DISTRIBUIÇÃO DA MÉDIA AMOSTRAL

Magalhães e Lima (7ª. edição) – Seção 7.1 e Seção 7.3

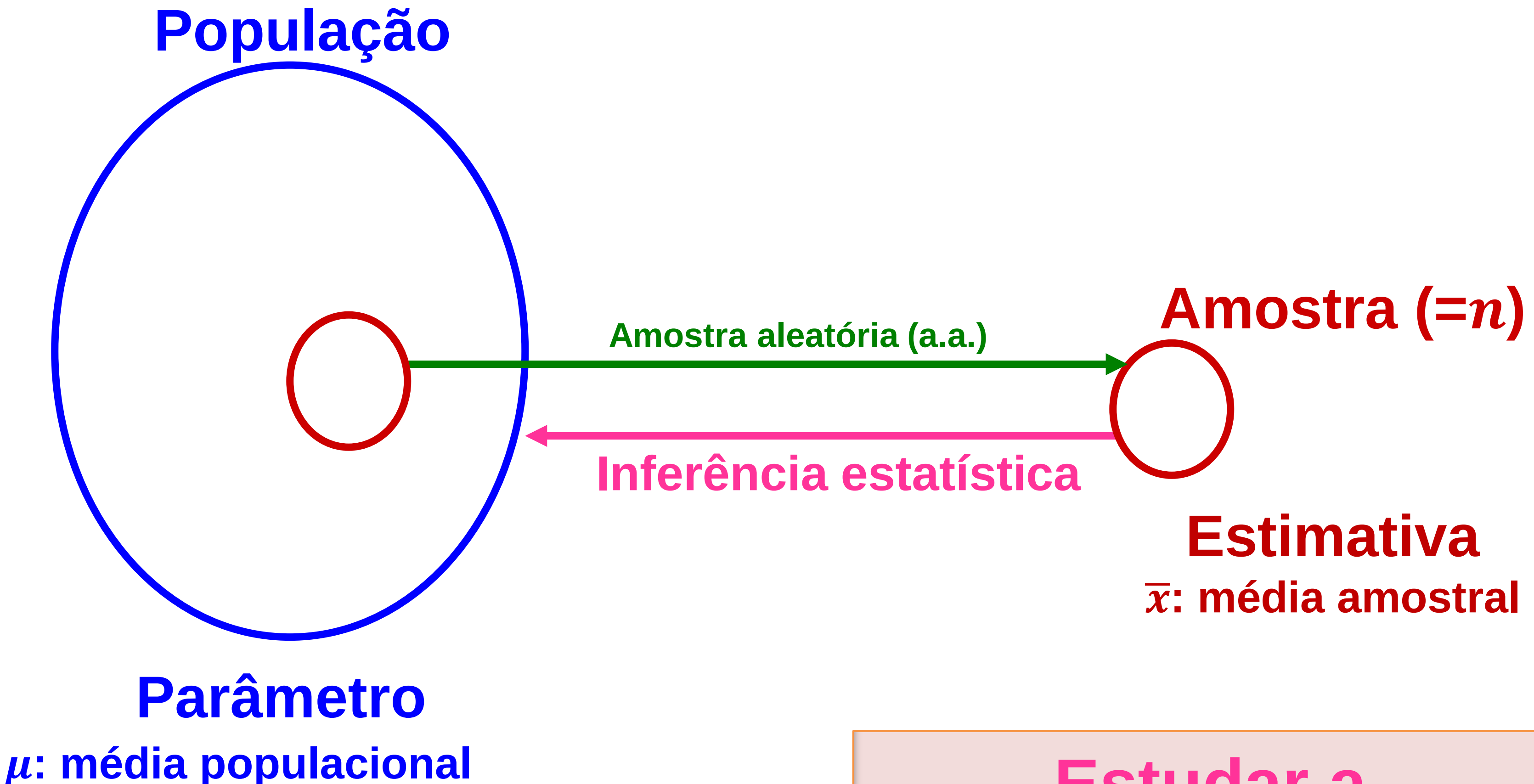
Objetivos de Aprendizagem

Os alunos devem ser capazes de:

- Compreender a média amostral como resultado aleatório necessário para discutir Inferência Estatística.
- Simular várias possíveis médias a partir de uma amostra de tamanho n e analisar o comportamento da distribuição da média amostral, ou seja, $\overline{\mathrm{X}} \sim ?$.

Acompanhe, previamente, o PLANO DE AULA no BLACKBOARD!

Introdução



X ~ Qualquer,
sendo σ^2: variância populacional

Estudar a distribuição da média amostral - $\bar{X}$!

Introdução

Dois conceitos importantes:

- **O que é Inferência Estatística:** utiliza dados obtidos de uma amostra para fazer estimativas (pontuais ou intervalares) ou testar hipóteses sobre características de interesse de uma população.

- **A média amostral $\bar{X}$ é uma variável aleatória**, já que amostras diferentes produzem, em geral, diferentes valores (estimativas) para a média amostral.

Introdução

Como a **média amostral é uma variável aleatória****, ela possui uma distribuição de probabilidades.**

Assim, estudaremos:

Distribuição de probabilidades da média amostral,

ou seja,

Distribuição de probabilidades de $\bar{X}$.

Distribuição da média amostral

A distribuição de probabilidades de $\bar{\mathbf{X}}$ é...

...baseada nas estimativas das médias amostrais provenientes de <u>todas as amostras possíveis</u> de uma população conhecida.

Insper

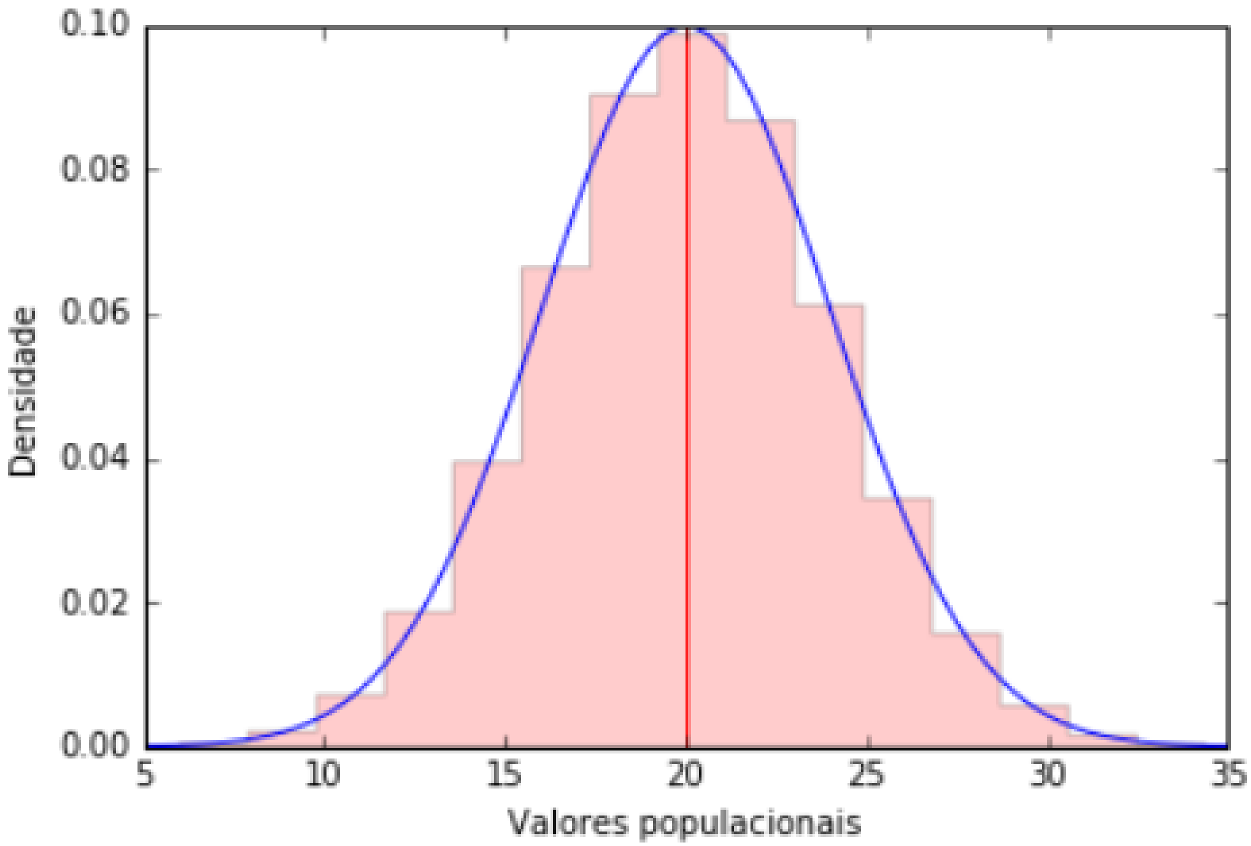
Dstribuição da média amostral quando variável de interesse é normal

Se $X \sim \text{Normal} \rightarrow \bar{X} \sim \text{Normal}$

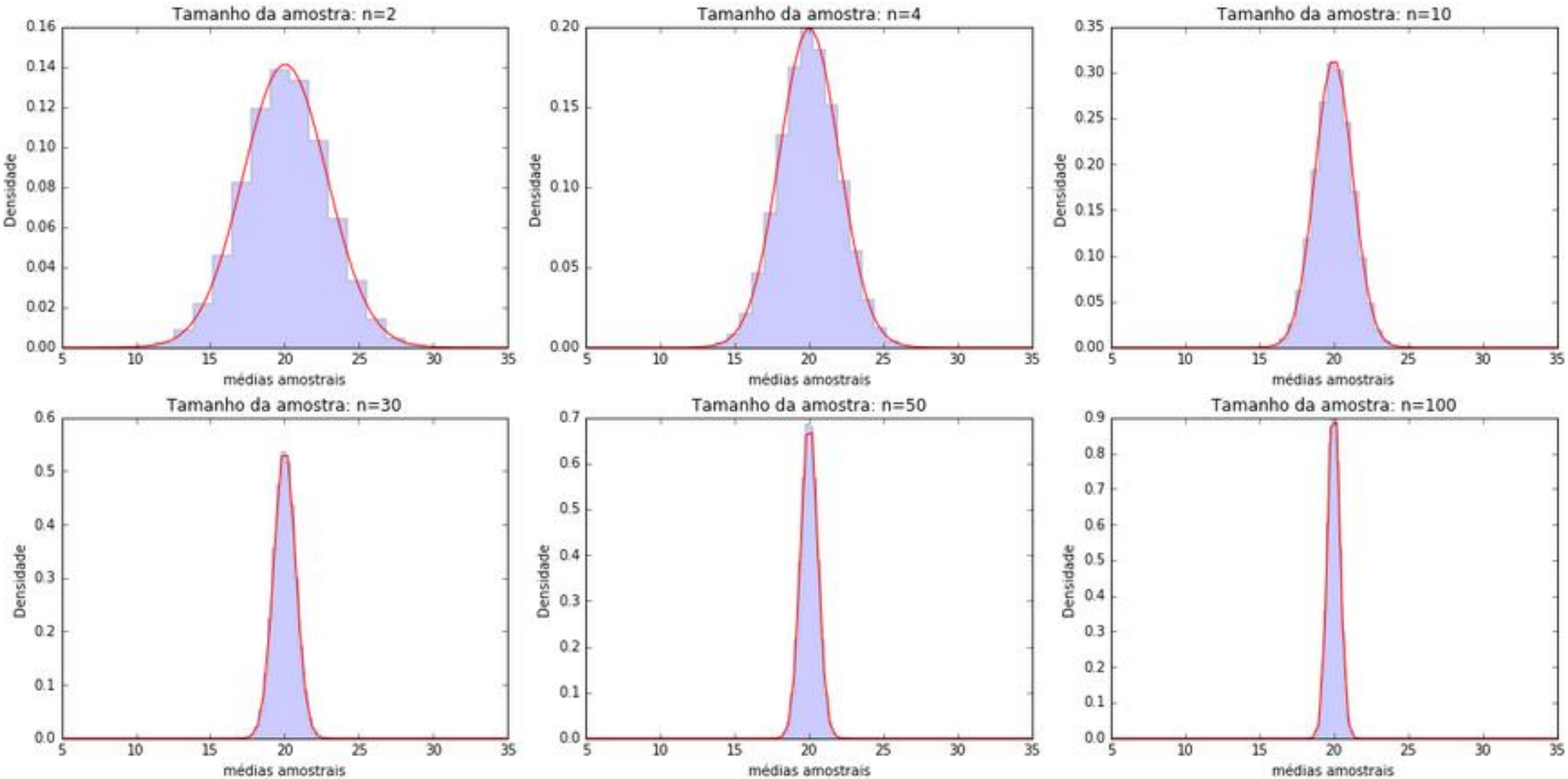
SIMULAÇÃO

Distribuição da variável de interesse - X

```
X segue uma distribuição Normal
Média de X ==> Valor de mu =  20.006
Variância de X ==> Valor de sigma2 = : 15.982
--------------------
```



Distribuição das médias amostrais - $\overline{X}$



Distribuição das médias amostrais - $\bar{X}$

```
Tamanho da amostra: n= 2
Média das médias amostrais: 20.028
Variância das médias amostrais:  7.973
---------------------
Tamanho da amostra: n= 4
Média das médias amostrais: 20.009
Variância das médias amostrais:  4.003
---------------------
Tamanho da amostra: n= 10
Média das médias amostrais: 20.005
Variância das médias amostrais:  1.599
---------------------
Tamanho da amostra: n= 30
Média das médias amostrais: 20.005
Variância das médias amostrais:  0.528
---------------------
Tamanho da amostra: n= 50
Média das médias amostrais: 20.006
Variância das médias amostrais:  0.317
---------------------
Tamanho da amostra: n= 100
Média das médias amostrais: 20.006
Variância das médias amostrais:  0.159
```

Caso particular de n=2

$X \sim \text{Normal} \Rightarrow$

$X_1 \sim \text{Normal} \Rightarrow$

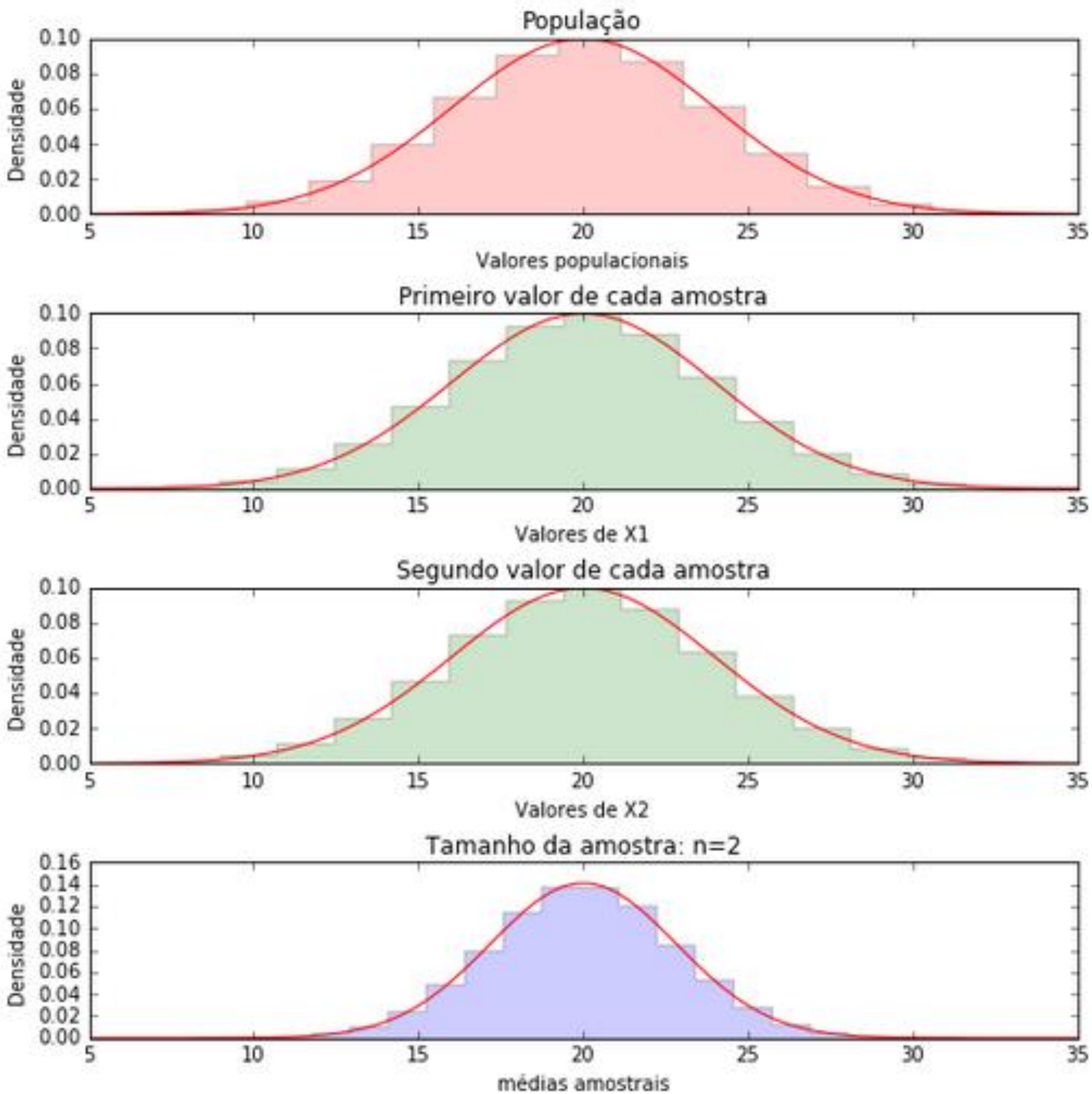
$\Downarrow$

$\text{Corr}(X_1, X_2) \cong 0$

$\Uparrow$

$X_2 \sim \text{Normal} \Rightarrow$

$\bar{X} \sim \text{Normal} \Rightarrow$



Resultados das Simulações:

Conclusões

Quando a variável de interesse **X é Normal**:

- a distribuição de $\bar{X}$ é **NORMAL** não importando o tamanho amostral;
- a média das médias amostrais sempre fica próxima à verdadeira média populacional;
- a variância das médias amostrais é, aproximadamente, a variância populacional dividida por n; e
- a sequência $X_1,...,X_n$ é uma amostra aleatória (independente e identicamente distribuída a X).

Insper

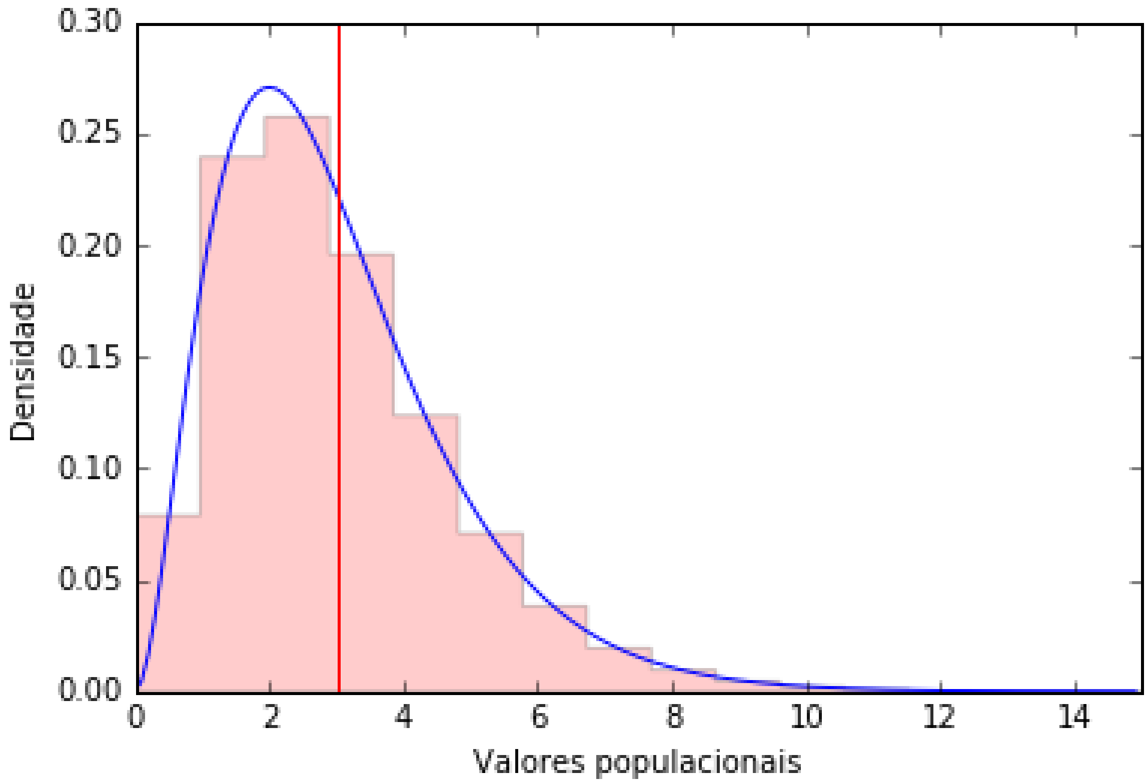
Dstribuição da média amostral quando variável de interesse não é normal

Se $\mathrm{X} \sim \mathrm{Qualquer} \rightarrow \bar{\mathrm{X}} \sim \mathrm{Normal}$

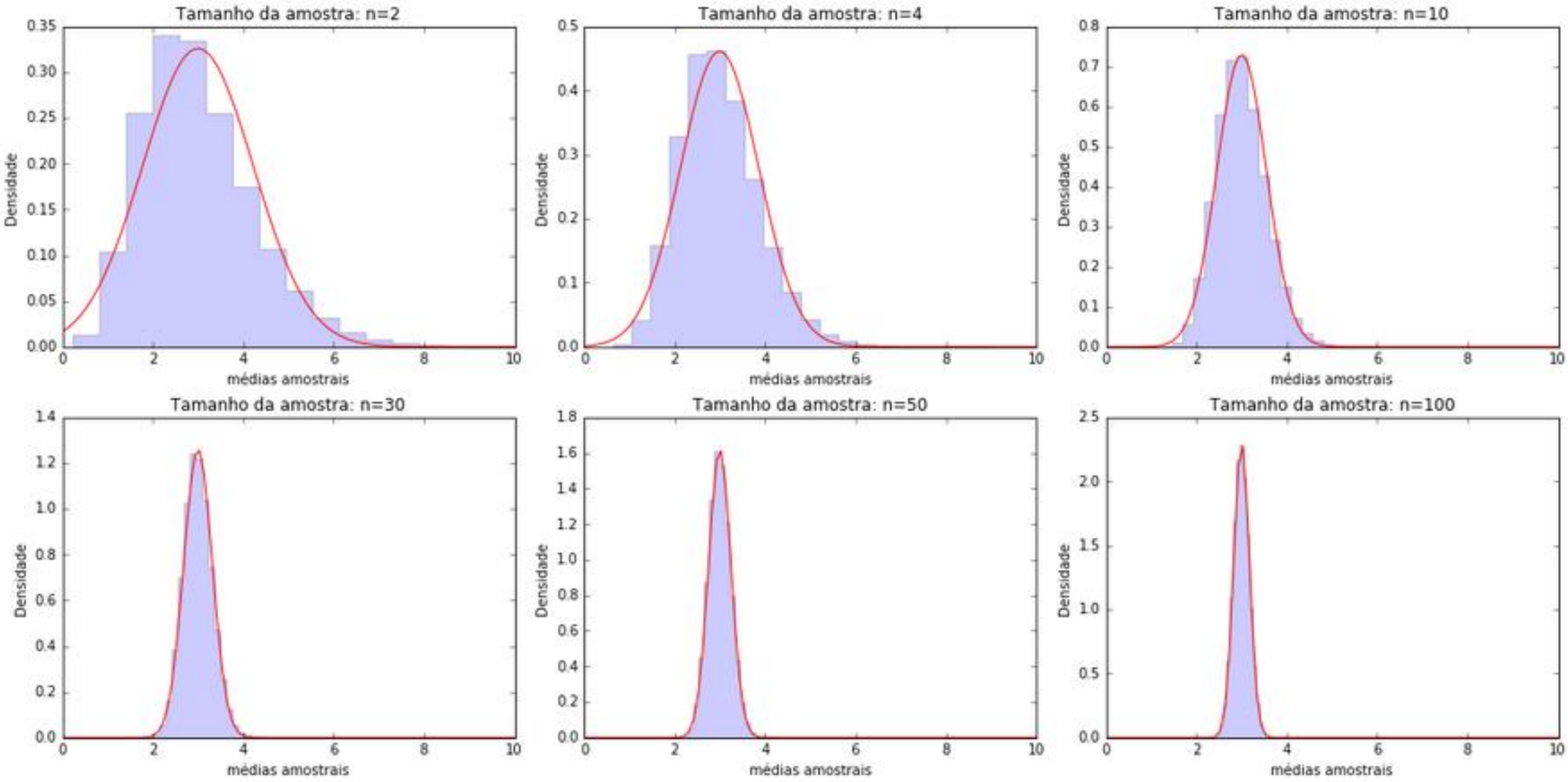
SIMULAÇÃO

Distribuição da variável de interesse - X

```
X segue uma distribuição Gamma
Média de X ==> Valor de mu =   3.000
Variância de X ==> Valor de sigma2 = :  2.998
--------------------
```



Distribuição das médias amostrais - $\bar{X}$



Distribuição das médias amostrais - $\bar{X}$

```
Tamanho da amostra: n= 2
Média das médias amostrais:  2.999
Variância das médias amostrais:  1.498
----------------------
Tamanho da amostra: n= 4
Média das médias amostrais:  2.995
Variância das médias amostrais:  0.748
----------------------
Tamanho da amostra: n= 10
Média das médias amostrais:  3.000
Variância das médias amostrais:  0.299
----------------------
Tamanho da amostra: n= 30
Média das médias amostrais:  3.000
Variância das médias amostrais:  0.100
----------------------
Tamanho da amostra: n= 50
Média das médias amostrais:  2.999
Variância das médias amostrais:  0.060
----------------------
Tamanho da amostra: n= 100
Média das médias amostrais:  2.999
Variância das médias amostrais:  0.030
----------------------
```

Caso particular de n=2

$X \sim \text{Gamma} \Rightarrow$

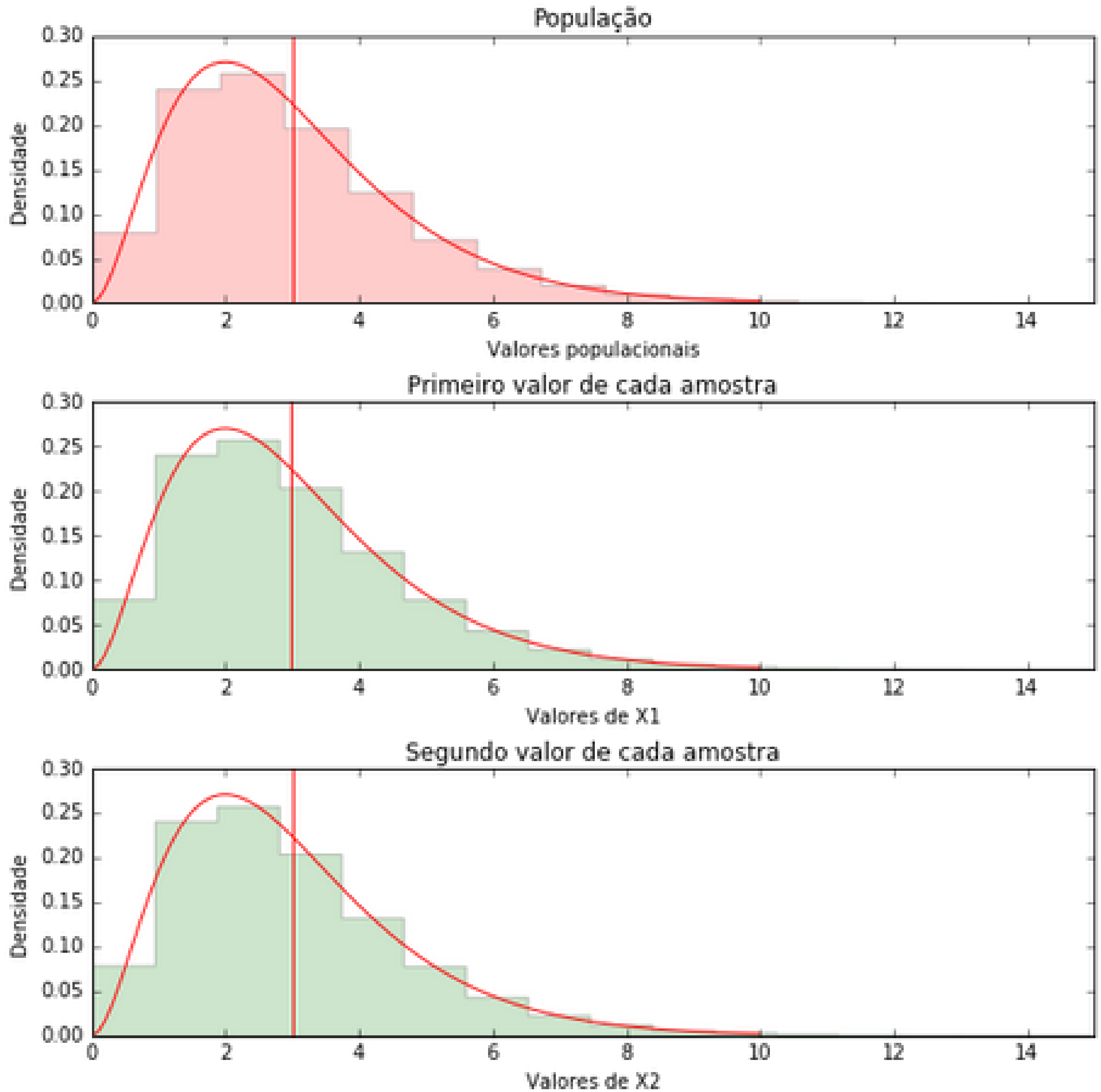
$X_1 \sim \text{Gamma} \Rightarrow$

$\Downarrow$

$\text{Corr}(X_1, X_2) \cong 0$

$\Uparrow$

$X_2 \sim \text{Gamma} \Rightarrow$



Resultados das Simulações:

Conclusões

Quando a variável de interesse **X não é Normal**:

- a distribuição de $\bar{X}$ é **NORMAL** para n suficientemente grande;
- a média das médias amostrais sempre fica próxima à verdadeira média populacional;
- a variância das médias amostrais é, aproximadamente, a variância populacional dividida por n; e
- a sequência $X_1,\ldots,X_n$ é uma amostra aleatória (independente e identicamente distribuída a X).

Insper

Dstribuição da média amostral

TEORIA

Distribuição da média amostral

Seja X_1, X_2, ... X_n uma AAS da v.a. X que apresenta média μ e variância σ^2.

Defina,

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

assim, é fácil provar que

$$\mathbf{E}(\bar{\mathbf{X}}) = \boldsymbol{\mu} \qquad \text{e} \qquad \mathbf{Var}(\bar{\mathbf{X}}) = \frac{\boldsymbol{\sigma}^2}{\boldsymbol{n}}$$

Distribuição da média amostral

Resultado Importante

Quando o tamanho da amostra aumenta, independentemente da distribuição de probabilidades da v.a. de interesse X, a distribuição de probabilidades da v.a. $\bar{X}$ (estimador da média populacional, μ), aproxima-se cada vez mais da distribuição Normal.

Insper

Distribuição da média amostral

O resultado anterior, fundamental na teoria de **Inferência Estatística**, é conhecido como **TEOREMA DO LIMITE CENTRAL (T. L. C.)**.

Insper

Distribuição da média amostral

Teorema do Limite Central (T.L.C.) –

Para uma AAS $\{X_1, X_2, \dots, X_n\}$ retirada de uma população com média μ e variância σ^2, a distribuição de

$$\overline{X} = \sum_{i=1}^{n} \frac{X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

aproxima-se de uma distribuição normal com média μ e variância ${}^{\sigma^2}/_{n}$, quando n tende ao infinito (suficientemente grande).

Insper

Distribuição da média amostral

Corolário 1 – Seja $\{X_1, X_2, \dots, X_n\}$ AAS da v.a. X com média μ $(E(X) = \mu)$ e variância σ^2 $(Var(X) = \sigma^2)$ e seja a média amostral

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

então

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{Var(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \underset{n \to \infty}{\sim} N(0; 1)$$

Distribuição da média amostral

Observação

Caso a variável aleatória de interesse, X, tiver distribuição normal, então

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

em que

$X_1, X_2, ..., X_n$ é uma a.a. da v.a. X,

terá distribuição normal exata.

Esse resultado segue do fato de que a distribuição de uma combinação linear de v.a.´s normais independentes também é normal.

Distribuição da média amostral

Corolário 1 – Seja $\{X_1, X_2, \dots, X_n\}$ amostra aleatória da v.a. X com média μ $(E(X) = \mu)$ e variância σ^2 $(Var(X) = \sigma^2)$ e seja a média amostral

$$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n},$$

então

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{Var(\bar{X})}} = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0; 1)$$

Insper

Exercícios

Exemplo 1

Assuma que a variável de interesse *Grau de satisfação de um passageiro com a experiência proporcionada ao longo da viagem no navio de cruzeiro* tenha média populacional igual a 0,705 e variância populacional igual a 0,1050.

a) Descreva a distribuição do *Grau médio de satisfação*, para uma amostra aleatória de 50 passageiros.

b) Qual a probabilidade de se ter uma média amostral abaixo de 0,65 a partir de uma amostra aleatória de 50 passageiros? Resp.: 11,51%

c) Qual deve ser o tamanho amostral para a probabilidade acima valer exatamente 5%? Resp.: n pelo menos 95

d) Fixando em 99% a probabilidade da média amostral não se afastar, para mais ou para menos, da média populacional, calcule essa distância (margem de erro - denotada por ε) para uma amostra de 50 passageiros. Resp.: ε=0,118

Exemplo 2

As especificações para a fabricação de um certo produto exigem μ = 25 mg e σ = 2 mg de um certo composto *A*. O fabricante deseja examinar uma amostra de cada lote produzido para saber se os produtos se encontram em conformidade com as exigências, no que diz respeito ao valor médio do composto *A*, uma vez que o desvio padrão se encontra dentro dos padrões requeridos.

Qual deve ser o tamanho da amostra que o fabricante precisará analisar, em cada lote, para que o erro amostral seja de, no máximo, 15% da verdadeira variância do composto *A*, para mais ou para menos, em 95% das vezes que tal inspeção for feita?

Resp.: $n \geq 43$

Insper

Exemplo 3

Um processo de produção é verificado periodicamente por um inspetor de controle de qualidade. O inspetor seleciona amostras aleatórias simples de 36 produtos acabados e calcula o peso médio (em kg) dos produtos da amostra.

a) Se os resultados dos testes depois de um longo período de tempo mostram que $P(1{,}90 \leq \bar{X} \leq 2{,}10) = 0{,}90$ e $P(\bar{X} > 2{,}10) = 0{,}05$, sendo $\bar{X}$ uma variável aleatória que denota o peso médio dos produtos acabados, quais são a média e o desvio-padrão para a população de produtos produzidos sob esse processo? Resp: $\mu = 2$ e $\sigma = 0{,}36585$

b) Para que o erro amostral não supere as 100 gramas, para mais ou para menos, em 99% das vezes que o inspetor de controle de qualidade verificar a produção, quantos produtos acabados deverão ser inspecionados? Caso necessite, utilize os resultados encontrados no item anterior. Resp: n pelo menos 89

c) Você teve que fazer algum tipo de suposição para responder aos itens anteriores? Em caso afirmativo, qual? Por quê? Justifique a sua resposta. Resp: X seja exatamente normal ou TLC.

Distribuição Normal : Valores de $P(Z \le z) = A(z)$

Parte inteira e primeira decimal de z

	Segunda decimal de z									
	0	1	2	3	4	5	6	7	8	9
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000