Revisão de Probabilidade e Estatística

1 Principais Ramos da Estatística

Estatística Descritiva

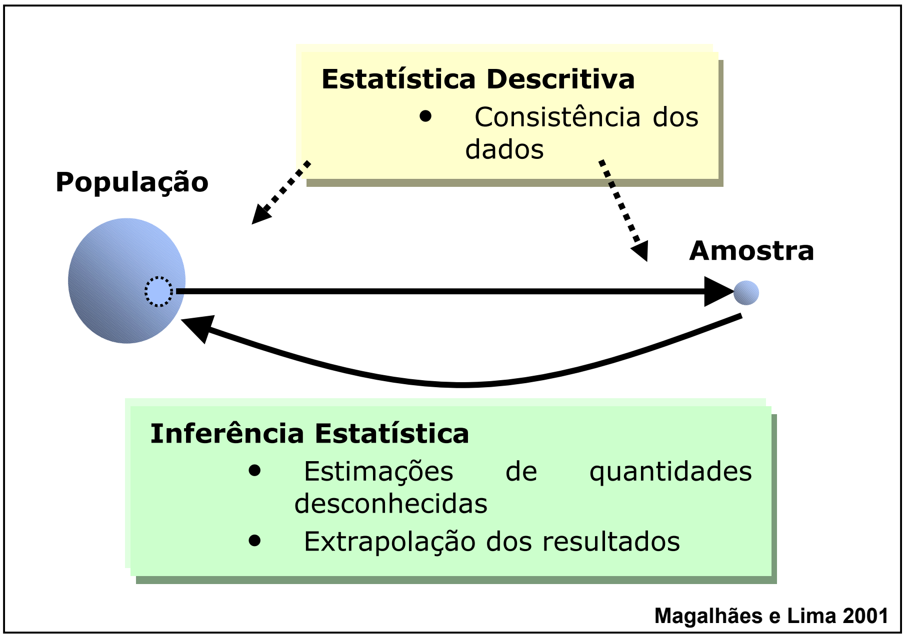
Utilizada na etapa inicial da análise de dados com o objetivo de tirar conclusões iniciais.

Probabilidade

Teoria matemática utilizada para estudar a incerteza decorrente de fenômenos de caráter aleatório.

Inferência Estatística

Estudo de técnicas que permitem a extrapolação, a um grande volume de dados, denominado **população**, de informações e conclusões obtidas de um subconjunto menor de valores, denominado **amostra**.



2 Conceitos de Probabilidades

Experimento

Experimento é um processo cuja saída não é conhecida com certeza. O conjunto de todos os valores possíveis do experimento são chamados de Espaço Amostral.

Espaço Amostral

Espaço Amostral, indicado por S, é o conjunto de todos resultados possíveis de um certo fenômeno aleatório.

Eventos

Eventos ou pontos amostrais, são subconjuntos do espaço amostral.

Variável aleatória

Variável aleatória é uma função que atribui um número real a cada ponto do espaço amostral.

Variável aleatória discreta

Variável aleatória discreta é uma variável aleatória que assume valores enumeráveis x_1, x_2, ..., isto é, existe uma correspondência 1 a 1 com conjunto dos números inteiros.

Exemplo 1:

No lançamento de uma moeda o espaço amostral é S ={cara,coroa}. A variável aleatória X que indica o valor do resultado pode ser igual a 1 se Cara e 2 se Coroa.

Exemplo 2:

No lançamento de dois dados o espaço amostral é S ={(1,1),(1,2),(1,3),...(6,6)}. A variável aleatória discreta X, definida como a soma dos dois valores, assume o valor 7 se o resultado for (3,4).

Exemplo 3:

Na chegada de clientes a um banco, o intervalo de tempo entre duas chegadas é uma variável aleatória que assume valores reais positivos.

Probabilidade

Uma função de P[.) é denominada **Probabilidade** se atribui valores numéricos aos eventos do espaço amostral de acordo com as seguintes condições:

(i) $0 \leq P[A] \leq 1$, qualquer $A \subset S$.

(ii) $P[S] = 1$

(iii) $P[\bigcup_{j=1}^{n} A_j] = \sum_{j=1}^{n} P[A_j]$ com os A_j disjuntos

Exemplo 4:

No lançamento de um dado podem ocorrer os valores 1, 2, 3, 4, 5 e 6. O espaço amostral S={1, 2, 3, 4, 5, 6}. A probabilidade de ocorrência de cada evento é p[1]=1/6, p[2]=1/6, p[3]=1/6, p[4]=1/6, p[5]=1/6, p[6]=1/6.

2.1 Propriedades de Probabilidades

Sejam A e B eventos de S:

Soma de Probabilidades

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

Produto de Probabilidades

$P[A \cap B] = P[A \mid B]\, P[B]$ quando $P[B]>0$

Probabilidade Condicional

$P[A \mid B] = P[A \cap B)/P[B)$ quando $P[B)>0$

Independência de Eventos

O evento A não depende do evento B quando $P[A \mid B) = P[A)$ quando $P[B]>0$

Partição do Espaço Amostral

$C_1, C_2, \ldots C_k$ é uma partição do Espaço Amostral S se eles não tem intercessão entre si e a união é S.

2.2 Funções de Probabilidade e de Distribuição

A **função de distribuição** $F(x)$ de uma variável aleatória X é definida como:

$F(x) = P[X \le x]$ para $-\infty<x< \infty$

A função distribuição $F(x)$ satisfaz às seguintes propriedades:

(i) $0\le F(X) \le 1$, para todo X

(ii) $F(X)$ é não decrescente, isto é, se $x_1 < x_2$ então $F(x_1) < F(x_2)$

(iii) $\lim_{x\to-\infty} F(x) = 0$ e $\lim_{x\to\infty} F(x) = 1$

2.3 Funções Discretas de Probabilidade e Distribuição

Função Massa de probabilidade

A **função (massa) de probabilidade** $p[x]$, no caso em que **X** é uma variável aleatória discreta, é definida como

$p[x_i] = P[X=x_i]$ $\quad$ $i=1,2,3,\ldots$

que satisfaz a $0 \le p[x_i] \le 1$ e $\sum_{i=1}^{\infty} p(x_i) = 1$

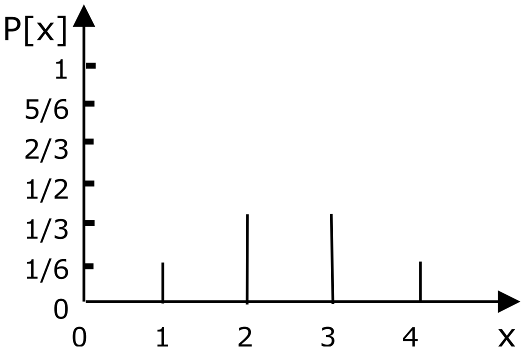
Se $I=[a,b]$ então $P(X \in I) = \sum_{a\le x_i \le b} p(x_i)$

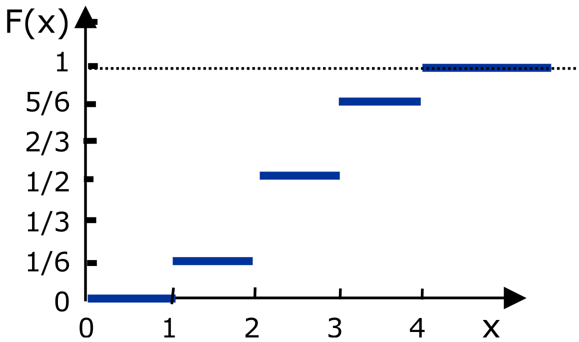
Função de Distribuição

A **função de distribuição F(x)** é dada por $F(x) = \sum_{x_i \le x} p(x_i)$ para -∞<x< ∞

Exemplo 5:

Em um sistema de estoque, a demanda por um produto é uma variável discreta que assume valores 1, 2, 3, 4 com probabilidades 1/6, 1/3, 1/3 e 1/6 respectivamente.





2.4 Funções Contínuas de Probabilidade e Distribuição

Função de densidade de probabilidade

Se X é uma variável aleatória contínua então para qualquer conjunto de números reais B existe a **função de densidade de probabilidade** f(x), tal que

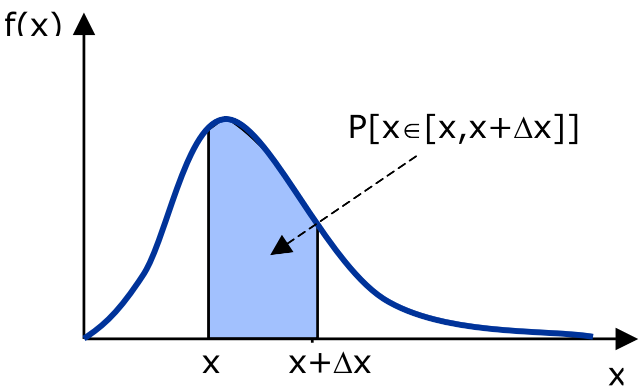
$$P[X \in B] = \int_B f(x)dx \qquad \text{e} \qquad \int_{-\infty}^{\infty} f(x)dx = 1$$

Qualquer x e $\Delta x > 0$ então $P(X \in [x, x + \Delta x]) = \int_x^{x+\Delta x} f(y)dy$

Função de Distribuição

A **função de distribuição** F(x) é dada por

$$F(x) = P[X \in [-\infty, x]] = \int_{-\infty}^{x} f(y)dy \qquad \text{para } -\infty<x<\infty$$



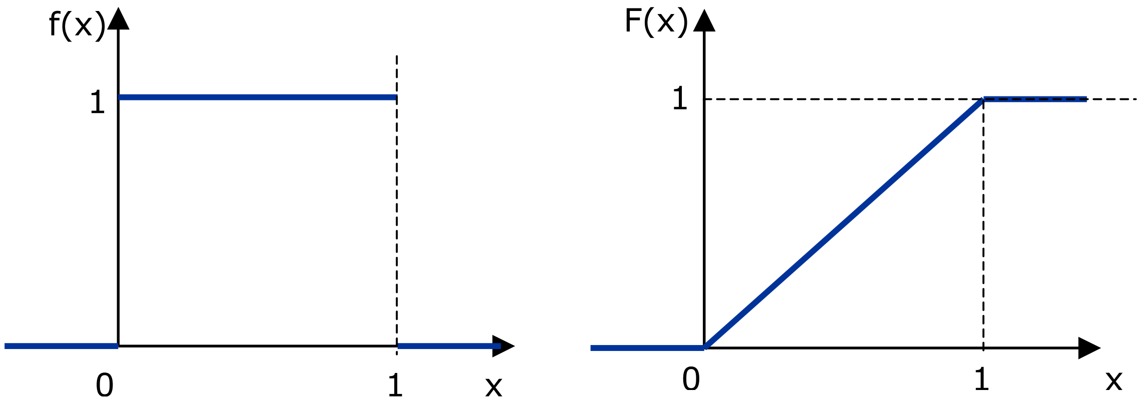
Exemplo 6:

Uma variável aleatória uniforme no intervalo [0,1] tem a função densidade de probabilidade:

$$F(x) = \begin{cases} 1 & \text{se } 0 \le x \le 1 \\ 0 & \text{Caso contrário} \end{cases}$$

No caso de $0 \le x \le 1$ tem-se $F(x) = \int_0^x f(y)dy = \int_0^x 1dy = x$

Os gráficos das funções do exemplo são os seguintes:



2.5 Função de Probabilidade Conjunta

Função de Probabilidade Conjunta – Caso Discreto

Se X e Y são variáveis aleatórias discretas então $p[x,y] = P[X=x, Y=y]$ para todo x,y onde $p[x,y]$ é denominada **função de probabilidade conjunta.**

X e Y são independentes se $p[x,y] = p_X[x]\, p_Y[y]$ para todo x,y onde

$$p_X[x] = \sum_{todo-y} p[x,y] \qquad \text{e} \qquad p_y[y] = \sum_{todo-x} p[x,y]$$

são as probabilidades (marginais) de X e Y.

Exemplo 7:

Supondo que X e Y sejam variáveis aleatórias discretas conjuntas com

$$p[x,y] = \begin{cases} \dfrac{xy}{27} & \text{para } x=1,2 \text{ e } y=2,3,4 \\ 0 & \text{caso contrário} \end{cases}$$

$$p_X[x] = \sum_{y=2}^{4} \frac{xy}{27} = \frac{x}{3}$$ para $x=1,2$

$$p_y[y] = \sum_{x=1}^{2} \frac{xy}{27} = \frac{y}{9}$$ para $y=2,3,4$

Considerando que $p[x,y]=xy/27=p_x[x]p_Y[y]$ para todo x,y, as variáveis aleatórias X e Y são independentes.

Função Densidade de Probabilidade Conjunta – Caso Contínuo

No caso em que X e Y são variáveis contínuas, se existe a função não negativa $f(x,y)$, chamada **função densidade de probabilidade conjunta** de X e Y, tal que para todos os conjuntos de números reais A e B tem-se

$$P[X \in A, Y \in B] = \int_A \int_B f(x,y)dxdy$$

Neste caso X e Y são independentes se $f(x,y) = f_X(x)f_Y(y)$ para todo x,y onde

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad \text{e} \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

são as funções densidade de probabilidade de X e Y.

Exemplo 8:

Sendo X e Y variáveis aleatórias contínuas conjuntas com

$$f(x,y) = \begin{cases} 24xy & \text{para } x \geq 0,\ y \geq 0,\ \text{e } x+y \leq 1 \\ 0 & \text{caso contrário} \end{cases}$$

então

$$f_X(x) = \int_0^{1-x} 24xydy = 12xy^2\Big|_0^{1-x} = 12x(1-x)^2 \qquad 0 \leq x \leq 1$$

e

$$f_Y(y) = \int_0^{1-y} 24xydx = 12x^2y\Big|_0^{1-y} = 12y(1-y)^2 \qquad 0 \leq y \leq 1$$

Como $f(\frac{1}{2},\frac{1}{2}) = 6 \neq \left(\frac{3}{2}\right)^2 = f_X(\frac{1}{2})f_Y(\frac{1}{2})$ então X e Y não são independentes.

2.6 Média e Desvio Padrão

A **média** ou **valor esperado** de uma variável aleatória X_i (onde i=1,2,...,n) é indicada como μ_i ou E[Xi] é definida como

$$\mu_i = \begin{cases} \sum_{j=1}^{\infty} x_j p_{x_i}(x_j) & \text{se } X_i \text{ é discreta} \\ \int_{-\infty}^{\infty} x f_{x_i}(x)dx & \text{se } X_i \text{ é contínua} \end{cases}$$

Exemplo 9:

Considerando a variável discreta que assume valores 1, 2, 3, 4 com probabilidades 1/6, 1/3, 1/3 e 1/6 do exemplo 5:

$$\mu = 1(\frac{1}{6}) + 2(\frac{1}{3}) + 3(\frac{1}{3}) + 4(\frac{1}{6}) = \frac{5}{6}$$

Exemplo 10:

Para a distribuição uniforme entre [0,1] do exemplo 6:

$$\mu = \int_0^1 xf(x)dx = \int_0^1 xdx = \frac{1}{2}$$

Propriedades da Média

1. E[cX] = cE[X]
2. $E[\sum_{j=1}^{n} c_i X_i] = \sum_{j=1}^{n} c_i E[X_i]$ mesmo se X_i forem dependentes

2.7 Variância

A **variância** de uma variável aleatória X_i (onde i=1,2,...,n) é indicada como σ_i^2 ou $Var[X_i]$ é definida como

$$\sigma_i^2 = E[(X_i - \mu_i)^2] = E[X_i^2] - \mu_i^2$$

Desvio Padrão é definido como σ_i

Exemplo 11:

Considerando os valores dos exemplos 5 e 9

$$E[X^2] = 1^2(\frac{1}{6}) + 2^2(\frac{1}{3}) + 3^2(\frac{1}{3}) + 4^2(\frac{1}{6}) = \frac{43}{6}$$

$$\mu = E[X] = 1(\frac{1}{6}) + 2(\frac{1}{3}) + 3(\frac{1}{3}) + 4(\frac{1}{6}) = \frac{5}{6}$$

$$Var[X] = E[X^2] - \mu^2 = \frac{43}{6} - (\frac{5}{2})^2 = \frac{11}{12}$$

Exemplo 12:

Para a distribuição uniforme entre [0,1] dos exemplos 6 e 10

$$E[X^2] = \int_0^1 x^2 f(x)dx = \int_0^1 x^2 dx = \frac{1}{3}$$

$$\mu = E[X] = \int_0^1 xf(x)dx = \int_0^1 xdx = \frac{1}{2}$$

$$Var[X] = E[X^2] - \mu^2 = \frac{1}{3} - (\frac{1}{2})^2 = \frac{1}{12}$$

Propriedades da Variância

1. $Var[X] \geq 0$
2. $Var[cX] = c^2 Var[X]$
3. $Var[\sum_{j=1}^{n} X_i] = \sum_{j=1}^{n} Var[X_i]$ se X_i forem independentes.

2.8 Covariância

A **covariância**, indicada como $\mathbf{C_{ij}}$ ou $Cov(X_i,X_j)$, é uma medida da dependência linear entre as variáveis aleatórias X_i e X_j $(i,j=1,2,...,n)$, sendo definida como

$$C_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] = E[X_i X_j] - \mu_i \mu_j$$

Exemplo 13:

Considerando-se as variáveis aleatórias conjuntas contínuas X e Y do exemplo 8

$$E[XY] = \int_0^1 \int_0^{1-x} xyf(x,y)dydx = \int_0^1 x^2 (\int_0^{1-x} 24y^2 dy)dx$$

$$= \int_0^1 8x^2(1-x^3)dx = \frac{2}{15}$$

$$E[X] = \int_0^1 xf_X(x)dx = \int_0^1 12x^2(1-x)^2 dx = \frac{2}{5}$$

$$E[Y]=\int_0^1 y f_Y(y)dy=\int_0^1 12y^2(1-y)^2dy=\frac{2}{5}$$

$$Cov(X,Y)=E[XY]-E[X]E[Y]=\frac{2}{15}-(\frac{2}{5})(\frac{2}{5})=\frac{2}{75}$$

Propriedades da Covariância

a) Se $C_{ij} = 0$ as variáveis aleatórias X_i e X_j são denominadas **não correlacionadas**.
b) Se $C_{ij} < 0$ as variáveis aleatórias X_i e X_j são denominadas **negativamente correlacionadas**.
c) Se $C_{ij} > 0$ as variáveis aleatórias X_i e X_j são denominadas **positivamente correlacionadas**.
d) Se X_i e X_j são variáveis aleatórias independentes então $C_{ij} = 0$. O inverso não é verdade.

O valor da correlação é um valor com dimensão. Para se obter um valor sem dimensão utiliza-se o índice de correlação ρ_{ij} definido como

$$\rho_{ij}=\frac{C_{ij}}{\sqrt{\sigma_i^2\sigma_j^2}}$$

2.9 Processos Estocásticos

Um **processo estocástico** é uma coleção de variáveis aleatórias similares ordenadas no tempo, todas definidas em um espaço amostral comum. O conjunto de todos os valores que estas variáveis podem assumir é denominado **espaço de estado**.

Se a coleção é $X_1, X_2, \ldots$, então o processo estocástico é de **tempo-discreto**.

Se a coleção é $\{ X(t), t \geq 0\}$, então o processo estocástico é de **tempo-contínuo**.

Exemplo 14:

Uma fila simples tal como a fila M/M/1 com tempos de chegada IID (Independentes e Identicamente Distribuídas) $A_1, A_2, \ldots$, e tempos de serviço IID $S_1, S_2, \ldots$, então podemos definir os atrasos na fila como o processo estocástico de tempo-discreto $D_1, D_2, \ldots$, onde

$D_1=0$
$D_{i+1} = \max\{D_i + S_i - A_{i+1},0\}$ para $i = 1,2,\ldots$

Desta forma a simulação mapeia as variáveis aleatórias de entrada em um processo estocástico de saída $D_1, D_2, \ldots$,. O espaço de estado é o conjunto de números reais não negativos. D_i e D_{i+1} são variáveis aleatórias positivamente correlacionadas.

Exemplo 15:

Na fila do Exemplo 14, seja $Q(t)$ o número de clientes na fila no instante t. Então $\{Q(t), t \geq 0\}$ é um processo estocástico de tempo-contínuo com espaço de estado 0, 1, 2, ...

Em alguns casos práticos, para tornar a análise estatística possível, supomos que algumas propriedades do processo estocástico são válidas, tais como a propriedade covariância-estacionária.

Um processo é dito de covariância-estacionária se

$\mu_i = \mu$ para $i=1,2,\ldots$ e $-\infty < \mu < \infty$.

$\sigma_i^2 = \sigma^2$ para $i=1,2,\ldots$ e $\sigma^2 < \infty$

e $C_{i,i+j} = Cov(X_i, X_{i+j})$ são independentes de i para $i=1,2,\ldots$

No caso de covariância-estacionária, a covariância e a correlação entre X_i e X_{i+j}, indicadas como C_j e ρ_j são

$$C_j = C_{i,i+j}$$

$$\rho_j = \frac{C_{i,i+j}}{\sqrt{\sigma_i^2 \sigma_{i+j}^2}} = \frac{C_j}{\sigma^2} = \frac{C_j}{C_0}$$

Se X_1, X_2, ..., é um processo estocástico começando no tempo 0, é provável que a covariância não seja estacionária. Entretanto, após algum tempo de simulação, isto é, para k suficientemente grande, X_k, X_{k+1}, ..., serão aproximadamente estacionários. O valor de k para atingir este ponto define o período de "aquecimento" do sistema ("warmup").

3 Estimadores e Estimativa

Parâmetros

Parâmetros são atributos da população, em geral desconhecidos, e sobre os quais temos interesse de estudo.

Estimador

Estimador é um representante de um parâmetro obtido através de uma amostra.

Estimativa

Estimativa é um valor numérico assumido pelo estimador.

Um **estimador** $\hat{\theta}$ de um parâmetro θ denomina-se **não viciado** se $E[\hat{\theta}] = \theta$

Serão estudados estimadores para dois casos diferentes:

- Variáveis aleatórias Independentes e Identicamente Distribuídas (IID).
- Variáveis aleatórias de um processo estocástico covariante-estacionário.

O segundo caso tem interesse à análise de dados de saída que, em geral, não são independentes.

3.1 Variáveis Aleatórias Identicamente Distribuídas IID

3.2 Estimativa da Média

Supondo que X_1, X_2,..., X_n sejam variáveis aleatórias IID com média da população finita μ e variância σ^2. Então a média

$$\overline{X}(n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

é um estimador não viciado de μ, isto é, $E[\overline{X}(n)] = \mu$

Intuitivamente, isto significa que se fizermos um número grande de experimentos independentes e calcularmos o $\overline{X}(n)$ para cada experimento, a média dos $\overline{X}(n)$ será μ.

Estimativa da Variância

De forma similar, a variância da amostra $S^2(n)$ é calculada como

$$S^2(n) = \frac{\sum_{i=1}^{n} [X_i - \overline{X}(n)]^2}{n-1}$$

é um estimador não viciado de σ^2 pois $E[S^2(n)] = \sigma^2$

Os estimadores $\overline{X}(n)$ e $S^2(n)$ são indicados muitas vezes como $\hat{\mu}$ e $\hat{\sigma}^2$.

A dificuldade de se trabalhar com estas estimativas é não se saber o quanto estão próximas do valor μ. Para isto será definido o intervalo de confiança.

Antes disso será feita a estimativa de $Var[\overline{X}(n)]$

$$Var[\overline{X}(n)] = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2} Var(\sum_{i=1}^{n} X_i)$$

Sendo X_i independentes

$$Var[\overline{X}(n)] = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

Da fórmula $Var[\overline{X}(n)] = \frac{\sigma^2}{n}$ pode-se observar que,quanto maior o valor de n, menor será $Var[\overline{X}(n)]$ e, em conseqüência, $\overline{X}(n)$ estará mais próximo de μ.

Além disso, podemos obter um estimador não viciado de $Var[\overline{X}(n)]$ substituindo σ^2 por $S^2(n)$.

$$\hat{Var}[\overline{X}(n)] = \frac{S^2(n)}{n} = \frac{\sum_{i=1}^{n}[X_i - \overline{X}(n)]^2}{n(n-1)}$$

3.3 Variáveis Aleatórias de um Processo Estocástico Covariante-estacionário

Quando as variáveis aleatórias $X_1, X_2, ..., X_n$ não forem IID mas definirem um processo estocástico com co-variância estacionária então a média amostral $\overline{X}(n)$ ainda é um estimador não viciado de μ mas a variância $S^2(n)$ não é mais um estimador não viciado de σ^2 pois pode-se mostrar que:

$$E[S^2(n)] = \sigma^2[1 - 2\frac{\sum_{j=1}^{n-1}(1 - j/n)\rho_j)}{n-1}]$$

Se $\rho_j > 0$ (correlação positiva), que é um caso comum na prática, então $E[S^2(n)] < \sigma^2$.

A estimativa da variância da média amostral $Var[\overline{X}(n)]$, quando $X_1, X_2, ..., X_n$ são variáveis aleatórias de um processo estocástico com co-variância estacionária, é:

$$Var[\overline{X}(n)] = \sigma^2 \frac{[1 + 2\sum_{j=1}^{n-1}(1 - j/n)\rho_j)]}{n}$$

Assim, estimar $Var[\overline{X}(n)]$ por $S^2(n)/n$ resulta em duas fontes de erros:

- $S^2(n)$ é um estimador viciado de σ^2 e
- Os termos de correlação foram negligenciados na fórmula acima.

As estimativas de ρ_j (para $j= 1,2,...,n-1$) podem ser calculadas como:

$$\hat{\rho}_j = \frac{\hat{C}j}{S^2(n)} \quad \text{e} \quad \hat{C}_j = \frac{\sum_{i=1}^{n-j}[X_i - \overline{X}(n)][X_i - \overline{X}(n)]}{n-j}$$

O problema com estes estimadores é que são viciados e possuem uma variância grande, a menos que n seja muito grande, e são correlacionados entre si, isto é, $Cov(\hat{\rho}_j, \hat{\rho}_k) \neq 0$.

Estas considerações mostram a dificuldade de se analisar os dados de saída por serem correlacionados.

4 Distribuição Normal

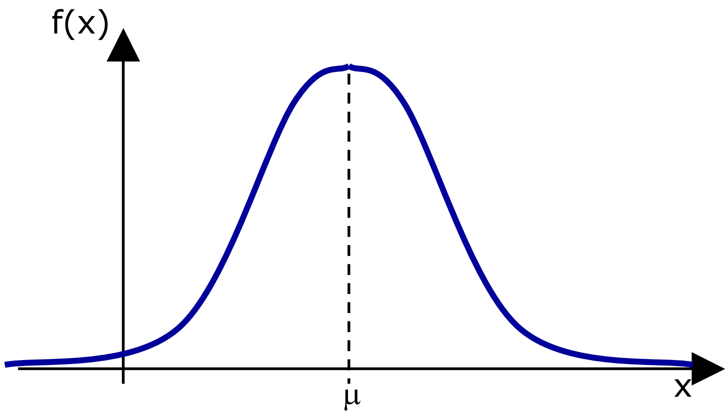
Uma variável aleatória X tem distribuição Normal com média μ e variância σ^2 se a sua função densidade é

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

Utiliza-se a notação $X \sim N(\mu,\sigma^2)$.

Propriedades:

a) $f(x)$ é simétrica em relação à μ
b) $f(x) \to 0$ quando $x \to \pm\infty$
c) O máximo de $f(x)$ ocorre em $x = \mu$



O cálculo de probabilidades é feito através da integral da função $f(x)$

$$P(a \le x \le b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

Esta integral tem solução aproximada e existe disponível a tabela de $N(0,1)$, isto é, função distribuição Normal com média 0 e variância 1, que é denominada **Normal Padrão** ou **Reduzida**.

Para obter as probabilidades considerando outros valores de média e variância deve ser feita uma transformação como a seguir.

Seja a variável X com distribuição $N(\mu,\sigma^2)$, isto é $E[X] = \mu$ e $Var[X] = \sigma^2$.
Definimos uma variável $Z=(X-\mu)/\sigma$ que tem média 0 e variância 1 como é mostrado a seguir:

$$E[Z]=E[(X-\mu)/\sigma]= E[X-\mu]/\sigma=(E[X]-\mu)/\sigma=0$$

$$Var[Z]=Var[(X-\mu)/\sigma]= Var[X-\mu]/\sigma2 =Var[X]/\sigma2=1$$

Pode-se verificar que z tem distribuição $Normal(0,1)$.
Para determinar $P[a \le X \le b]$ faz-se

$$P(a \le X \le b) = P(a-\mu \le X-\mu \le b-\mu)$$
$$= P(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}) = P(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma})$$

Exemplo 16:

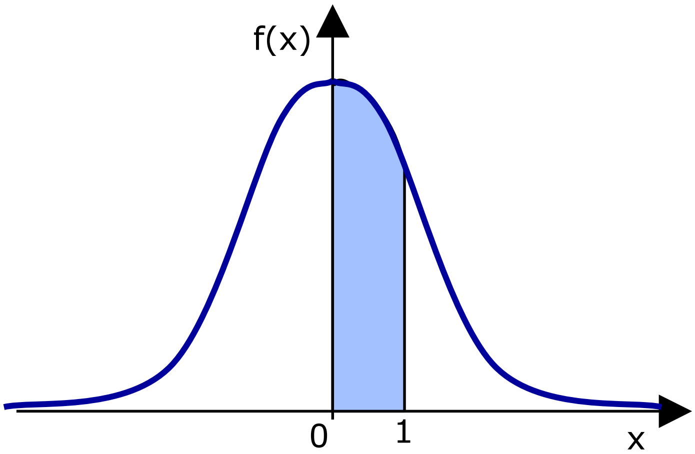
Considerando que X tem distribuição $N(2,9)$, para determinar $P[2 \le X \le 5]$ faz-se:

$$P(2 \le X \le 5) = P(\frac{2-2}{\sqrt{9}} \le Z \le \frac{5-2}{\sqrt{9}}) = P(0 \le Z \le 1)$$

Neste caso basta procurar o valor correspondente na tabela que é $P[0 \le Z \le 1] = 0{,}3413$.

Notar que as tabelas $N(0,1)$ fornecem $P[0 \le Z \le c]$.

Distribuição Normal Padrão



5 Intervalo de Confiança e Testes de Hipóteses

Sejam X_1, X_2, ..., X_n variáveis aleatórias IID com média da população finita μ e variância σ^2 ($\sigma^2>0$) . Veremos como construir um intervalo de confiança de μ e como testar a hipótese de $\mu = \mu_0$.

5.1 Teorema Central do Limite

Seja a variável aleatória Z_n definida como $z_n = \frac{\overline{X}(n)-\mu}{\sqrt{\sigma^2/n}}$ e seja $F_n(z)$ a função de distribuição de Z_n para uma amostra de tamanho n, isto é, $F_n(z)=P[Z_n \le z]$.
Então $F_n(z) \to \Phi(z)$ quando $n \to \infty$, onde $\Phi(z)$ é a função de distribuição normal de uma variável aleatória Z com média 0 e variância 1, isto é, $N(0,1)$.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-y^2}{2}} dy \qquad \text{para } -\infty<z<\infty$$

Na prática o teorema diz que quando n for suficiente grande, a variável aleatória Z_n será distribuída aproximadamente como uma variável com distribuição normal, independente da distribuição das variáveis X_i.
Também pode ser demonstrado que quando n for grande então a média amostral $\overline{X}(n)$ tem distribuição aproximadamente normal com média μ e variância σ^2/n.
A dificuldade de utilizar estes resultados é não se conhecer o valor da variância σ^2. Neste caso se utilizará $S^2(n)$ que converge para σ^2 quando n se torna grande.

O novo enunciado do teorema, com esta alteração, ficará "Quando n for suficientemente grande, a variável $t_n = \dfrac{\overline{X}(n)-\mu}{\sqrt{S^2(n)/n}}$ terá distribuição aproximada à de uma variável com distribuição normal $N(0,1)$."

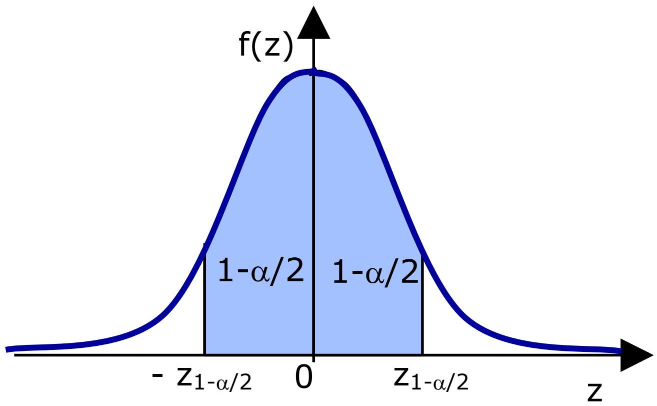
5.2 Intervalo de Confiança

Será considerada a variável aleatória Z_n, definida no Teorema Central do Limite, com distribuição Normal $N(0,1)$.
Fixado um valor α tal que $0 < \alpha < 1$, podemos encontrar um valor $z_{1-\alpha/2}$ tal que

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

Neste caso, dado α procura-se na tabela de $N(0,1)$ o valor de $z_{1-\alpha/2}$ tal que

$$P[-z_{1-\alpha/2}< Z_n<z_{1-\alpha/2})=1- \alpha.$$



Em lugar de z_n usaremos t_n definido pela fórmula

$$t_n = \frac{\overline{X}(n)-\mu}{\sqrt{S^2(n)/n}}$$

Neste caso tem-se

$$P[-z_{1-\alpha/2} \le \frac{\overline{X}(n)-\mu}{\sqrt{S^2(n)/n}} \le z_{1-\alpha/2}] = 1-\alpha$$

que pode re-escrita como

$$P[\overline{X}(n) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{S^2(n)}{n}} \le \mu \le \overline{X}(n) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{S^2(n)}{n}}] = 1-\alpha$$

Assim, para n suficientemente grande, o intervalo com $100(1-\ \alpha)$ porcento de confiança para μ, é definido como

$$\overline{X}(n) \pm z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}$$

ou

$$[\overline{X}(n) - z_{1-\frac{\alpha}{2}}\sqrt{\frac{S^2(n)}{n}}, \overline{X}(n) + z_{1-\frac{\alpha}{2}}\sqrt{\frac{S^2(n)}{n}}]$$

Dado o conjunto X_1, X_2, ..., X_n de variáveis, chamamos

$l(n,\alpha) = \overline{X}(n) - z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}$ limite inferior do intervalo de confiança e

$u(n,\alpha) = \overline{X}(n) + z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}$ limite superior do intervalo de confiança e

o intervalo de confiança será $[l(n,\ \alpha), u(n,\ \alpha)]$.

A interpretação para o intervalo de confiança é:

"Se construirmos um número grande de intervalos de confiança 100(1- α), independentes e baseados em n observações, para n suficientemente grande, a proporção desses intervalos que contem μ é (1-α). Esta proporção define a cobertura do intervalo de confiança"

O intervalo de confiança dá uma idéia de quão preciso é o valor de μ.

A construção do intervalo de confiança depende da escolha de um n "suficientemente grande". Quanto mais assimétrica for a distribuição dos X_i's maior deve ser o valor de n. Se n não for suficientemente grande, o intervalo de confiança será aproximado.

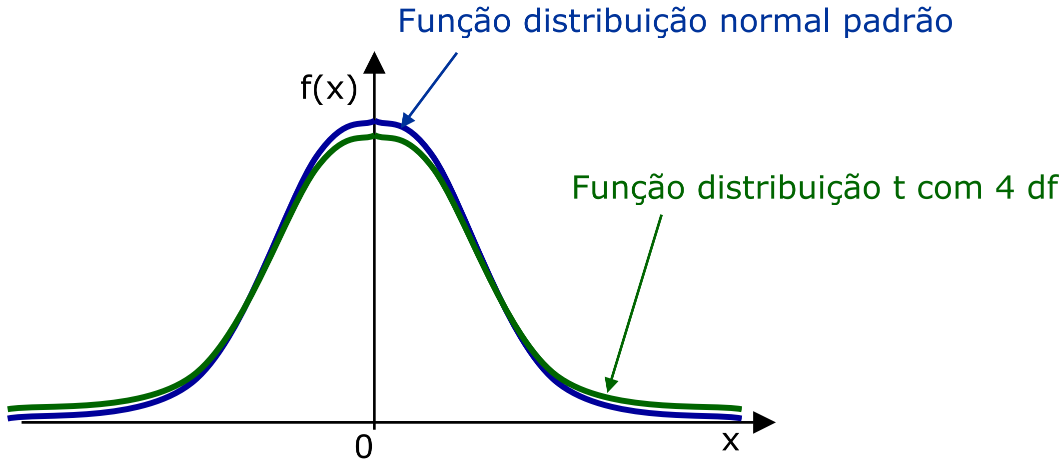
Tem-se uma forma alternativa para determinar o intervalo de confiança. Se as variáveis X_i's têm distribuição normal, então $t_n = [\overline{X}(n) - \mu] / \sqrt{S^2(n)/n}$ tem distribuição t (t-Student) com $n-1$ graus de liberdade (df).

Neste caso, um intervalo de confiança exato para μ com porcentagem 100(1- α), para $n \ge 2$, é dado por

$$\overline{X}(n) \pm t_{n-1,1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}$$

O valor de $t_{n-1,1-\alpha/2}$ é obtido da tabela da distribuição t.

Pode-se observar, pela forma das curvas nos gráficos que $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$.



Na prática os X_i's raramente são normais e o intervalo de confiança dado pela fórmula

$$\bar{X}(n) \pm t_{n-1,1-\alpha/2}\sqrt{S^2(n)/n}$$

é aproximado.

Pelo fato que $t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$, o intervalo obtido é mais largo que o obtido com a fórmula

$$\bar{X}(n) \pm z_{1-\alpha/2}\sqrt{S^2(n)/n}$$

e, portanto, está mais próximo de cobrir o nível $(1-\alpha)$ desejado.

Deve ser observado que $t_{n-1,1-\alpha/2} \rightarrow z_{1-\alpha/2}$ quando $n \rightarrow \infty$.

Exemplo 17:

Supondo que 10 observações 1.20, 1.50, 1.68, 1.89, 0.95, 1.49, 1.58, 1.55, 0.50, e 1,09 foram feitas e apresentaram distribuição normal com média μ desconhecida e queremos construir um intervalo de confiança com 90% (α=0.10) para μ. Dos dados calculamos

$$\bar{X}(10) = 1{,}34$$

$$S^2(10) = 0{,}17$$

Com estes resultados e consultando a tabela da distribuição t calculamos

$$\bar{X}(10) \pm t_{9,0.95}\sqrt{S^2(10)/10} = 1.34 \pm 1.83\sqrt{0.17/10} = 1.34 \pm 0.24$$

Com este resultado podemos dizer com 90 % de confiança que μ está no intervalo [1.10,1.58].

A cobertura do intervalo de confiança pode ser afetada pelas distribuições dos X_i's como mostra o experimento a seguir.

Exemplo 18:

Foram realizados 500 experimentos independentes para cada tamanho de amostra n=5, 10, 20 e 40, com distribuições normal, exponencial, chi-quadrado com 1df (normal ao quadrado), lognormal (e^Y onde Y é normal) e hiperexponencial ($F(x)=0.9F_1(x)+0.1F_2(x)$ sendo F_1 e F_2 exponenciais com médias 0.5 e 0.55).

As estimativas de cobertura dos intervalos de confiança de 90% são mostrados na tabela a seguir.

Distribuição	Assimetria υ	N=5	N=10	N=20	N=40
Normal	0.0	0.910	0.902	0.898	0.900
Exponencial	2.0	0.854	0.878	0.870	0.890
Chi-quadrado	2.83	0.810	0.830	0.848	0.890
Lognormal	6.18	0.758	0.768	0.842	0.852
Hiperexponencial	6.43	0.584	0.586	0.682	0.774

A assimetria υ é definida como:

$$\upsilon = \frac{E[(X-\mu)^3]}{(\sigma^2)^{3/2}} \qquad \text{para } -\infty < \upsilon < \infty$$

Pela tabela pode-se observar que a cobertura atinge valores próximos de 90% quando n se torna maior. Além disso, esta aproximação demora mais a ocorrer nos casos de distribuições mais assimétricas.

5.3 Teste de Hipóteses

Dado o conjunto $X_1, X_2, \ldots, X_n$ de variáveis aleatórias normalmente distribuídas, queremos testar a hipótese nula H_0, de que $\mu = \mu_0$, onde μ_0 é um hipotético valor de μ.

Intuitivamente espera-se que, se $|\overline{X}(n) - \mu_0|$ for um valor grande então é provável que H_0 seja falsa.

Entretanto, para desenvolver um teste com propriedades estatísticas precisamos de uma estatística dos X_i's cuja distribuição seja conhecida e onde H_o seja verdade.

Do que foi discutido, se H_0 é verdadeira a estatística

$$t_n = [\overline{X}(n) - \mu_0] / \sqrt{S^2(n)/n}$$

terá distribuição t com $n-1$ graus de liberdade.

Desta forma, indo de encontro à intuição, o teste de hipótese para $\mu = \mu_0$ é

$$\text{Se} \quad |t_n| \begin{cases} > t_{n-1,1-\alpha/2} & \text{rejeitar } H_0 \\ \le t_{n-1,1-\alpha/2} & \text{aceitar } H_0 \end{cases}$$

O conjunto de todos os x tais que $|t_n| > t_{n-1,1-\alpha/2}$, isto é, correspondem a rejeitar H_0, é chamado **região crítica** para o teste.
A probabilidade α que a estatística t_n caia na região crítica, considerando que H_0 é verdadeira, é chamada **nível do teste**.

Quando realizamos um teste podem ocorrer dois tipos de erros :

a) **Erro tipo I**: Se rejeitamos a hipótese H_0 quando H_0 é verdadeira. A probabilidade deste tipo de erro é α e está sob controle.
b) **Erro tipo II**: Se aceitamos a hipótese H_0 quando H_0 é falsa. Para um nível α e uma amostra de tamanho n, a probabilidade deste erro, que indicaremos por β, depende do μ que é verdadeiro e pode ser desconhecido.

Chamamos $\delta = 1- \beta$ o **poder do teste** sendo a probabilidade de rejeitar H_0 quando H_0 é falso.

Exemplo 19:

Para os dados do exemplo 18, suponhamos que queremos testar a hipótese nula H_0 tal que $\mu= 1$ no nível $\alpha=0.10$.

Como $t_{10} = \dfrac{\overline{X}(10)-1}{\sqrt{S^2(10)/10}} = \dfrac{0.34}{\sqrt{0.17/10}} t_{10} = 2.65 > 1.83 = t_{9,0.95}$

então H_0 será rejeitada.

Existe uma relação próxima entre o intervalo de confiança definido por

$$\overline{X}(n) \pm t_{n-1,1-\alpha/2}\sqrt{S^2(n)/n}$$

e o teste de hipóteses definido por

$$|t_n| \begin{cases} > t_{n-1,1-\alpha/2} & \text{rejeitar } H_0 \\ \leq t_{n-1,1-\alpha/2} & \text{aceitar } H_0 \end{cases}$$

A rejeição da hipótese nula H_0 de que $\mu= \mu_0$, é equivalente a μ_0 não estar contido no intervalo de confiança para μ, assumindo o mesmo valor de α tanto para o teste de hipóteses quanto para o intervalo de confiança.

Lei dos Grandes Números: Teorema

Sejam $X_1, X_2, ..., X_n$ variáveis aleatórias IID com média finita μ. Então $\overline{X}(n) \rightarrow \mu$ com probabilidade 1, quando $n \rightarrow \infty$.

6 Bibliografia

[1] Magalhães, M. N., Lima, A. C. P., “Noções de Probabilidade e Estatística”, 3 ed,. IME-USP, São Paulo, 2001, 375p.
[2] Law, A. M., Kelton, W. D., "Simulation Modeling and Analysis", 3rd ed., McGraw-Hill Companies Inc, 2000,ISBN 0-07-059292-6, 760p.
[3] Cassandras, C. G., ”Discrete Event Systems: Modeling and Performance Analysis”, Aksen Associates Incorporated Publishers, 1993 , ISBN: 0-256-11212-6, 790p.