Intro à Probabilidade

Eduardo Horta

Aula 14

Variáveis Aleatórias Contínuas

Variáveis Aleatórias Contínuas

- Vamos iniciar com um **exemplo**.
- Consideremos um relógio analógico cujo ponteiro dos segundos "salta" de 6 em 6 graus. Esse relógio é alimentado por uma pilha AA que pode se esgotar a qualquer momento, e o instante exato em que o relógio para é, para todos os efeitos, aleatório.
- Seja X a variável aleatória "ângulo formado pelo ponteiro dos segundos no instante em que o relógio parar". Então X é uma v.a. discreta cujos possíveis valores são $0°, 6°, 12°, \ldots, 348°$ e $354°$. Vamos chamar esses possíveis valores de **sítios** ocupados pelo ponteiro.

Variáveis Aleatórias Contínuas

- Qual é a função massa de probabilidade de X?
- Se considerarmos que o momento em que a pilha se esgota "não privilegia" nenhum instante em particular, então a probabilidade de o ponteiro se encontrar em qualquer dos sítios deve ser a mesma: X é variável aleatória Uniforme discreta, isto é, sua função massa de probabilidade é dada por

$$p_X(x) = \begin{cases} \frac{1}{60} & \text{se } x = 0°, 6°, \ldots, 354° \\ 0 & \text{demais valores de } x. \end{cases}$$

Variáveis Aleatórias Contínuas

- Consideremos agora um relógio de alta precisão: por exemplo, um relógio que mede o tempo em nanosegundos (1 nanosegundo = 10^{-9} segundos).
- Nesse caso, o número total de sítios que o ponteiro pode ocupar é de $60 \cdot 10^9$.
- De fato, o ponteiro inicia no sítio $0°$, "salta" para o sítio $1° \cdot 6 \cdot 10^{-9}$, depois para o sítio $2° \cdot 6 \cdot 10^{-9}$ e assim sucessivamente.
- Isto é, os possíveis sítios ocupados pelo ponteiro são dados por

$$x_k = k \cdot \frac{360°}{60 \cdot 10^9}, \qquad k = 0, 1, 2, \ldots, \, 60 \cdot 10^9 - 1$$

Variáveis Aleatórias Contínuas

- Ainda assumindo que o momento em que a pilha se esgota “não privilegia” nenhum instante em particular, vemos que a probabilidade de encontrarmos, no momento da falha, tal ponteiro em um sítio fixado qualquer é de apenas $1/\big(60 \cdot 10^9\big)$.
- Em tal situação, visto que é extremamente improvável que encontremos o ponteiro em um sítio específico, talvez seja mais razoável nos indagarmos sobre a probabilidade de o encontrarmos em determinadas *regiões* do relógio.

Variáveis Aleatórias Contínuas

- Por exemplo, qual é a probabilidade de o ponteiro se encontrar no primeiro quadrante (quer dizer, entre $0°$ e $90°$), no momento da falha?
- Como o relógio tem quatro quadrantes, uma resposta razoável é que a probabilidade requerida deva ser igual a $1/4$. Vejamos como obter essa resposta de maneira rigorosa:
- Considere dois sítios x_i e x_j quaisquer, com $0° \leq x_i < x_j < 360°$. Então, vale a igualdade

$$\mathbb{P}\Big(x_i < X \leq x_j\Big) = \frac{x_j - x_i}{360} = \frac{j - i}{60 \cdot 10^9}. \tag{1}$$

Variáveis Aleatórias Contínuas

- Essa quantidade corresponde, precisamente, ao número de sítios compreendidos entre x_i (exclusive) e x_j (inclusive), dividido pela quantidade total de sítios.
- Assim, obtemos

$$\mathbb{P}\Big(0^\circ < X \leq 90^\circ\Big) = \frac{1}{4}.$$

conforme havíamos intuído.

Variáveis Aleatórias Contínuas

- É importante notar que $\mathbb{P}[0^\circ \leq X \leq 90^\circ] = 1/4 + 1/(60 \cdot 10^9)$, e consequentemente incluir ou não o sítio 0° na nossa definição de "primeiro quadrante" faz pouca diferença: as probabilidades obtidas diferem por somente $1/\left(60 \cdot 10^9\right)$.
- Também é importante perceber que a equação (1) vale independentemente da precisão que estamos considerando. Quer dizer, a relação ali expressa vale para qualquer relógio cujo ponteiro dos segundos "salta" entre sítios equiespaçados $x_0, x_1, \ldots, x_N$.

Variáveis Aleatórias Contínuas

- Consideremos agora um relógio cujo ponteiro dos segundos se move continuamente. Tal relógio pode ser pensado como um mecanismo ideal que mede o passar do tempo com *precisão arbitrária*.
- Uma primeira questão a ser colocada seria a seguinte: quais são os possíveis sítios ocupados pelo ponteiro?
- Uma resposta razoável é que qualquer ângulo maior ou igual a $0°$ e menor do que $360°$ é um sítio possível.
- Em outras palavras: o conjunto dos possíveis sítios deve ser o intervalo $[0°, 360°)$.

Detour

- Um resultado matemático profundo nos diz que *é impossível enumerar todos os elementos de um intervalo de números reais!* (já falamos sobre isto anteriormente)
- Em particular, se formarmos uma lista *arbitrária* de números reais pertencentes ao intervalo $[0°, 360°)$, digamos $\{a_1, a_2, a_3 \ldots\}$, então algum número b em $[0°, 360°)$ não estará nessa lista!
- Quer dizer, $b \neq a_1$, $b \neq a_2$, e assim por diante[1].
- Esse fato demonstra a inadequação do conceito de variável aleatória discreta para descrever fenômenos como aquele que propusemos aqui.

[1] Você pode se perguntar o que ocorreria se adicionássemos o número b à nossa lista original. Não adiantaria: existiria algum número c no intervalo $[0°, 360°)$ que estaria fora dessa nova lista. Esse procedimento nunca acabaria!

- De que maneira, então, podemos propor um modelo para descrever as probabilidades associadas à variável aleatória X quando o ponteiro dos relógios se move continuamente?
- O ponto de partida será a equação (1). Dados quaisquer números a e b, com $0° \leq a < b < 360°$, podemos *definir*

$$\mathbb{P}\Big(a < X \leq b\Big) = \frac{b-a}{360}. \tag{2}$$

- Note que, nesse caso, dado qualquer número x no intervalo $[0°, 360°)$, temos que $\mathbb{P}[X = x] = 0$.
- Isto é, a probabilidade de encontrarmos o ponteiro, no instante em que o relógio falha, em um sítio específico qualquer, é nula!

Variáveis Aleatórias Contínuas

- Isso pode soar estranho à primeira vista, mas lembre-se que nosso relógio mede o tempo com precisão *arbitrária*.
- Note também que, dessa forma, conseguimos incorporar a ideia inicial de que o momento em que a pilha se esgota "não privilegia" nenhum instante em particular.
- De fato, a probabilidade acima depende apenas do comprimento do arco que vai de a até b, mas não da posição particular desse arco sobre o círculo: se transladarmos ambos os números, a e b, pela mesma quantidade, a probabilidade na equação (2) permanece inalterada. Por exemplo,

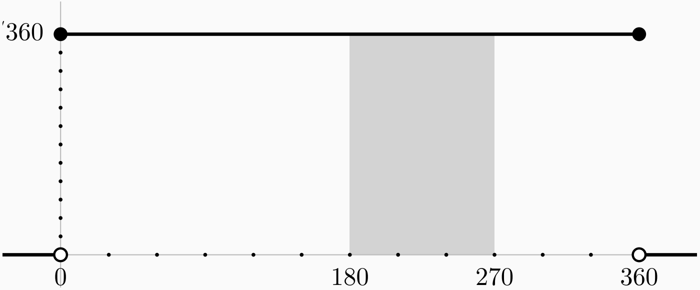
$$\mathbb{P}\Big(7^\circ < X \leq 97^\circ\Big) = \frac{97-7}{360} = \frac{90-0}{360} = \mathbb{P}\Big(0^\circ < X \leq 90^\circ\Big).$$

- Vamos agora *definir* uma função real de uma variável real, que denotaremos por f_X e chamaremos de **função densidade de probabilidade de X**.
- Essa função será utilizada para calcular probabilidades associadas à variável aleatória X. No presente exemplo, f_X é assim definida:

$$f_X(x) = \begin{cases} \frac{1}{360} & \text{se } 0 \leq x \leq 360 \\ 0 & \text{se } x < 0 \text{ ou } x > 360. \end{cases} \tag{3}$$

Variáveis Aleatórias Contínuas

- A Figura abaixo mostra o gráfico da função f_X definida acima. Em cinza vemos a região compreendida abaixo do gráfico de f_X e acima do eixo das abscissas, entre os números $a = 180$ e $b = 270$.



- Evidentemente, a área dessa região é igual a $(270 - 180) \times 1/360 = 1/4$, a qual corresponde precisamente a $\mathbb{P}(180 < X \leq 270)$.
- De fato, o mesmo procedimento pode ser usado para calcular $\mathbb{P}(a < X \leq b)$ para valores arbitrários de a e b (desde que $0 \leq a < b \leq 360$).
- *A propriedade de que* $\mathbb{P}(a < X \leq b)$ *corresponde a uma área compreendida abaixo do gráfico de* alguma *função* f_X *e acima do eixo das abcissas é precisamente o que caracteriza as variáveis aleatórias contíinuas*.

Variáveis Aleatórias Contínuas: definição

- Dizemos que uma variável aleatória X é (absolutamente) **contínua** se, *para alguma* função não negativa f_X, vale a seguinte igualdade:

$$\mathbb{P}\Big(a < X \leq b\Big) = \int_a^b f_X(x)\,\mathrm{d}x, \tag{4}$$

para quaisquer números reais $a < b$.

Variáveis Aleatórias Contínuas

- *Formalismos à parte*, podemos pensar uma variável aleatória contínua, digamos X, da seguinte maneira:

1. X é um característico numérico de um experimento aleatório;
2. Somente temos interesse em calcular probabilidades do tipo $\mathbb{P}(X \in A)$, onde A é ou um intervalo, ou uma união de intevalos, etc;
3. Podemos calcular essas probabilidades utilizando *alguma* função f_X, através da área $\int_a^b f_X(x)\,\mathrm{d}x$.

- A função f_X é chamada a **função densidade de probabilidade** de X. Variáveis aleatórias contínuas, portanto, são aquelas variáveis aleatórias para as quais utilizamos sua função densidade de probabilidade para calcular probabilidades de interesse, através da equação (4).

Variáveis Aleatórias Contínuas

- Os diferentes modelos probabilísticos para variáveis aleatórias contínuas se resumem, consequentemente, a diferentes escolhas de f_X. Se você "desenhar" uma função não negativa, tal que a área total compreendida entre seu gráfico e o eixo das abcissas seja igual a 1, então podemos dizer que essa função representa *algum* modelo probabilístico de uma v.a. contínua X.
- É importante ressaltar mais uma vez que, quando estudamos variáveis aleatórias contínuas, o tipo de probabilidade que estaremos interessados em calcular serão como as acima: a probabilidade de a variável em questão se situar em alguma *região* (intervalos, uniões de intervalos, etc).
- Essa mudança de enfoque é necessária porque variáveis aleatórias contínuas possuem massa zero em qualquer ponto x. De fato, se x é *qualquer* número real, então $\mathbb{P}[X = x]$ é igual à área de uma figura geométrica cuja base tem comprimento zero.

Variáveis Aleatórias Contínuas: FDA

- Uma relação importante segue imediatamente da definição de variável aleatória contínua: vemos que, nesse contexto, a função de distribuição acumulada é dada pela expressão

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,\mathrm{d}u, \qquad x \in \mathbb{R}.$$

- De fato, basta notar que vale $\mathbb{P}(X \leq x) = \lim_{a \to -\infty} \mathbb{P}(a < X \leq b)$.

Alguns Modelos Probabilísticos Contínuos

Alguns Modelos Probabilísticos Contínuos

- Quando estudamos variáveis aleatórias *discretas*, alguns modelos – devido à sua importância – recebem um nome especial e merecem ser estudados com maior detalhe (por exemplo, os modelos Bernoulli, Binomial, Poisson, etc).
- Em cada um desses casos, *determinar o modelo se resume a especificar a função massa de probabilidade* da variável aleatória em questão.
- Por exemplo, uma variável aleatória discreta X é dita **Binomial com parâmetros** $n \in \mathbb{N}$ **e** $\alpha \in [0, 1]$ se sua FMP é dada por

$$p_X(x) = \binom{n}{x} \alpha^x (1-\alpha)^{n-x}, \qquad x = 0, 1, \ldots n \tag{5}$$

  Utilizamos a notação $X \sim \mathrm{Binomial}(n, \alpha)$ para resumir esse fato (quer dizer, a expressão acima deve ser lida como “X é uma variável aleatória Binomial com parâmetros n e α”).

Alguns Modelos Probabilísticos Contínuos

- Agora note que, na verdade, a equação (5) especifica uma família de modelos: cada escolha para os parâmetros n e α determina uma diferente FMP.
- Dito de outra forma, se sabemos que X é uma variável aleatória Binomial com parâmetros n e α e quisermos calcular, digamos, $\mathbb{P}(X = 1)$, então para que obtenhamos o valor *numérico* correspondente a essa probabilidade é preciso atribuir um valor numérico específico para os parâmetros n e α.
- Por exemplo, se $X \sim \mathrm{Binomail}(4, {}^3\!/_4)$, então essa probabilidade é igual a 0.046875.

Alguns Modelos Probabilísticos Contínuos

- No caso de variáveis aleatórias *contínuas*, essa estrutura de apresentação irá se repetir. A diferença é que, agora, os diferentes modelos probabilísticos correspondem a diferentes expressões algébricas para a função *densidade* de probabilidade.
- Aqui, como como no caso discreto, *essas expressões vão depender de certos parâmetros*.
- Consequentemente, cada um dos modelos abaixo apresentados corresponde, na verdade, a uma família *paramétrica* de modelos.

Alguns Modelos Probabilísticos Contínuos

- Um último comentário sobre semelhanças e diferenças entre modelos discretos e contínuos: muitas vezes, modelos discretos podem ser deduzidos diretamente da estrutura do problema proposto.
- Por exemplo, se considerarmos a variável aleatória X = "número de caras em 10 lançamentos de uma moeda honesta", então com algum esforço e noções de análise combinatória obteremos a função massa de probabilidade correspondente ao modelo Binomial$(n = 10, \alpha = 1/2)$.

Alguns Modelos Probabilísticos Contínuos

- No caso de variáveis aleatórias contínuas, os modelos são construções teóricas (as densidades de probabilidade são meramente funções para as quais se tem uma fórmula dada): sua derivação, em geral, não decorre imediatamente da estrutura do problema proposto.
- Em muitos casos, a justificativa para se utilizar determinada função densidade de probabilidade é teórica (por exemplo, o Teorema Central do Limite nos dá indicativos de quando a distribuição normal poderia ser adequada).
- Em outros casos, a adequação desses modelos ao estudo de fenômenos concretos decorre de um conhecimento empírico.
- Por exemplo, sabe-se que a altura de indivíduos em uma população é bem descrita por um modelo Normal; sabe-se que o tempo até o decaimento do núcleo de um isótopo radiativo de Urânio é bem descrito por um modelo Exponencial, etc.

Distribuição Normal

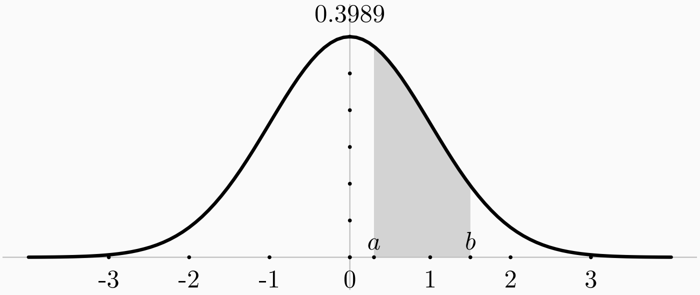
- Dizemos que uma variável aleatória contínua X tem distribuição **Normal com parâmetros** $\mu \in \mathbb{R}$ **e** $\sigma^2 > 0$ se sua função densidade de probabilidade é dada por

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(x-\mu)^2/2\sigma^2}, \qquad x \in \mathbb{R}. \tag{6}$$

- Notação: $X \sim N\left(\mu, \sigma^2\right)$
- Quando $\mu = 0$ e $\sigma^2 = 1$, dizemos que X é **Normal padrão**.

Distribuição Normal

- A Figura abaixo mostra o gráfico da função densidade de probabilidade f_X de uma variável aleatória Normal padrão.



Distribuição Normal

- O gráfico também mostra, em cinza, a região compreendida abaixo do gráfico de f_X e acima do eixo das abscissas, entre os números $a = 0.3$ e $b = 1.5$.
- A área dessa região é igual a 0.3152814 (aproximadamente), isto é,

$$\int_{0.3}^{1.5} \frac{1}{\sqrt{2\pi}} \, \mathrm{e}^{-x^2/2} \, \mathrm{d}x \approx 0.3152814.$$

- Portanto, se X é uma v.a. Normal padrão, podemos usar a equação (4) para concluir que $\mathbb{P}(0.3 < X \leq 1.5) \approx 0.3152814$.
- Alterando-se os valores de a e b, podemos assim obter o valor numérico de $\mathbb{P}(a < X \leq b)$ em qualquer caso de nosso interesse.

Distribuição Exponencial

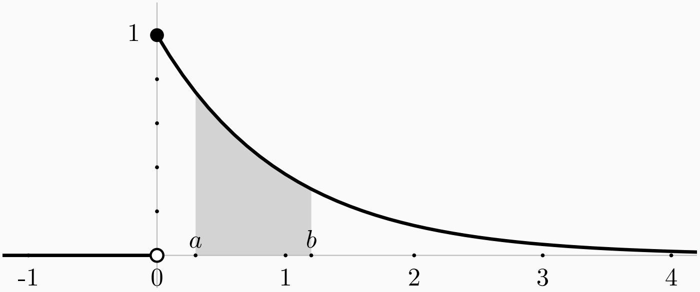
Dizemos que uma variável aleatória contínua X tem distribuição **Exponencial com parâmetro** $\lambda > 0$ se sua função densidade de probabilidade é dada por

$$f_X(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x} & \text{se } x \geq 0 \\ 0 & \text{se } x < 0. \end{cases} \tag{7}$$

- Notação: $X \sim \mathrm{Exponencial}(\lambda)$.
- **Importante**: alguns autores exprimem a fórmula da densidade exponencial em termos do parâmetro $\beta = 1/\lambda$.

Distribuição Exponencial

- A Figura abaixo mostra o gráfico da função densidade de probabilidade f_X de uma v.a. $X \sim \text{Exponencial}(1)$.



Distribuição Exponencial

- O gráfico também mostra, em cinza, a região compreendida abaixo do gráfico de f_X e acima do eixo das abscissas, entre os números $a = 0.3$ e $b = 1.2$.
- A área dessa região é igual a 0.439624. Portanto, $\mathbb{P}(0.3 < X \leq 1.2) = 0.439624$. Compare com o exemplo semelhante dado do modelo Normal padrão.

Valor esperado de variáveis aleatórias contínuas

- Se X é uma variável aleatória contínua com função densidade de probabilidade f_X, definimos o **valor esperado de** X como sendo o número real $\mathbb{E}X$ dado por

$$\mathbb{E}X = \int_{-\infty}^{+\infty} x\, f_X(x)\, \mathrm{d}x.$$

- Assim como no caso discreto, vale a *Law of the Unconscious Statistician*:

Teorema (Law of the Unconscious Statistician) Seja X uma variável aleatória contínua com função densidade de probabilidade f_X, e seja $h\colon \mathbb{R} \to \mathbb{R}$. Então

$$\mathbb{E}h(X) = \int_{-\infty}^{+\infty} h(x)\, f_X(x)\, \mathrm{d}x.$$

Aula 15

Variância de variáveis aleatórias contínuas

- Se X é uma variável aleatória contínua com função densidade de probabilidade f_X e valor esperado $\mathbb{E}X = \mu$, definimos a **variância de** X como sendo o número real $\mathrm{Var}(X)$ dado por

$$\mathrm{Var}(X) = \mathbb{E}\big\{(X - \mathbb{E}X)^2\big\} = \int_{-\infty}^{+\infty} (x-\mu)^2 \, f_X(x) \, \mathrm{d}x,$$

onde a última igualdade decorre da LotUS.

Lei dos Grandes Números e Teorema Central do Limite

Lei dos Grandes Números

- Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias **iid**[2]. Isto é, para quaisquer $x_1, \ldots x_n \in \mathbb{R}$ e quaisquer $1 \leq i, j \leq n$, tem-se

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{k=1}^{n} \mathbb{P}(X_k \leq x_i), \qquad \text{(indepenentes)}$$

$$X_i \stackrel{\mathrm{d}}{=} X_j, \qquad \text{(identicamente distribuídas)}$$

- Consideremos ainda a variável aleatória **média amostral**:

$$\bar{X}_n \stackrel{\text{def}}{=} \frac{X_1 + \cdots + X_n}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- O que podemos afirmar sobre a distribuição de $\bar{X}_n$?

[2] **iid** é o acrônimo para *independentes e identicamente distribuídas*.

Lei dos Grandes Números

- Por simplicidade, vamos nos ater ao caso em que X_1 é discreta.
- Sabemos que o valor esperado é linear: em particular, vemos que

$$\mathbb{E}(\bar{X}_n) = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}X_i.$$

- Como, por hipótese, as variáveis aleatórias $X_1, \ldots, X_n$ são iid, temos $\mu \stackrel{\text{def}}{=} \mathbb{E}X_1 = \mathbb{E}X_2 = \cdots \mathbb{E}X_n$. Fazendo a substituição, obtemos

$$E(\bar{X}_n) = \frac{\mu + \cdots + \mu}{n} = \frac{n}{n}\mu = \mu$$

- Isto é, a variável aleatória *média amostral* tem o mesmo valor esperado[3] que as variáveis aleatórias $X_1, \ldots, X_n$

[3] Por essa razão, no contexto de inferência estatística a média amostral $\bar{X}_n$ é dita um **estimador não-enviesado para** μ.

Lei dos Grandes Números

- Quanto à variância de $\bar{X}_n$, podemos utilizar o fato de que as variáveis aleatórias $X_1, \ldots, X_n$ são iid para concluir que

$$\mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i).$$

- Assim, escrevendo $\sigma^2 = \mathrm{Var}(X_1)$, obtemos

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2 + \cdots + \sigma^2}{n^2} = \frac{n}{n^2}\sigma^2 = \frac{\sigma^2}{n}.$$

- Segue que $\bar{X}_n$ distribui-se em torno do valor esperado μ, e que sua variância é “pequena” quando n é suficientemente “grande”.

Lei dos Grandes Números

- Esperamos, portanto, que ao computar a média amostral de n variáveis aleatórias iid (com n relativamente grande), obtenhamos um valor não muito distante de μ: a FMP de $\bar{X}_n$ deve atribuir a maior parte da sua massa a valores próximos de μ.
- Essa intuição pode ser refinada utilizando-se a **desigualdade de Chebyshev**:

$$\mathbb{P}\left(\left|\bar{X}_n - \mu\right| \geq x\right) \leq \frac{\sigma^2}{nx^2}, \qquad x > 0$$

- Na desigualdade acima, tipicamente tomamos um valor de x próximo de 0 e um valor de n suficientemente grande de tal forma que o lado direito da inequação seja tão pequeno quanto desejarmos.

Lei dos Grandes Números: Exemplo

- **Exemplo**: Considere n variáveis aleatórias iid $X_1, \ldots, X_n$ com $X_1 \sim \text{Bernoulli}(\alpha)$.
- Nesse caso, $\mathbb{E}X_1 = \alpha$ e $\text{Var}(X_1) = \alpha(1-\alpha)$.
- Já vimos em um exercício que a variável aleatória $S_n = X_1 + \cdots + X_n$ tem distribuição $\text{Binomial}(n, \alpha)$, com $\mathbb{E}S_n = n\alpha$ e $\text{Var}(S_n) = n\alpha(1-\alpha)$.
- Vamos interpretar S_n como sendo o "*número* de **caras** em n lançamentos independentes de de uma moeda (possivelmente desequilibrada)", onde a probabilidade de ocorrência de **cara** em cada lançamento é dada pelo *parâmetro de desequilíbrio* α.
- Sendo assim, $\bar{X}_n \equiv S_n/n$ é a "*proporção* de **caras** em n lançamentos independentes de uma moeda".

Lei dos Grandes Números: Exemplo

- Vemos que $\mathbb{E}(\bar{X}_n) = \alpha$ e $\mathrm{Var}(\bar{X}_n) = \alpha(1-\alpha)/n$.
- Agora, a função $f(\alpha) = \alpha(1-\alpha)$, com $0 \leq \alpha \leq 1$, é tal que $0 \leq f(\alpha) \leq {}^1\!/_4$ para qualquer α.
- Usando a desigualdade de Chebyshev, no presente contexto, nos dá

$$\mathbb{P}\left(\left|\bar{X}_n - \alpha\right| \geq x\right) \leq \frac{\alpha(1-\alpha)}{nx^2} \leq \frac{1}{4nx^2}$$

- Digamos agora que queremos inferir o valor de α utilizando a média amostral $\bar{X}_n$ como estimativa.

Lei dos Grandes Números: Exemplo

- Se desejarmos, com essa estimativa, cometer um erro de inferência com uma margem de erro de no máximo $x = 0.05$, a desigualdade acima nos dá

$$\mathbb{P}\left(\left|\bar{X}_n - \alpha\right| \geq 0.05\right) \leq \frac{400}{n}.$$

- Note que a desigualdade não tem utilidade se $n < 400$, pois nesse caso o lado direito é sempre maior do que 1.
- **Interpretação**: Digamos que iremos lançar a moeda $n = 4000$ vezes. Nesse caso, sabemos que a proporção de **caras** $\bar{X}_n$ que observarmos nesses n lançamentos estará no intervalo que vai de $\alpha - x$ até $\alpha + x$ com probabilidade maior ou igual a 90%.

Lei dos Grandes Números

- A Lei (Forte) dos Grandes Números refina a heurística acima garantindo que, de fato, para valores de n suficientemente grandes sempre observa-se $\bar{X}_n$ muito próximo de μ:

Teorema (Lei Forte dos Grandes Números): Seja $X_1, X_2, \ldots$ uma sequência de variáveis aleatórias iid, com $\mathbb{E}X_1 = \mu$. Então, com probabilidade 1,

$$\lim_{n \to \infty} \bar{X}_n = \mu.$$

- Em palavras: à medida que o *tamanho da amostra* aumenta, a *média amostral* se aproxima da *média populacional* (e isso ocorre com probabilidade 1).

Lei dos Grandes Números

- Outras maneiras de escrever a convergência dada na LFGN:

$$\mathbb{P}\left(\bar{X}_n \underset{n\to\infty}{\longrightarrow} \mu\right) = 1,$$

$$\frac{S_n}{n} \to \mu \text{ quase certamente ao } n \to \infty,$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu \text{ com probabilidade 1 ao } n \to \infty$$

etc.

Teorema Central do Limite

Teorema Central do Limite: preliminares

- Considere variáveis aleatórias iid $X_1, \ldots, X_n$ com $\mathbb{E}X_1 = \mu$ e $\mathrm{Var}(X_1) = \sigma^2$.
- O Teorema Central do Limite estuda a distribuição assintótica da **estatística padronizada**

$$Z_n \stackrel{\text{def}}{=} \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}.$$

- Note que $\mathbb{E}Z_n = 0$ e $\mathrm{Var}(Z_n) = 1$. (**Verifique!**)
- O TCL afirma - surpreendentemente - que, para valores de n suficientemente grandes, Z_n tem distribuição aproximadamente Normal padrão, *não importando qual a distribuição original* das variáveis aleatórias $X_1, X_2, \ldots, X_n$.

Teorema Central do Limite

Teorema Central do Limite (Berry-Esseen): Sejam $X_1, X_2, \ldots, X_n$ variáveis aleatórias iid, com $\mathbb{E}X_1 = \mu$, $0 < \mathrm{Var}(X_1) = \sigma^2 < \infty$ e $\mathbb{E}(|X_1|^3) = \gamma < \infty$. Então

$$\mathbb{P}(Z_n \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-u^2/2}\, \mathrm{d}u + R_n(x),$$

onde o "resto" $R_n(x)$ satisfaz a desigualdade

$$\left|R_n(x)\right| \leq \frac{\gamma}{2\sigma^3\sqrt{n}}$$

para todo $x \in \mathbb{R}$.

Teorema Central do Limite: comentários

- Uma maneira de reescrever o TCL: nas condições do Teorema acima, para valores de n suficientemente grande, vale que

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \overset{\text{aprox.}}{\sim} N(0,1)$$

- Note que a expressão acima pode ainda ser reescrita como

$$\bar{X}_n \overset{\text{aprox.}}{\sim} N(\mu, {}^{\sigma^2}/_{n})$$

- Essa última expressão complementa a heurística que seguimos na exposição da Lei dos Grandes Números: agora sabemos que não apenas a média amostral distribui-se em torno da média populacional, com variância inversamente proporcional ao tamanho da amostra, mas que, além disso, essa distribuição é aproximadamente normal.

Teorema Central do Limite: comentários

- Em alguns contextos, convém também interpretar o TCL em termos da distribuição da soma parcial $S_n = X_1 + \cdots + X_n$: note que as expressões anteriores são equivalentes a

$$S_n \overset{\text{aprox.}}{\sim} N(n\mu, n\sigma^2)$$

- No caso em que $\mu = 0$, obtemos uma interpretação particularmente interessante: a *Lei dos Grandes Números* afirma que, se dividirmos a soma parcial S_n pelo fator n então, à medida que n aumenta, a quantidade S_n/n aproxima-se de zero (embora S_n/n seja sempre uma variável aleatória, sua variância torna-se muito pequena).
- Já o *Teorema Central do Limite* estabelece qual o comportamento assintótico que obtemos se decidirmos "folgar" um pouco na padronização, dividindo S_n não por n mas pelo fator $\sqrt{n}$.
- Surpreendentemente, nesse caso a variável aleatória $S_n/\sqrt{n}$ tem distribuição aproximadamente Normal.

Teorema Central do Limite: Exemplo (continuação)

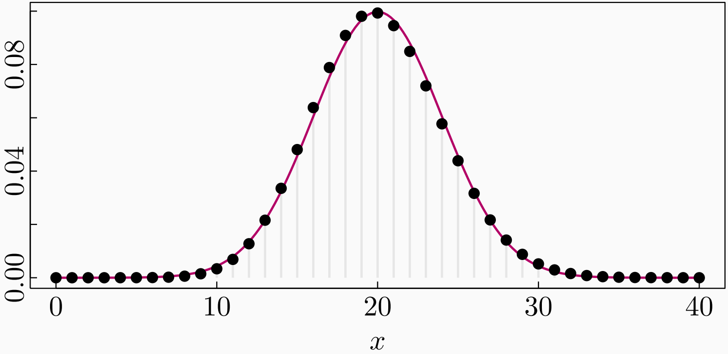
- Considere n variáveis aleatórias iid $X_1, \ldots, X_n$ com $X_1 \sim \mathrm{Bernoulli}(\alpha)$.
- Nesse caso, $\mathbb{E}X_1 = \alpha$ e $\mathrm{Var}(X_1) = \alpha(1-\alpha)$.
- Já vimos em um exercício que a variável aleatória $S_n = X_1 + \cdots + X_n$ tem distribuição $\mathrm{Binomial}(n, \alpha)$, com $\mathbb{E}S_n = n\alpha$ e $\mathrm{Var}(S_n) = n\alpha(1-\alpha)$.
- Pelo TCL, S_n tem distribuição aproximadamente Normal. Quer dizer, vale a aproximação

$$\sum_{u \leq x} \binom{n}{u} \alpha^u (1-\alpha)^{n-u} \approx \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(u-\mu)^2}{2\sigma^2} \right) \mathrm{d}u,$$

com $\mu = n\alpha$ e $\sigma^2 = n\alpha(1-\alpha)$

Teorema Central do Limite: Exemplo (continuação)

- A figura abaixo mostra a função massa de probabilidade correspondente à distribuição $\mathrm{Binomial}(n = 100,\ \alpha = {}^1\!/_5)$ (donde $\mathbb{E}S_n = 20$ e $\mathrm{Var}(S_n) = 16$) juntamente com a função densidade de probabilidade correspondente à distribuição $N(\mu = 20,\ \sigma^2 = 16)$.




I THOUGHT I WAS INTERESTED IN UNCERTAINTY BUT NOW I'M NOT SO SURE
TOSHIO