Avaliação e Desempenho
Aula 3 - Simulação

- Validando resultados da simulação
- Média e variância amostral
- Teorema do Limite Central
- Intervalo de confiança
- Organizando as execuções da simulação

Verificando Resultados da Simulação

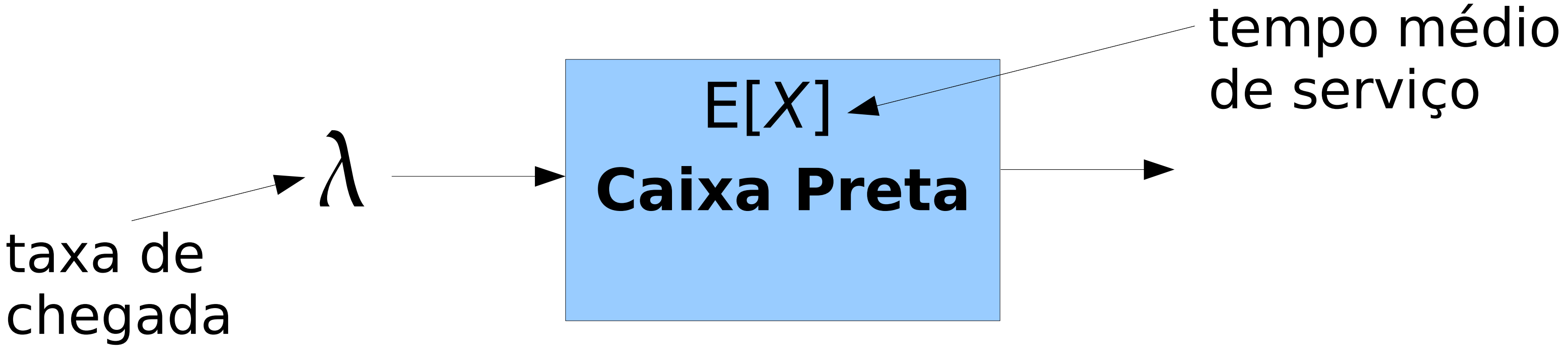
- Como saber que resultado da simulação está correto?
- eterno problema com simulação...

Validar com resultados analíticos conhecidos!

Verificando via Utilização

- Medir fração de tempo sistema ocupado
- Comparar com valor teórico

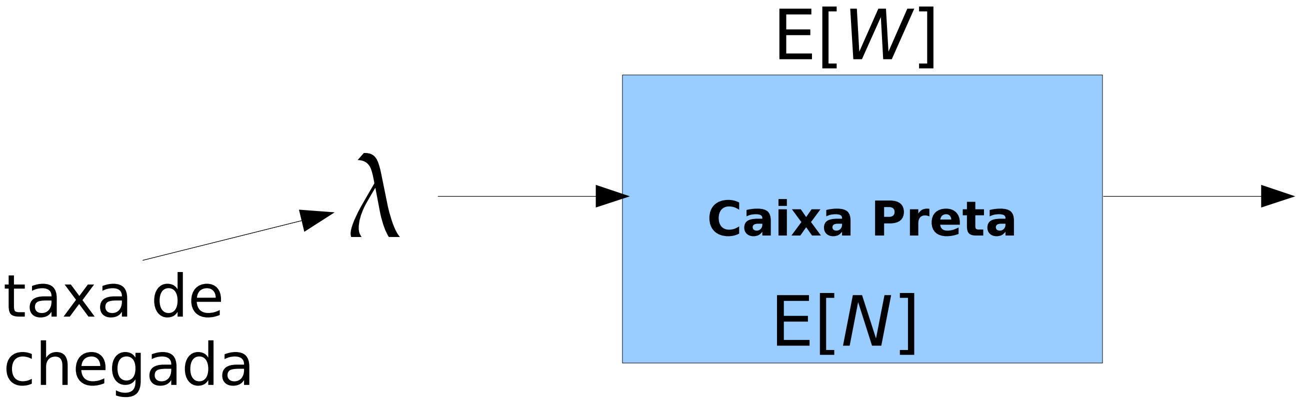
- Qual é a utilização do sistema?



- Utilização: $\lambda \, \mathrm{E}[X]$

Verificando via Little

- Estimar E[W] e E[N] com simulador
- Verificar que resultado de Little é válido



- Resultado de Little: $E[N] = \lambda E[W]$

Resultado da Simulação

- Medida de interesse *X*
  - ex. *X* = “número de pessoas que chegaram e viram a fila vazia nas primeiras 2 horas”
- Resultado da simulação fornece uma amostra de *X*
- O que acontece se executarmos o simulador novamente?
- Outro valor de *X*
  - se usarmos outra semente
  - *independente* do primeiro valor
- X_i : resultado da *i*-ésima execução

Estimando a Medida de Interesse

- Supor medida de interesse X
  - X é uma variável aleatória
- E[X] não é conhecido
- X_i : resultado da i-ésima execução
  - supor n execuções do simulador
- **Como estimar E[X]?**
  - valor esperado de X

Média Amostral

- X_i : resultado da *i*-ésima execução
- $X_1, \ldots, X_n$: sequência de v.a. iid

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Média amostral

variável aleatória:
soma de v.a. é uma outra v.a.

- Quanto vale $E[\overline{X_n}]$?

aplicar definição de valor esperado

$$E[\overline{X_n}] = \mu$$

Qualidade do Estimador

- $\overline{X}_n$: estimador para μ
- Sem tendência (*unbiased*)
  - pois $E[\overline{X}_n] = \mu$
- Qualidade do estimador
  - depende da variância

$$E[(\overline{X}_n - \mu)^2] = Var[\overline{X}_n] = \frac{\sigma^2}{n}$$

variância de *X*

diminui com *n*

- Bom estimador para μ

Soma de Variáveis Aleatórias

- Seja $X_1, X_2, \ldots, X_n$ uma sequencia de v.a. iid com valor esperado $\mu < \infty$ e variância $\sigma^2 < \infty$
- Seja $S_n = \sum_{i=1}^{n} X_i$ a soma da sequência
- Para onde vai esta soma?
  - soma possui valor único (para um dado *n*)?
- Sabemos: $S_n = n\,\overline{X_n} \longrightarrow E[S_n] = n\,\mu \quad Var[S_n] = n\,\sigma^2$
- E sua distribuição?

$$P[S_n \leqslant z] = ?$$

lei dos grandes números

o que podemos afirmar?

Teorema do Limite Central

- Distribuição da soma de v.a. iid converge para distribuição Normal
  - Resultado fundamental em probabilidade

- Seja X_1, X_2, ..., X_n uma sequencia de v.a. iid com valor esperado $\mu < \infty$ e variância $\sigma^2 < \infty$

- Seja $S_n = \sum_{i=1}^{n} X_i$ a soma da sequência

- Então

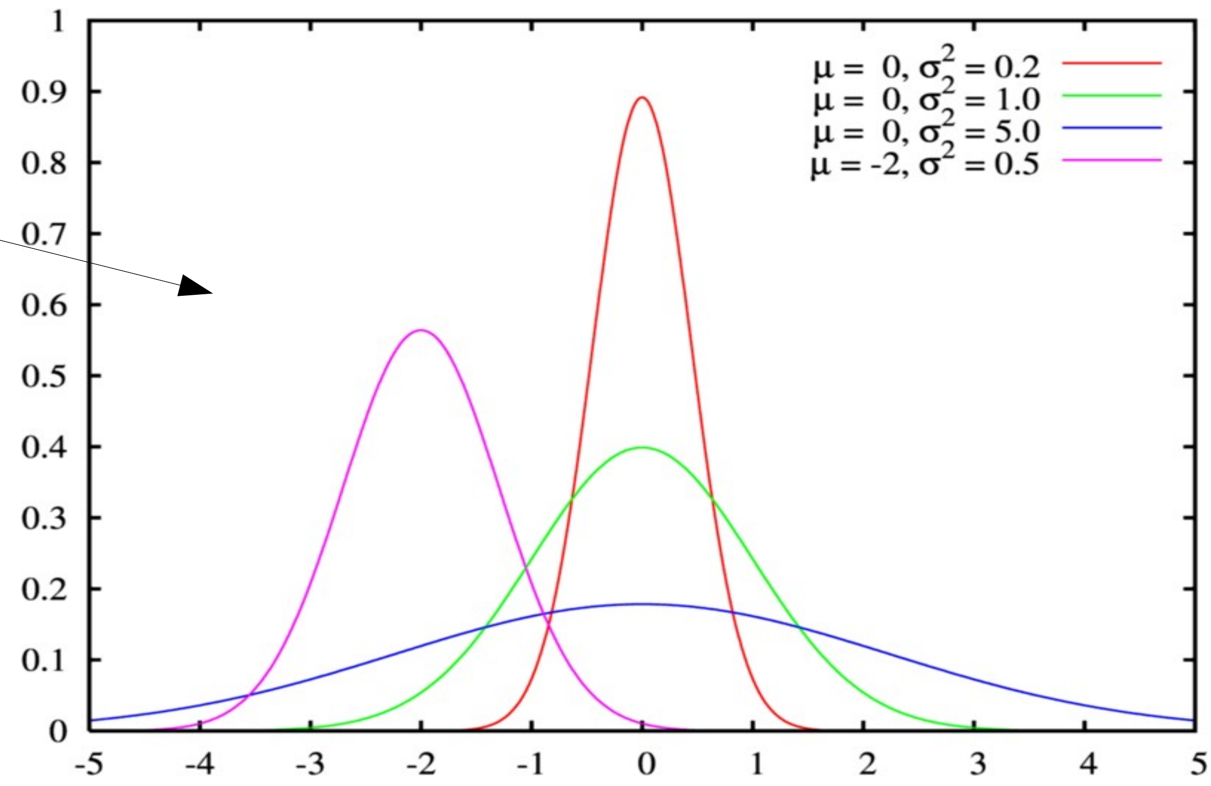
Normal com mesma média e variância

$$P[S_n \leq z] \rightarrow N(n\mu, n\sigma^2)$$ quando $n \rightarrow \infty$

Distribuição Normal

- Importante distribuição (conhecida também como distribuição de Gauss)
- Dois parâmetros: média $\mu < \infty$ e variância $\sigma^2 < \infty$
- Normal padrão: $\mu = 0$, $\sigma^2 = 1$

Exemplos da distribuição Normal (função de densidade) – diferentes parâmetros



Normalização

- Seja X_1, X_2, ..., X_n uma sequencia de v.a. iid com valor esperado $\mu < \infty$ e variância $\sigma^2 < \infty$
- Transformar a soma S_n numa v.a. com valor esperado 0 e variância 1
- Seja $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ a nova sequência
- Sabemos: $E[Z_n] = 0$ $\quad Var[Z_n] = 1$
- Então

Normal padrão – N(0, 1)

$$P[Z_n \leqslant z] \rightarrow \phi(z)$$ quando $n \rightarrow \infty$

Teorema do Limite Central

- Seja $X_1, X_2, \ldots, X_n$ uma sequencia de v.a. iid com valor esperado $\mu < \infty$ e variância $\sigma^2 < \infty$

dividindo numerador e denominador por *n*

média amostral

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

onde $\overline{X}_n = \frac{1}{n} S_n$

- Resultado

função de distribuição cumulativa da Normal padrão

$$\lim_{n\to\infty} P\left[\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leqslant z\right] = \phi(z)$$

convergência em probabilidade

Comparação com Grandes Números

- Qual é a diferença?
- Lei dos grandes números
  - média converge para seu valor esperado

$$\lim_{n \to \infty} P\left[\left| \overline{X}_n - \mu \right| < \epsilon \right] = 1$$

- Teorema do Limite Central
  - distribuição converge para Normal

$$\lim_{n \to \infty} P\left[\frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \leqslant z \right] = \phi(z)$$

Teorema do Limite Central na Prática

- Na prática *n* é finito
  - TLC vale no limite
- Com *n* grande, mas finito resultado é aproximado
  - usaremos esta aproximação

$$P\left[S_n \leq z\right] \rightarrow N\left(n\mu, n\sigma^2\right)$$

quando *n* é *suficientemente* grande (ex. *n* > 30)

Estimando Variância da Medida de Interesse

- $\overline{X_n}$ estimativa para valor esperado da medida de interesse
  - E[X] não é conhecido (objetivo é estimá-lo!)
- Como você estimaria a variância da medida de interesse?
  - Var[X] também não é conhecida

Variância Amostral

- Seja X_1, X_2, ..., X_n uma sequencia de v.a. iid com valor esperado $\mu < \infty$ e variância $\sigma^2 < \infty$
- Seja $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X}_n)^2$
- Então $E[S^2] = \sigma^2$
- S^2 é um estimador não-tendencioso (unbiased) da variância de X

Intervalo de Confiança

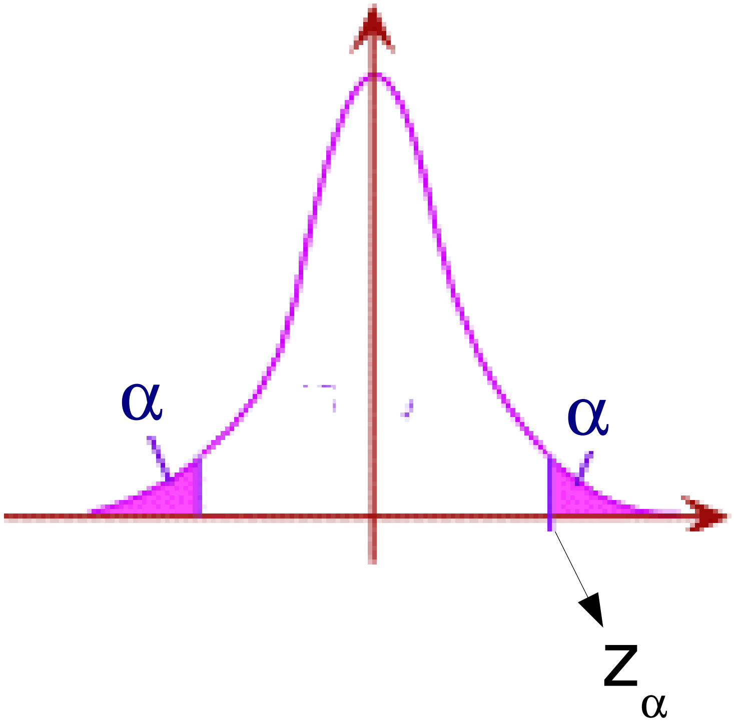
- Média amostral não é igual ao valor esperado
  - *tende* ao valor esperado

Idéia

- Utilizar a média para calcular um *intervalo* onde o valor esperado *pode* estar
- Probabilidade de obter um intervalo que contém o valor esperado
  - depende do tamanho do intervalo
  - intervalo menor, menor probabilidade
- *Confiança* do intervalo

Distribuição Normal (0,1)

- P [Z > z_α] = α
  - Onde α é a área da curva



- A partir da simetria da distribuição Normal temos:

$$P\left[-z_\alpha < Z_n < z_\alpha\right] = 1 - 2\alpha$$

$$P\left[-z_{\alpha/2} < Z_n < z_{\alpha/2}\right] = 1 - \alpha$$

Calculando o Intervalo

- Temos então

$$P[-z_{\alpha/2} < Z_n < z_{\alpha/2}] \sim 1-\alpha$$

$$P[-z_{\alpha/2} < \frac{\overline{X}_n - \mu}{S/\sqrt{n}} < z_{\alpha/2}] \sim 1-\alpha$$

$$P[-z_{\alpha/2} < \frac{\mu - \overline{X}_n}{S/\sqrt{n}} < z_{\alpha/2}] \sim 1-\alpha$$

$$P[\overline{X}_n - \frac{z_{\alpha/2} S}{\sqrt{n}} < \mu < \overline{X}_n + \frac{z_{\alpha/2} S}{\sqrt{n}}] \sim 1-\alpha$$

- A confiança do intervalo $\overline{X}_n \pm z_{\alpha/2} S/\sqrt{n}$ é ~ $1-\alpha$

Aproximando a Normal

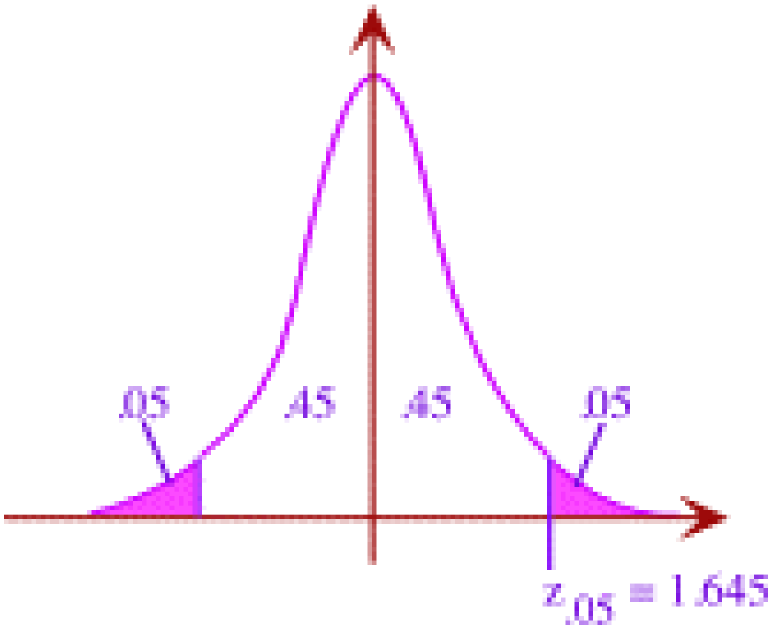
- Seja $Z_n = \dfrac{\overline{X_n} - \mu}{S / \sqrt{n}}$ $\qquad P[Z_n \leqslant z] \sim \phi(z)$

outra aproximação! *S* é uma estimativa do desvio padrão

aproximadamente N(0,1) quando *n* é *suficientemente* grande

Como obter o valor de $z_{\alpha/2}$?

- Suponha intervalo de confiança de 90%
  - Temos que 1-α=0.9, logo α/2=0.05



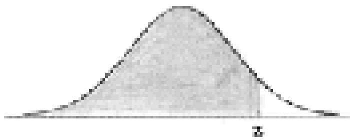
$$P\left[Z_n > z_{0.05}\right] = 0.05$$

$$P\left[Z_n \leq z_{0.05}\right] = 0.95$$

$$z_{0.05} = 1.645$$

$$F_{Z_n}\left(z_{0.05}\right) = 0.95$$

Tabela da Distribuição Normal



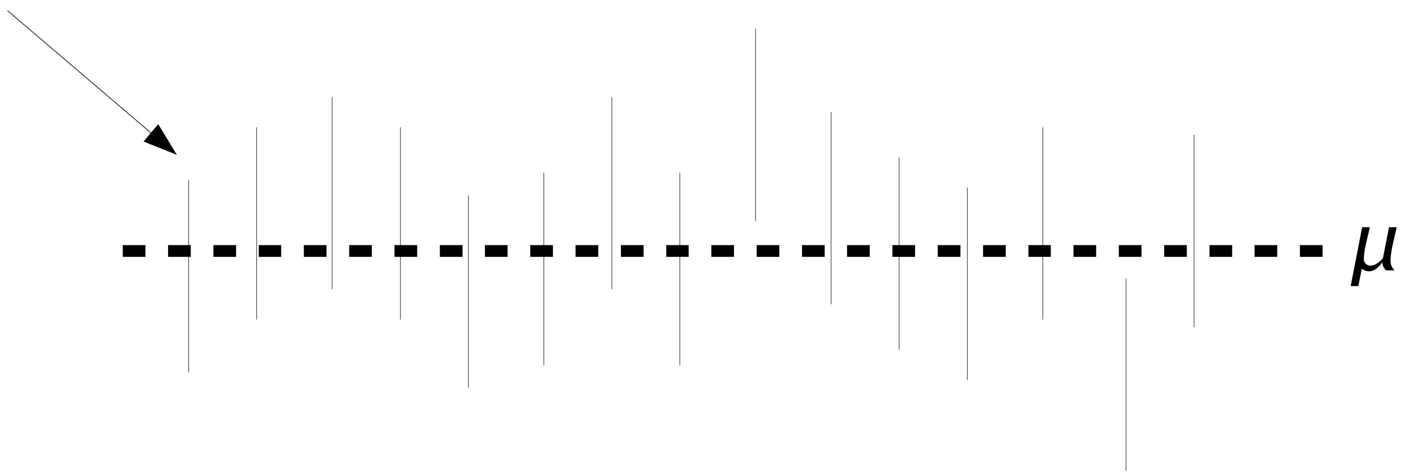
Probability Content from -oo to Z ← $P\left[N(0,1) \leqslant z\right] = \phi(z)$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	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

Explicando Intervalo de Confiança

- Com probabilidade $1-\alpha$ o intervalo gerado contém o valor esperado μ

Intervalos de confiança gerados (dado *n* amostras)



- Invervalo de confiança de 95%
  - boa chance do intervalo gerado conter o valor esperado, mas não é garantido!

Organizando as execuções da simulação

- Objetivo: Estudar o sistema em estado estacionário
- Descartar o período transiente
  - Descartar amostras até que o evento de menor taxa tenha ocorrido uma centena de vezes.
- Requisitos quando aproximamos $Z_n = \dfrac{\overline{X_n} - \mu}{S/\sqrt{n}}$ por uma v.a. Normal
  - As v.a. X_1, X_2, ..., X_n devem ser iid e ter média e variância finitas (teorema do limite central)

Método de Replicações Independentes

- Executar a simulação diversas vezes de forma que os valores coletados para os X_i's sejam independentes.
- Execução 1 → $X_1 = 1/m \sum_{j=1}^{m} X_{1j}$
- Execução 2 → $X_2 = 1/m \sum_{j=1}^{m} X_{2j}$
- Execução n → $X_n = 1/m \sum_{j=1}^{m} X_{nj}$

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X_n})^2$$

Critério de Parada

- Escolher valores iniciais para α (confiança do intervalo) e l (largura do intervalo)
- Calcule S e $\overline{X}_n$ a partir das n amostras de tamanho m:

$$\underbrace{X_{11} \ldots X_{1m}}_{X_1} \quad \underbrace{X_{21} \ldots X_{2m}}_{X_2} \quad \quad \underbrace{X_{n1} \ldots X_{nm}}_{X_n}$$

- Se $2 z_{\alpha/2} S / \sqrt{n} < l$ então parar a simulação
- senão aumentar m e/ou n