(Y \geq 7) = 1 - \mathrm{P}(Y < 7) &= 1 - \sum_{y=1}^{7-1} (1-p)^{y-1} \cdot p \\ &= 1 - (p(1) + \cdots + p(6)) \\ &= 0.016. \end{aligned}$$

Distribuição Binomial Negativa

Características de uma v.a. com distribuição Binomial Negativa

É uma **generalização** da distribuição **Geométrica**.

Um experimento aleatório consistem em fazer **tentativas de Bernoulli**, de modo que

1. As tentativas sejam **independentes**.
2. Cada tentativa apresente apenas um de dois resultados possíveis (0: *fracasso* ou 1: *sucesso*).
3. A probabilidade de um sucesso em cada tentativa, $0 < p < 1$, é **constante**.
4. Seja Y a variável aleatória que conta o **número de tentativas** feitas até r**-ésimo sucesso**.
5. Definida nestas condições, Y tem distribuição Binomial Negativa com parâmetros p e r.
6. Quando $r = 1$, obtém-se a distribuição Geométrica.

Figura 5. Dardos em um alvo. Foto de Hasan Albari no Pexels.

Distribuição Binomial Negativa

A variável aleatória Y tem distribuição Binomial Negativa com parâmetros $r > 0$ e $0 < p < 1$ se sua função de probabilidades é dada por alguma das duas parametrizações a seguir.

1) Parametrização do **número de tentativas**

$$p(y) = \binom{y-1}{r-1} \cdot (1-p)^{y-r} \cdot p^r,$$
$$y \in \{r, r+1, r+2, \ldots\}.$$

Essa parametrização apresenta:

- $\mu = \text{E}(Y) = \frac{r}{p}$.
- $\sigma^2 = \text{V}(V) = \frac{r \cdot (1-p)}{p^2}$.

2) Parametrização do **número de fracassos**

$$p(y) = \binom{r+y-1}{y} \cdot (1-p)^{y} \cdot p^r,$$
$$y \in \{0, 1, 2, \ldots\}.$$

Essa parametrização apresenta:

- $\mu = \text{E}(Y) = \frac{r(1-p)}{p}$.
- $\sigma^2 = \text{V}(Y) = \frac{r(1-p)}{p^2}$.

Denotamos por $Y \sim \text{BNeg}(r, p)$.

Gráfico da distribuição Binomial Negativa

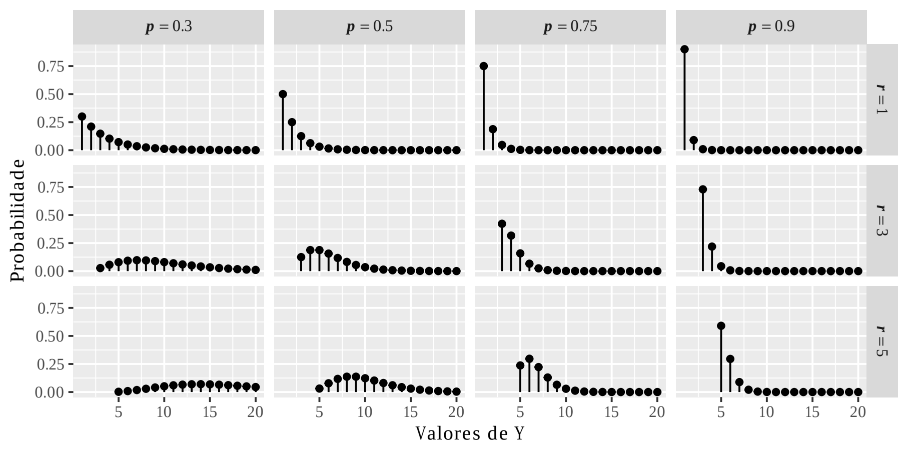


Figura 6. Gráficos para a distribuição Binomial Negativa com a parametrização do número de tentativas.

Exemplo: vitórias no vídeo game

Um jogador de vídeo game é confrontado com uma série de oponentes independentes em um jogo online. Dada sua experiência, ele tem 25% de chance de perder para um oponente. O jogador contínua a enfrentar oponentes até perder 3 vezes para então o jogo encerrar. O resultado com cada oponente é independente de confrontos prévios.

1. Qual a probabilidade de sair do jogo sem uma vitória?
2. Qual a probabilidade de enfrentar 5 oponentes para sair do jogo?
3. Qual o número esperado de oponentes em um jogo?

Figura 7. Jogo de futebol no vídeo game. Foto de EVG Culture no Pexels.

Solução

1. Fazendo uso da função de probabilidade, tem-se

$$p(3) = \binom{3-1}{3-1} \cdot (1-0.25)^{3-3} \cdot 0.25^3 = 0.25^3 = 0.016.$$

2. Novamente, aplicando a função de probabilidade

$$p(5) = \binom{5-1}{3-1} \cdot (1-0.25)^{5-3} \cdot 0.25^3 = 0.053.$$

3. O número médio de oponentes usa a expressão para o valor esperado

$$\mu = r/p = 3/0.25 = 12.$$

Relações com a Binomial Negativa

- A distribuição Binomial Negativa tem importante aplicação como alternativa à distribuição de Poisson na modelagem de **número de eventos por unidade** de tempo, espaço, etc.
- Observe a **dualidade** entre Binomial e Binomial Negativa. Na primeira, o número de tentativas é fixo e a v.a. é o número de sucessos. Na segunda, o número de sucessos é fixo e a v.a. é o número de tentativas.
- Como visto, a distribuição **Geométrica** é **caso particular** da distribuição Binomial Negativa quando $r = 1$.
- Pela propriedade da falta de memória da Geométrica, a distribuição Binomial Negativa pode ser obtida considerando a **soma** de r v.a. independentes de distribuição **Geométrica** de parâmetro p (soma das tentativas),

$$\begin{aligned} Y_{\text{BN}} &= Y_1 + \cdots + Y_r, \quad Y_i \sim \text{Geo}(p), \text{então} \\ Y_{\text{BN}} &\sim \text{BNeg}(r, p). \end{aligned}$$

Distribuição Hipergeométrica

Características de uma v.a. com distribuição Hipergeométrica

- Considere o experimento aleatório de retirar **sem reposição** $r > 0$ elementos de um conjunto de $m + n$ elementos em que $m > 0$ são de um tipo (sucesso) e $n > 0$ de outro (fracasso).
- Considere que todos os elementos têm **igual** probabilidade de serem retirados.
- Defina Y como o **número** de elementos de um tipo, e.g. sucesso, contidos na amostra retirada.
- Sob essas condições, Y tem distribuição **hipergeométrica** com parâmetros m, n e r.

Figura 8. Cartas de um baralho. Foto de Midhun Joy no Pexels.

Usos da Hipergeométrica

- Uma das principais aplicação da distribuição hipergeométrica é em situações envolvendo **amostragem** aleatória simples **sem reposição**.
- Na área de Controle Estatístico de Qualidade, aplica-se em problemas de amostragem de **aceitação de lotes**.
- Uma aplicação bastante interessante é na **estimação de tamanho de população** usando captura e recaptura, como visto no vídeo de motivação desta Unidade Didática.
  - Na pesca: tamanho de cardumes ou estoque pesqueiro.
  - Na segurança pública: determinar o número de pessoas participando de manifestações populares.

Figura 9. Pescador puxando a rede. Foto de Quang Nguyen Vinh no Pexels.

Distribuição Hipergeométrica

Uma variável aleatória Y tem distribuição hipergeométrica de parâmetros m, n e r se sua função de probabilidades é dada por

$$p(y) = \frac{\binom{m}{y} \cdot \binom{n}{r-y}}{\binom{m+n}{r}}, \quad y \in \{\max(0, r-n), \cdots, \min(r, m)\}.$$

Denotamos por $Y \sim \text{Hip}(m, n, r)$.

A v.a. Hipergeométrica tem média e variância dadas por

- $\mu = \text{E}(Y) = rp$ em que $p = \frac{m}{m+n}$.
- $\sigma^2 = \text{V}(Y) = rp(1-p)\left(\frac{(m+n)-r}{(m+n)-1}\right)$.

Gráfico da distribuição Hipergeométrica

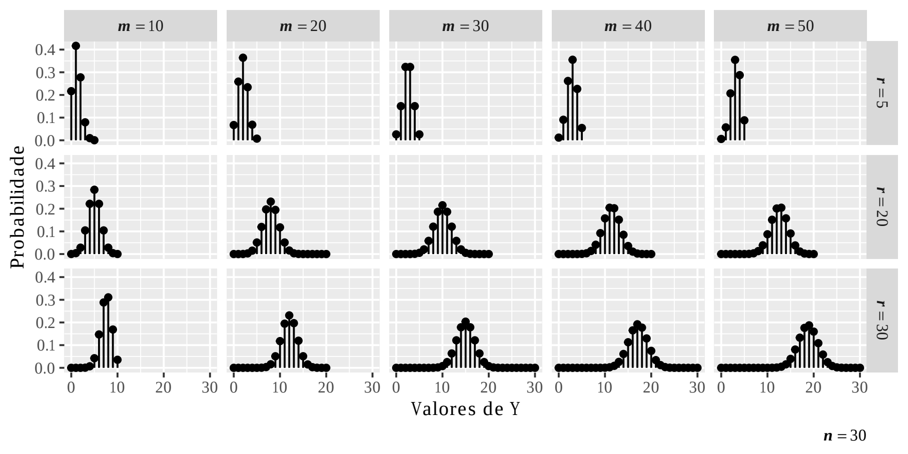


Figura 10. Gráficos para a distribuição Hipergeométrica.

Exemplo: questões no Moodle

No ambiente virtual de aprendizado Moodle, uma professora pode selecionar ao acaso sentenças para apresentar uma questão de verdadeiro e falso. Suponha que para um assunto, ela tenha 30 sentenças ao todo, das quais 10 são falsas.

1. Qual a probabilidade de um aluno receber na prova uma questão com 6 sentenças e todas serem falsas (e o aluno não ter o que marcar)?
2. Qual a probabilidade de ter 3 verdadeiras e 3 falsas?
3. Qual o número médio de questões falsas na prova?

1. Considere a seguinte amostra de dados:

33 37 41 43 46 47 55 59

Assinale as sentenças verdadeiras.

[x] A média é 45,12.
[ ] A mediana é 44.
[ ] A amplitude é 30.
[x] O desvio-padrão amostral é 8,69.
[x] A média geométrica é 44,40.
[ ] O desvio absoluto médio da mediana é 6.
[x] Se o maior valor for acrescido em 8 unidades, a média resultante irá aumentar 1 unidade.
[ ] Se os dados forem multiplicados por 2, o desvio-padrão será multiplicado por 4.

Figura 11. Uma questão de verdadeiro ou falso.

Solução

1. Faz-se o uso da função de probabilidade,

$$p(6) = \frac{\binom{10}{6} \cdot \binom{20}{6-6}}{\binom{10+20}{6}} = 3.5366932 \times 10^{-4}.$$

2. Idem ao anterior,

$$p(3) = \frac{\binom{10}{3} \cdot \binom{20}{6-3}}{\binom{10+20}{6}} = 0.23.$$

3. Usa-se a expressão da esperança matemática,

$$\mu = \frac{mr}{m+n} = \frac{106}{10+20} = 2.$$

Relação entre Binomial e Hipergeométrica

Sejam $Y_1 \sim \text{Bin}(r, p)$ e $Y_2 \sim \text{Hip}(m, n, r)$. Sabe-se pelas expressões que

$$\text{E}(Y_1) = \text{E}(Y_2) = rp, \quad \text{sendo } p = \frac{rm}{m+n} \text{ e } N = m + n.$$

$$\text{V}(Y_2) = \text{V}(Y_1)\left(\frac{N-r}{N-1}\right) = rp(1-p)\underbrace{\left(\frac{N-r}{N-1}\right)}_{(1)}$$

À medida que $N \gg r$, tem-se que $\left(\frac{N-r}{N-1}\right) \approx 1$ e com isso $\text{V}(Y_2) = \text{V}(Y_1)$. O termo (1) é conhecido como **fator de correção** para população finita.

O que isso significa na prática?

Que a distribuição Binomial pode ser usada como alternativa à distribuição Hipergemométrica, de forma a aproximá-la para o cálculo de probabilidades, quando o tamanho da população é muito grande ($N \to \infty$).

Considerações finais

Distribuições discretas adicionais

Modelos adicionais

- Existem muitos modelos probabilísticos discretos.
- Vários são **generalizações** dos modelos apresentados.
- Outros são construções considerando outras **premissas** ou **necessidades**.
- Modelos probabilísticos fazem **suposições** que podem ser **limitantes** ou **frágeis** em certos contextos.
- O **emprego adequado** de uma distribuição em uma situação prática é fundamental para a **utilidade** dos resultados obtidos.

Alguns modelos

- Lei de Benford.
- Beta-binomial.
- Poisson-binomial.
- Poisson generalizada.
- COM-Poisson.
- Gamma Count.
- Hipergeométrica não central de Fisher.
- Hipergeométrica não central de Wallenius.
- E outros.

Considerações finais

Neste vídeo

- Modelos probabilísticos discretos adicionais.
  - Geométrica.
  - Binomial Negativa.
  - Hipergeométrica.
- Fundamentação e propriedades.
- Exemplos de aplicação.