Introdução à Probabilidade e Estatística

Teorema Central do Limite

Carlos Trucíos ctruciosm.github.io

Universidade Federal do ABC (UFABC)

Semana 10

Introdução TCL IID TCL

Introdução

 Um dos resultados mais interessantes e utilizados na teoria de probabilidades

- Um dos resultados mais interessantes e utilizados na teoria de probabilidades
- Anteriormente vimos que se X_1, X_2, \ldots, X_n são *iid* com $X_i \sim N(\mu, \sigma)$, então

$$ar{X}_n \sim extstyle N \Big(\mu, rac{\sigma}{\sqrt{n}} \Big)$$
 ou equivalentemente $rac{\sqrt{n}(ar{X}_n - \mu)}{\sigma} \sim extstyle N(0,1)$

- Um dos resultados mais interessantes e utilizados na teoria de probabilidades
- Anteriormente vimos que se X_1, X_2, \ldots, X_n são *iid* com $X_i \sim N(\mu, \sigma)$, então

$$ar{X}_n \sim extstyle N \Big(\mu, rac{\sigma}{\sqrt{n}} \Big)$$
 ou equivalentemente $rac{\sqrt{n}(ar{X}_n - \mu)}{\sigma} \sim extstyle N(0,1)$

• Agora veremos que, se X_1, X_2, \ldots, X_n são *iid* com $E(X_1) = \mu$ e $V(X_1) = \sigma^2$, então

$$rac{\sqrt{n}(ar{X}_n - \mu)}{\sigma} \sim_{aprox} N(0,1)$$

- Um dos resultados mais interessantes e utilizados na teoria de probabilidades
- Anteriormente vimos que se X_1, X_2, \ldots, X_n são *iid* com $X_i \sim N(\mu, \sigma)$, então

$$ar{X}_n \sim extstyle N \Big(\mu, rac{\sigma}{\sqrt{n}} \Big)$$
 ou equivalentemente $rac{\sqrt{n}(ar{X}_n - \mu)}{\sigma} \sim extstyle N(0,1)$

• Agora veremos que, se X_1, X_2, \ldots, X_n são *iid* com $E(X_1) = \mu$ e $V(X_1) = \sigma^2$, então

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

• Com isso, podemos aproximar probabilidades

Introdução TCL IID TCL

TCL IID

TCL IID

Teorema

Sejam X_1, X_2, \dots, X_n (para n grande) v.as **independentes** e identicamente distribuidas com $E(X_1) = \mu$ e $V(X_i 1) = \sigma^2 < \infty$. Então,

$$rac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \sim_{\mathit{aprox}} \mathit{N}(0,1)$$

Isto implica que:

$$\bullet \frac{\sqrt{n}(ar{X}_n - \mu)}{\sigma} \sim_{aprox} N(0, 1)$$

$$\begin{array}{l} \bullet \ \, \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \sim_{aprox} N(0,1) \\ \bullet \ \, \lim_{n \to \infty} P\Big(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq a\Big) = \Phi(a) \end{array}$$

Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?

- Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?
- X_i : lado superior da moeda no i-ésimo lançamento (1: cara, 0:coroa)

- Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?
- X_i : lado superior da moeda no i-ésimo lançamento (1: cara, 0:coroa)
- $X_i \sim bernoulli(p = 0.5)$

- Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?
- X_i : lado superior da moeda no i-ésimo lançamento (1: cara, 0:coroa)
- $X_i \sim bernoulli(p = 0.5)$
- $E(X_1) = p = 0.5$

- Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?
- X_i : lado superior da moeda no i-ésimo lançamento (1: cara, 0:coroa)
- $X_i \sim bernoulli(p = 0.5)$
- $E(X_1) = p = 0.5$
- $V(X_1) = pq = 0.25$

- Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?
 - X_i : lado superior da moeda no i-ésimo lançamento (1: cara, 0:coroa)
- $X_i \sim bernoulli(p = 0.5)$
- $E(X_1) = p = 0.5$
- $V(X_1) = pq = 0.25$
- TCL: $\frac{X_1 + \cdots + X_n n\mu}{\sigma \sqrt{n}} \sim_{aprox} N(0,1)$

- Suponha que uma moeda honesta é lançada 900 vezes. Qual a probabilidade de obter mais de 495 caras?
 - X_i : lado superior da moeda no i-ésimo lançamento (1: cara, 0:coroa)
 - $X_i \sim bernoulli(p = 0.5)$
 - $E(X_1) = p = 0.5$
 - $V(X_1) = pq = 0.25$
- TCL: $\frac{X_1 + \cdots + X_n n\mu}{\sigma \sqrt{n}} \sim_{aprox} N(0, 1)$

•
$$P(\sum_{i=1}^{n} X_i > 495) = P(\underbrace{\frac{\sum_{i=1}^{n} X_i - 900 \times 0.5}{\sqrt{0.25} \times \sqrt{900}}}_{Z}) > \underbrace{\frac{495 - 900 \times 0.5}{\sqrt{0.25} \times \sqrt{900}}}_{3})$$

```
1-pbinom(495, 900, prob = 0.5)

## [1] 0.001200108

1-pnorm(3)

## [1] 0.001349898
```

② Suponha que selecionamos uma amostra aleatoria de tamanho n=12 de uma U[0,1]. Calcule $P(|\bar{X}_n-1/2|\leq 0.1)$

- ② Suponha que selecionamos uma amostra aleatoria de tamanho n=12 de uma U[0,1]. Calcule $P(|\bar{X}_n-1/2|\leq 0.1)$
- $X_i \sim U[0,1]$

- ② Suponha que selecionamos uma amostra aleatoria de tamanho n=12 de uma U[0,1]. Calcule $P(|\bar{X}_n-1/2|\leq 0.1)$
- $X_i \sim U[0,1]$
- $E(X_1) = 1/2$

- ② Suponha que selecionamos uma amostra aleatoria de tamanho n=12 de uma U[0,1]. Calcule $P(|\bar{X}_n-1/2|\leq 0.1)$
- $X_i \sim U[0,1]$
- $E(X_1) = 1/2$
- $V(X_1) = 1/12$

- ② Suponha que selecionamos uma amostra aleatoria de tamanho n=12 de uma U[0,1]. Calcule $P(|\bar{X}_n-1/2|\leq 0.1)$
- $X_i \sim U[0,1]$
- $E(X_1) = 1/2$
- $V(X_1) = 1/12$
- TCL: $\frac{\sqrt{n}(\bar{X}_n \mu)}{\sigma} \sim_{aprox} N(0, 1)$

- ② Suponha que selecionamos uma amostra aleatoria de tamanho n=12 de uma U[0,1]. Calcule $P(|\bar{X}_n-1/2|\leq 0.1)$
- $X_i \sim U[0,1]$
- $E(X_1) = 1/2$
- $V(X_1) = 1/12$
- TCL: $\frac{\sqrt{n}(\bar{X}_n \mu)}{\sigma} \sim_{aprox} N(0, 1)$
- $P(|\bar{X}_n 1/2| \le 0.1) = P(-0.1 \le \bar{X}_n 1/2 \le 0.1)$

$$P(\frac{-0.1\sqrt{12}}{1/\sqrt{12}} \le \frac{\sqrt{12}(\bar{X}_n - 1/2)}{1/\sqrt{12}} \le \frac{0.1\sqrt{12}}{1/\sqrt{12}})$$

$$P(\frac{-0.1\sqrt{12}}{1/\sqrt{12}} \le \frac{\sqrt{12}(\bar{X}_n - 1/2)}{1/\sqrt{12}} \le \frac{0.1\sqrt{12}}{1/\sqrt{12}})$$

$$P(\underbrace{-0.1 \times 12}_{-1.2} \le \underbrace{\frac{\sqrt{12}(\bar{X}_n - 1/2)}{1/\sqrt{12}}}_{Z} \le \underbrace{0.1 \times 12}_{1.2})$$

$$P(\frac{-0.1\sqrt{12}}{1/\sqrt{12}} \le \frac{\sqrt{12}(\bar{X}_n - 1/2)}{1/\sqrt{12}} \le \frac{0.1\sqrt{12}}{1/\sqrt{12}})$$

$$P(\underbrace{-0.1 \times 12}_{-1.2} \le \frac{\sqrt{12}(\bar{X}_n - 1/2)}{1/\sqrt{12}} \le \underbrace{0.1 \times 12}_{1.2})$$

$$pnorm(1.2) - pnorm(-1.2)$$

[1] 0.7698607

Introdução TCL IID TCL

TCL

• A versão apresentada do TCL é utilizada quando as variaveis são *iid* com $E(X) = \mu$ e $V(X) = \sigma^2 < \infty$.

- A versão apresentada do TCL é utilizada quando as variaveis são *iid* com $E(X) = \mu$ e $V(X) = \sigma^2 < \infty$.
- Se tivermos X_1, \ldots, X_n independentes mas não identicamente distribuidas o TCL apresentado não poder ser aplicado (mesmo que $E(X) = \mu$ e $V(X) = \sigma^2 < \infty$)

- A versão apresentada do TCL é utilizada quando as variaveis são *iid* com $E(X) = \mu$ e $V(X) = \sigma^2 < \infty$.
- Se tivermos X_1, \ldots, X_n independentes mas não identicamente distribuidas o TCL apresentado não poder ser aplicado (mesmo que $E(X) = \mu$ e $V(X) = \sigma^2 < \infty$)
- A. Liapounov em 1901 prova que ainda podemos ter uma versão do TCL sem precisar da suposição das v.as. de ser identicamente distribuidas

TCL (Liapounov)

Sejam X_1, X_2, \ldots, X_n v.as **independentes** com $E(X_i) = \mu_i$ e $V(X_i) = \sigma_i^2 < \infty$ com pelo menos um $\sigma_i^2 > 0$. Seja $S_n = X_1 + \ldots + X_n$ e $s_n^2 = \sigma_1^2 + \ldots + \sigma_n^2$, então

$$\frac{S_n - E(S_n)}{\sqrt{s_n^2}} \sim_{aprox} N(0,1)$$

Desde que, existe $\delta > 0$, t.q.

$$\lim_{n \to \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E|X_i - \mu_i|^{2+\delta} = 0$$

3 Sejam X_1, \ldots, X_n v.as independentes, $X_i \sim U[-i, i]$. Mostre que,

$$\frac{S_n - E(S_n)}{s_n} \sim_{aprox} N(0,1)$$

3 Sejam X_1, \ldots, X_n v.as independentes, $X_i \sim U[-i, i]$. Mostre que,

$$rac{S_n - E(S_n)}{s_n} \sim_{aprox} N(0,1)$$

Precisamos verificar as condições de Liapounov

•
$$E(X_i) = \frac{i-i}{2} = 0$$

3 Sejam X_1, \ldots, X_n v.as independentes, $X_i \sim U[-i, i]$. Mostre que,

$$rac{S_n - E(S_n)}{s_n} \sim_{aprox} N(0,1)$$

Precisamos verificar as condições de Liapounov

$$\bullet E(X_i) = \frac{i-i}{2} = 0$$

•
$$V(X_i) = \frac{(i+i)^2}{12} = \frac{i^2}{3} < \infty$$

Sejam X_1, \ldots, X_n v.as independentes, $X_i \sim U[-i, i]$. Mostre que,

$$\frac{S_n - E(S_n)}{S_n} \sim_{aprox} N(0,1)$$

Precisamos verificar as condições de Liapounov

$$\bullet E(X_i) = \frac{i-i}{2} = 0$$

•
$$V(X_i) = \frac{(i+i)^2}{12} = \frac{i^2}{3} < \infty$$

•
$$V(X_i) = \frac{(i+i)^2}{12} = \frac{i^2}{3} < \infty$$

• $s_n^2 = \sum_{k=1}^n \frac{k^2}{3} = \frac{n(n+1)(2n+1)}{18}$

Temos calculado s_n^2 mas de fato não precisamos calcular, apenas saber a ordem grandeza

Limite

$$orall \ lpha > 0$$
, $\lim_{n o \infty} rac{1}{n^{lpha+1}} \sum_{k=1}^n k^{lpha} = rac{1}{lpha+1}$

• s_n^2 é da ordem n^3

Temos calculado s_n^2 mas de fato não precisamos calcular, apenas saber a ordem grandeza

Limite

$$\forall \ \alpha > 0$$
, $\lim_{n \to \infty} \frac{1}{n^{\alpha+1}} \sum_{k=1}^{n} k^{\alpha} = \frac{1}{\alpha+1}$

- s_n^2 é da ordem n^3
- Falta verificar que: $\lim_{n \to \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n E|X_i \mu_i|^{2+\delta} = 0$

Escolhemos $\delta = 1$

•
$$E|X_i|^3 = \int_{-i}^i \frac{|x|^3}{2i} dx = \frac{1}{i} \int_0^i x^3 dx = \frac{1}{i} \frac{x^4}{4} \Big|_0^i = \frac{i^3}{4}$$

Escolhemos $\delta=1$

•
$$E|X_i|^3 = \int_{-i}^i \frac{|x|^3}{2i} dx = \frac{1}{i} \int_0^i x^3 dx = \frac{1}{i} \frac{x^4}{4} \Big|_0^i = \frac{i^3}{4}$$

• $\sum_{i=1}^{n} E|X_i|^3$ é da ordem n^4

Escolhemos $\delta = 1$

•
$$E|X_i|^3 = \int_{-i}^i \frac{|x|^3}{2i} dx = \frac{1}{i} \int_0^i x^3 dx = \frac{1}{i} \frac{x^4}{4} \Big|_0^i = \frac{i^3}{4}$$

• $\sum_{i=1}^{n} E|X_i|^3$ é da ordem n^4

•
$$\lim_{n \to \infty} \frac{1}{\underbrace{s_n^{2+1}}_{\text{ordem } n^{9/2}}} \underbrace{\sum_{i=1}^n E|X_i|^{2+1}}_{\text{ordem } n^4} = 0$$

Escolhemos $\delta=1$

•
$$E|X_i|^3 = \int_{-i}^i \frac{|x|^3}{2i} dx = \frac{1}{i} \int_0^i x^3 dx = \frac{1}{i} \frac{x^4}{4} \Big|_0^i = \frac{i^3}{4}$$

• $\sum_{i=1}^{n} E|X_i|^3$ é da ordem n^4

$$\bullet \lim_{n\to\infty} \frac{1}{\underbrace{s_n^{2+1}}_{\text{ordem } n^{9/2}}} \underbrace{\sum_{i=1}^n E|X_i|^{2+1}}_{\text{ordem } n^4} = 0$$

Então,

$$\frac{S_n - E(S_n)}{S_n} \sim_{aprox} N(0,1)$$

• Utilizando o TCL de Liapounov podemos provar que:

• Utilizando o TCL de Liapounov podemos provar que:

TCL v.as. Bernoulli

Sejam X_1, \ldots, X_n v.as **independentes** t.q. $X_i \sim bernoulli(p_i)$. Então

$$\frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} p_i}{\sqrt{\sum_{i=1}^{n} p_i q_i}} \sim_{aprox} N(0,1)$$

Desde que
$$\lim_{n\to\infty}\sum_{i=1}^n p_iq_i=\infty$$

•
$$E(X_i) = p_i$$
, $V(X_i) = p_i q_i < \infty$

- $E(X_i) = p_i$, $V(X_i) = p_i q_i < \infty$
- Provaremos que TCL Bernoulli implica que as condições do TCL Liapounov são satisfeitas

- $E(X_i) = p_i, \ V(X_i) = p_i q_i < \infty$
- Provaremos que TCL Bernoulli implica que as condições do TCL Liapounov são satisfeitas
- $\bullet \ E(|X_i p_i|^3) = (1 p_i)^3 \underbrace{p_i}_{P(X_i = 1)} + p_i^3 \underbrace{q_i}_{P(X_i = 0)} = p_i q_i (q_i^2 + p_i^2) \le p_i q_i$

- $E(X_i) = p_i$, $V(X_i) = p_i q_i < \infty$
- Provaremos que TCL Bernoulli implica que as condições do TCL Liapounov são satisfeitas

•
$$E(|X_i - p_i|^3) = (1 - p_i)^3 \underbrace{p_i}_{P(X_i=1)} + p_i^3 \underbrace{q_i}_{P(X_i=0)} = p_i q_i (q_i^2 + p_i^2) \le p_i q_i$$

$$\frac{\sum_{i=1}^{n} E(|X_i - p_i|^3)}{\left(\sum_{i=1}^{n} p_i q_i\right)^{3/2}} \le \frac{\sum_{i=1}^{n} p_i q_i}{\left(\sum_{i=1}^{n} p_i q_i\right)^{3/2}}$$

- $E(X_i) = p_i$, $V(X_i) = p_i q_i < \infty$
- Provaremos que TCL Bernoulli implica que as condições do TCL Liapounov são satisfeitas

•
$$E(|X_i - p_i|^3) = (1 - p_i)^3 \underbrace{p_i}_{P(X_i = 1)} + p_i^3 \underbrace{q_i}_{P(X_i = 0)} = p_i q_i (q_i^2 + p_i^2) \le p_i q_i$$

$$\frac{\sum_{i=1}^{n} E(|X_i - p_i|^3)}{\left(\sum_{i=1}^{n} p_i q_i\right)^{3/2}} \le \frac{\sum_{i=1}^{n} p_i q_i}{\left(\sum_{i=1}^{n} p_i q_i\right)^{3/2}}$$

• Então,
$$\lim_{n \to \infty} \frac{1}{s_n^3} \sum_{i=1}^n E(|X_i - p_i|^3) = 0$$

- $E(X_i) = p_i, \ V(X_i) = p_i q_i < \infty$
- Provaremos que TCL Bernoulli implica que as condições do TCL Liapounov são satisfeitas

•
$$E(|X_i - p_i|^3) = (1 - p_i)^3 \underbrace{p_i}_{P(X_i = 1)} + p_i^3 \underbrace{q_i}_{P(X_i = 0)} = p_i q_i (q_i^2 + p_i^2) \le p_i q_i$$

$$\frac{\sum_{i=1}^{n} E(|X_i - p_i|^3)}{\left(\sum_{i=1}^{n} p_i q_i\right)^{3/2}} \le \frac{\sum_{i=1}^{n} p_i q_i}{\left(\sum_{i=1}^{n} p_i q_i\right)^{3/2}}$$

• Então,
$$\lim_{n \to \infty} \frac{1}{s_n^3} \sum_{i=1}^n E(|X_i - p_i|^3) = 0$$

• Logo,
$$\frac{\sum_{i=1}^{n} X_{i} - \sum_{i=1}^{n} p_{i}}{\sqrt{\sum_{i=1}^{n} p_{i} q_{i}}} \sim_{aprox} N(0,1)$$

Uma prova consiste de 99 questões. Suponha que a probabilidade de que um estudante responda corretamente a primeira pergunta é 0.99, a segunda, 0.98, (em geral para a i-ésima pergunta $p_i = 1 - i/100$). O estudante precisa responder corretamente pelo menos 60 questões para aprovar o examen. Assumindo independência entre as questões. Qual a probabilidade do estudante aprovar o exame?

- ① Uma prova consiste de 99 questões. Suponha que a probabilidade de que um estudante responda corretamente a primeira pergunta é 0.99, a segunda, 0.98, (em geral para a i-ésima pergunta $p_i = 1 i/100$). O estudante precisa responder corretamente pelo menos 60 questões para aprovar o examen. Assumindo independência entre as questões. Qual a probabilidade do estudante aprovar o exame?
- X_i : a i-ésima questão é respondida corretamente (Sim: 1, Não: 0)

- ① Uma prova consiste de 99 questões. Suponha que a probabilidade de que um estudante responda corretamente a primeira pergunta é 0.99, a segunda, 0.98, (em geral para a i-ésima pergunta $p_i = 1 i/100$). O estudante precisa responder corretamente pelo menos 60 questões para aprovar o examen. Assumindo independência entre as questões. Qual a probabilidade do estudante aprovar o exame?
- X_i : a i-ésima questão é respondida corretamente (Sim: 1, Não: 0)
- $X_i \sim bernoulli(p_i = 1 i/100)$

- ① Uma prova consiste de 99 questões. Suponha que a probabilidade de que um estudante responda corretamente a primeira pergunta é 0.99, a segunda, 0.98, (em geral para a i-ésima pergunta $p_i = 1 i/100$). O estudante precisa responder corretamente pelo menos 60 questões para aprovar o examen. Assumindo independência entre as questões. Qual a probabilidade do estudante aprovar o exame?
- X_i : a i-ésima questão é respondida corretamente (Sim: 1, Não: 0)
- $X_i \sim bernoulli(p_i = 1 i/100)$
- $E(X_i) = 1 i/100$, $V(X_i) = (1 i/100)(i/100) < \infty$

- Uma prova consiste de 99 questões. Suponha que a probabilidade de que um estudante responda corretamente a primeira pergunta é 0.99, a segunda, 0.98, (em geral para a i-ésima pergunta $p_i = 1 - i/100$). O estudante precisa responder corretamente pelo menos 60 questões para aprovar o examen. Assumindo independência entre as questões. Qual a probabilidade do estudante aprovar o exame?
- X_i : a i-ésima questão é respondida corretamente (Sim: 1, Não: 0)
- $X_i \sim bernoulli(p_i = 1 i/100)$

•
$$E(X_i) = 1 - i/100$$
, $V(X_i) = (1 - i/100)(i/100) < \infty$

•
$$E(X_i) = 1 - i/100$$
, $V(X_i) = (1 - i/100)(i/100) < \infty$
• $\sum_{i=1}^{99} p_i = 99 - \frac{1}{100} \sum_{i=1}^{99} = 99 - \frac{1}{100} \frac{99 \times 100}{2} = 49.5$

[1] 0.005055645

• Pelo TCL:
$$\frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} p_i}{\sqrt{\sum_{i=1}^{n} p_i q_i}} \sim_{aprox} N(0,1)$$

$$\sum_{i=1}^{99} p_i q_i = \underbrace{\frac{1}{100} \sum_{i=1}^{99} i}_{49.5} - \underbrace{\frac{1}{100^2} \sum_{i=1}^{99} i^2}_{\underbrace{99 \times 100 \times 199}_{200}} = 16.665$$

• Pelo TCL:
$$\frac{\sum_{i=1}^{n} X_{i} - \sum_{i=1}^{n} p_{i}}{\sqrt{\sum_{i=1}^{n} p_{i}q_{i}}} \sim_{aprox} N(0,1)$$

• Pelo TCL:
$$\frac{1}{\sqrt{\sum_{i=1}^{n} p_{i}q_{i}}} \sim_{aprox} N(0, 1)$$

• Queremos: $P(\sum_{i=1}^{99} X_{i} \ge 60) = P(\underbrace{\frac{\sum_{i=1}^{99} -49.5}{\sqrt{16.665}}}_{Z} X_{i} \ge \underbrace{\frac{60 - 49.5}{\sqrt{16.665}}}_{2.572})$

1-pnorm(2.572)

[1] 0.005055645

Leituras recomendadas

Leituras recomendadas

- Degroot & Schervish (6.3)
- Ross (8.3)

Para praticar

• Lista 9 do gradmat.ufabc.edu.br/disciplinas/ipe/listas/