

Estatística

Professora: Patrícia Ferreira Paranaíba

Distribuições Discretas

Considere as seguintes situações por exemplo:

- ❶ (a) Lança-se uma moeda e observa-se o resultado obtido e (b) pergunta-se a um eleitor se ele vai votar no candidato A ou B;
- ❷ (a) Lança-se uma moeda n vezes e observa-se o número de caras obtidas e (b) de uma grande população, extrai-se uma amostra de n eleitores e pergunta-se a cada um deles em qual dos candidatos A ou B eles votarão e conta-se o número de votos do candidato A;
- ❸ (a) De uma urna com P bolas vermelhas e Q bolas brancas, extraem-se n bolas sem reposição e conta-se o número de bolas brancas e (b) de uma população com P pessoas a favor do candidato A e Q pessoas a favor do candidato B, extrai-se uma amostra de tamanho n sem reposição e conta-se o número de pessoas a favor do candidato A.

- O que veremos nesse capítulo são alguns modelos de variáveis aleatórias discretas que podem descrever situações como as listadas.
- Nesse contexto, um modelo será definido por uma variável aleatória e sua função de distribuição de probabilidade.
- De posse desses elementos, poderemos analisar diferentes situações práticas para tentar "encaixá-las" em algum dos modelos dados.

Distribuição Uniforme Discreta

- É a mais simples das distribuições discretas e recebe o nome de uniforme porque todos os valores da variável aleatória são assumidos com a mesma probabilidade.
- A função de probabilidade é dada por:

$$P(X) = \frac{1}{k}, \text{ para } x = x_1, x_2, x_3, \dots, x_k$$

k número de termos.

- Exemplo: O lançamento de um dado não viciado, definindo como X, a variável aleatória que representa a face voltada para cima, X assume os valores $x = 1, 2, 3, 4, 5, 6$ com a mesma probabilidade $1/6$.

Distribuição de Bernoulli

- Considere o lançamento de uma moeda. A característica desse experimento aleatório é que ele possui apenas dois resultados possíveis.
- Um experimento de Bernoulli é um experimento aleatório com apenas dois resultados possíveis; por convenção, um deles é chamado "sucesso" e o outro "fracasso".

- A distribuição de Bernoulli é a distribuição de uma v.a. X associada a um experimento de Bernoulli, onde se define $X = 1$ se ocorre sucesso e $X = 0$ se ocorre fracasso. Chamando de p a probabilidade de sucesso ($0 < p < 1$), a distribuição de Bernoulli é:

x	0	1
$P(X = x)$	$1 - p$	p

- Obviamente, as condições definidoras de uma f.d.p. são satisfeitas, uma vez que $p > 0$, $1 - p > 0$ e $p + (1 - p) = 1$. O valor de p é o único valor que precisamos conhecer para determinar completamente a distribuição; ele é, então, chamado parâmetro da distribuição de Bernoulli. Vamos denotar a distribuição de Bernoulli com parâmetro p por $Bern(p)$.

- A função de distribuição acumulada é dada por:

$$F_X(x) = \begin{cases} 0, & \text{se } x < 0; \\ 1 - p, & \text{se } 0 \leq x < 1; \\ 1, & \text{se } x \geq 1. \end{cases}$$

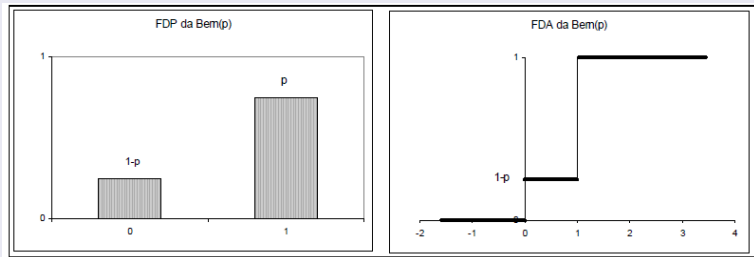


Figura: Distribuição de Bernoulli com parâmetro p .

- **Esperança:** Seja $X \sim \text{Bern}(p)$ (lê-se: a variável aleatória X tem distribuição de Bernoulli com parâmetro p).

$$X \sim \text{Bern}(p) \Rightarrow E(X) = p$$

- **Variância:** Seja $X \sim \text{Bern}(p)$.

$$X \sim \text{Bern}(p) \Rightarrow V(X) = p(1 - p)$$

Distribuição Binomial

- Consideremos n repetições independentes de um experimento de Bernoulli com parâmetro p (pense em n lançamentos de uma moeda com probabilidade p de cara). Vamos definir a seguinte v.a. associada a este experimento:

X = número de sucessos obtidos nas n repetições

- Os valores possíveis de X são 0 (só ocorrem fracassos), 1 (ocorre apenas 1 sucesso), 2 (ocorrem 2 sucessos), . . . , n (ocorrem apenas sucessos).

- Seja k o número de sucessos, logo,

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

A fórmula de cálculo de uma combinação é a seguinte:

$$C_k^n = \binom{n}{k} = \frac{n!}{k! (n - k)!}$$

Essa é a distribuição binomial ; note que para determiná-la precisamos conhecer os valores de n e p , que são os parâmetros da distribuição. Vamos usar a seguinte notação: $X \sim \text{bin}(n; p)$.

- **Esperança:**

$$X \sim \text{bin}(n, p) \Rightarrow E(X) = np$$

- **Variância:**

$$X \sim \text{bin}(n, p) \Rightarrow V(X) = np(1 - p)$$

Exemplo

Numa família com $n = 5$ filhos, qual a probabilidade de não haver homens? Qual a probabilidade de haver dois homens?

$$n = 5, p = \frac{1}{2}, q = \frac{1}{2}$$

$$P(X) = C_k^5 p^k q^{5-k}; \quad k = 0, 1, 2, 4, 5$$

A variável aleatória representa o número de homens (filhos do sexo masculino) encontrado em famílias de 5 filhos.

- $x = 0$ homem

$$\begin{aligned} P(X = 0) &= C_0^5 p^0 q^{5-0} \\ &= \frac{5!}{0! (5-0)!} \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 \\ &= \frac{1}{32} = 0,0313 \text{ ou } 3,13\% \end{aligned}$$

- $x = 2$ homens

$$\begin{aligned}P(X = 2) &= C_2^5 p^2 q^{5-2} \\&= \frac{5!}{2! (5-2)!} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 \\&= \frac{20}{2} \frac{1}{4} \frac{1}{8} = \frac{10}{32} = 0,3125 \text{ ou } 31,25\%\end{aligned}$$

Distribuição Hipergeométrica

- Considere uma população de tamanho N dividida em 2 classes, uma composta de r "sucessos" e a outra composta de $N - r$ "fracassos". Dessa população, vamos extrair uma amostra de tamanho n sem reposição.

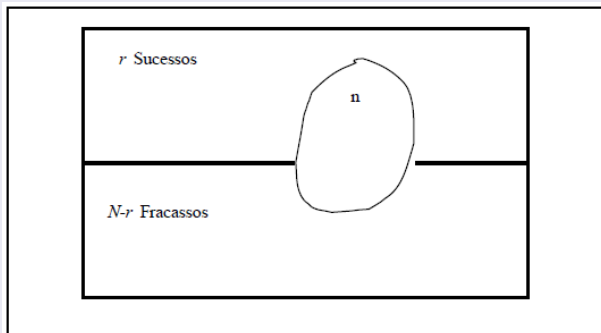


Figura: Ilustração do experimento definidor da v.a. hipergeométrica.

- Vamos considerar a seguinte v.a. para esse experimento:

X = número de sucessos em uma amostra retirada sem reposição

- A distribuição hipergeométrica tem a seguinte função de probabilidade:

$$P(X = k) = \frac{\binom{r}{k} \binom{N-r}{n-k}}{\binom{N}{n}}, \quad k = 0, 1, 2, \dots, n.$$

- **Esperança:**

$$X \sim \text{hiper}(N, r, n) \Rightarrow E(X) = n \frac{r}{N}$$

- **Variância:**

$$X \sim \text{hiper}(N, r, n) \Rightarrow V(X) = n \frac{r}{N} \frac{N-r}{N} \frac{N-n}{N-1}$$

Exemplo

Na Mega-Sena, um apostador escolhe 6 dezenas dentre 60. Qual a probabilidade dele acertar as 6 dezenas corretas?

- $N = 60, r = 6, N - r = 60 - 6 = 54, n = 6$

$$P(X = 6) = \frac{\binom{6}{6} \binom{60-6}{6-6}}{\binom{60}{6}} = \frac{1}{50063860} = 1,9974 \times 10^{-8}$$

Na Mega-Sena, um apostador escolhe 7 dezenas dentre 60. Qual a probabilidade dele acertar as 6 dezenas corretas?

- $N = 60, r = 6, N - r = 60 - 6 = 54, n = 7$

$$P(X = 6) = \frac{\binom{6}{6} \binom{60-6}{7-6}}{\binom{60}{7}} = \frac{54}{386206920} = 1,3982 \times 10^{-7}$$

- Dividindo a probabilidade de acertar 6 jogando 7, com a probabilidade de acertar 6 jogando 6, tem-se:

$$\frac{1,3982 \times 10^{-7}}{1,9974 \times 10^{-8}} \approx 7$$

- Isto significa que, jogando 7 dezenas, tem-se uma chance 7 vezes maior de acertar as 6 dezenas corretas. Com efeito, o preço pago por um cartão de 7 dezenas é 7 vezes maior que o preço de um cartão com 6 dezenas!

Distribuição de Poisson

- A distribuição de Poisson é empregada em experimentos nos quais se está interessado no número de sucessos ocorridos durante um intervalo contínuo, que pode ser um intervalo de tempo, espaço, comprimento, área, ou volume.
- A distribuição de Poisson tem a seguinte função de probabilidade:

$$P(X = K) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, 3, \dots$$

em que:

- e : base dos logaritmos neperianos (2,718...)
- λ : número médio de ocorrência do evento de interesse em um intervalo unitário.

- **Esperança:**

$$X \sim Poi(\lambda) \Rightarrow E(X) = \lambda$$

- **Variância:**

$$X \sim Poi(\lambda) \Rightarrow V(X) = \lambda$$

Exemplo

Uma central telefônica recebe uma média de 5 chamadas por minuto. Supondo que as chamadas que chegam constituam uma distribuição de Poisson, qual é a probabilidade da central não receber nenhuma chamada em um minuto?

- Seja X = número de chamadas por minuto. Então, $X \sim Poi(5)$.
Logo,

$$P(X = 0) = \frac{5^0}{0!} e^{-5} = 0,00673795$$

Distribuição Exponencial

- A distribuição de Poisson conta o número de eventos discretos em um intervalo ou espaço, que está intimamente ligado à distribuição exponencial, que (entre outras aplicações) mede o intervalo ou espaço para a ocorrência de um evento.
- Esta distribuição tem sido bastante utilizada como um modelo para o tempo de vida de certos produtos e materiais.
- Uma variável aleatória contínua X tem distribuição exponencial com parâmetro λ , se sua função densidade de probabilidade é dada por:

$$f(x) = \lambda e^{-\lambda x}, x > 0$$

em que $\lambda > 0$ representa a taxa de ocorrência por unidade de medida.

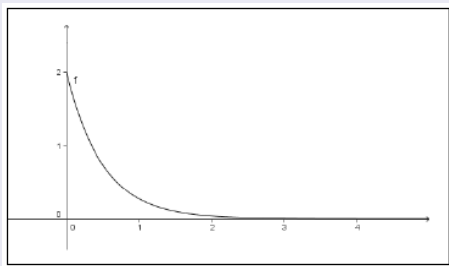


Figura: Representação gráfica da f.d.p. da Exponencial ($\lambda = 2$).

- Assim, a função de distribuição é dada por:

$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

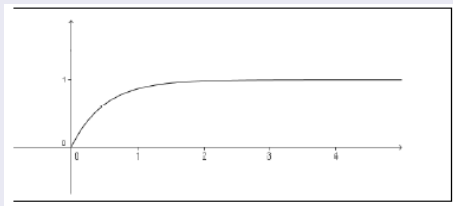


Figura: Representação gráfica da f.d.a. da Exponencial ($\lambda = 2$).

- **Esperança:**

$$X \sim \exp(\lambda) \Rightarrow E(X) = \frac{1}{\lambda}$$

- **Variância:**

$$X \sim \exp(\lambda) \Rightarrow V(X) = \frac{1}{\lambda^2}$$

Exemplo

O tempo até a falha do ventilador de motores a diesel tem uma distribuição Exponencial com parâmetro $\lambda = \frac{1}{28700}$ horas. Qual a probabilidade de um destes ventiladores falhar nas primeiras 24000 horas de funcionamento?

$$P(X \leq 24000) = F(24000) = 1 - e^{-\left(\frac{1}{28700}\right)24000} = 0,567$$

Ou seja, a probabilidade de um destes ventiladores falhar nas primeiras 24000 horas de funcionamento é de, aproximadamente, 56,7%.

Distribuição Normal

- A distribuição Normal corresponde a mais importante distribuição de variáveis aleatórias contínuas, em razão da sua enorme aplicação nos mais variados campos do conhecimento.
- Uma variável aleatória contínua X tem distribuição normal com parâmetros μ e σ^2 , se sua função de densidade de probabilidade é dada por

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$

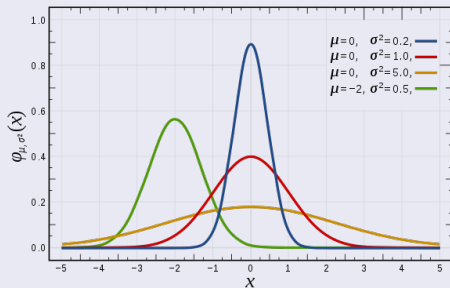


Figura: Representação gráfica da fdp da Normal $N(\mu, \sigma^2)$.

A distribuição normal apresenta as seguintes propriedades:

- É simétrica em relação a μ ;
- O ponto máximo de $f(x)$ ocorre em $x = \mu$;
- A área compreendida abaixo da curva normal e a acima do eixo x vale 1 ou 100%.

- **Esperança:**

$$X \sim N(\mu, \sigma^2) \Rightarrow E(X) = \mu$$

- **Variância:**

$$X \sim N(\mu, \sigma^2) \Rightarrow V(X) = \sigma^2$$

- A função de distribuição da Normal não possui uma forma analítica e é representada pela integral abaixo:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(u - \mu)^2}{2\sigma^2} \right\} du$$

Devido a dificuldade de resolução dessa integral, procurou-se métodos alternativos para obtenção das probabilidades.

- Se X tiver uma distribuição Normal $N(\mu, \sigma^2)$, e se $Z = \frac{X - \mu}{\sigma}$, então Z terá distribuição Normal $N(0, 1)$.
- A distribuição Normal com média $\mu = 0$ e variância $\sigma^2 = 1$ é conhecida como distribuição Normal reduzida ou padronizada. Uma variável aleatória com essa distribuição geralmente é simbolizada pela letra Z .
- Uma das formas mais utilizadas para determinar probabilidades de uma distribuição Normal é por meio de tabela de probabilidades de uma distribuição Normal padrão (Z).

Exemplos

Exemplo 1: Seja uma variedade de milho em que a altura é um variável X com distribuição normal com média $\mu = 200cm$ e variância $\sigma^2 = 100cm^2$. Qual a probabilidade de uma planta desta variedade ter altura entre 190 e 195cm?

Para obter $P(190 < X < 195)$, primeiro vamos padronizar esta variável, sendo $\sigma = \sqrt{\sigma^2} = \sqrt{100} = 10$

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{190 - 200}{10} = -1$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{195 - 200}{10} = -0,5$$

Então,

$$\begin{aligned}P(190 < X < 195) &= P(-1 < Z < -0,5) = \Phi(-0,5) - \Phi(-1) \\&= \Phi(1) - \Phi(0,5) = 0,84134 - 0,6915 \\&= 0,14988\end{aligned}$$

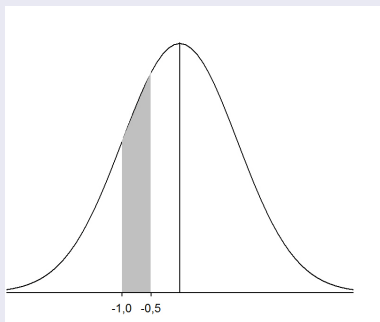


Tabela 6.1: Tabela da Distribuição Acumulada da Normal Padrão $\Phi(z) = P(Z \leq z)$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,50000	0,50399	0,50798	0,51197	0,51595	0,51994	0,52392	0,52790	0,53188	0,53586
0,1	0,53983	0,54380	0,54776	0,55172	0,55567	0,55962	0,56356	0,56749	0,57142	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,59871	0,60257	0,60642	0,61026	0,61409
0,3	0,61791	0,62172	0,62552	0,62930	0,63307	0,63683	0,64058	0,64431	0,64803	0,65173
0,4	0,65542	0,65910	0,66276	0,66640	0,67003	0,67364	0,67724	0,68082	0,68439	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,70540	0,70884	0,71226	0,71566	0,71904	0,72240
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,74215	0,74537	0,74857	0,75175	0,75490
0,7	0,75804	0,76115	0,76424	0,76730	0,77035	0,77337	0,77637	0,77935	0,78230	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,80234	0,80511	0,80785	0,81057	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,82894	0,83147	0,83398	0,83646	0,83891
1,0	0,84134	0,84375	0,84614	0,84850	0,85083	0,85314	0,85543	0,85769	0,85993	0,86214
1,1	0,86433	0,86650	0,86864	0,87076	0,87286	0,87493	0,87698	0,87900	0,88100	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,89435	0,89617	0,89796	0,89973	0,90147
1,3	0,90320	0,90490	0,90658	0,90824	0,90988	0,91149	0,91309	0,91466	0,91621	0,91774
1,4	0,91924	0,92073	0,92220	0,92364	0,92507	0,92647	0,92786	0,92922	0,93056	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,93943	0,94062	0,94179	0,94295	0,94408

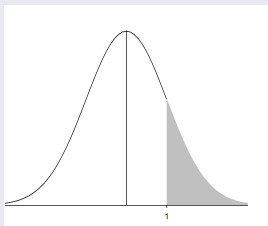
Exemplo 2: A produção diária de uma fabricante de tintas é uma variável aleatória X com distribuição normal com média $\mu = 10000$ galões e variância $\sigma^2 = 1000000$ galões². A direção dessa fábrica quer criar um bônus de incentivo aos funcionários, que será pago se a produção média diária exceder 11000 galões. Qual a probabilidade da empresa pagar o bônus?

Quero saber $P(X > 11000)$, primeiro vamos padronizar esta variável, sendo $\sigma = \sqrt{\sigma^2} = \sqrt{1000000} = 1000$:

$$z = \frac{x - \mu}{\sigma} = \frac{11000 - 10000}{1000} = 1,0$$

Assim,

$$P(X > 11000) = (Z > 1,0)$$



Como a tabela me fornece apenas o valor de que está entre 0 e z, então temos

$$\begin{aligned} P(X > 11000) &= P(Z > 1,0) = 0,5 - P(0 < Z < 1,0) \\ &= 0,5 - 0,3413 = 0,1587. \end{aligned}$$

Assim a probabilidade da empresa pagar o bonus é de 0,1587.

Tabela 6.2: Distribuição Normal - probabilidade do valor de z padronizado estar entre 0 e o valor tabulado nas margens

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,34134	0,34375	0,34614	0,34850	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41309	0,41466	0,41621	0,41774
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42786	0,42922	0,43056	0,43189