Variável Aleatória

Variável aleatória é função!!!

Espaço amostral

Informações que coletamos na natureza.)?

Ideia: Conectar probabilidade e dados.

$X: \Omega \xrightarrow{\text{Conexão}} \mathbb{R}$

Nos reais!

+?: Vida real! Prática! cotidiano!

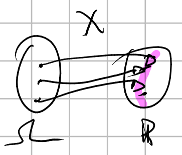
X: Fala sobre toda população através de probabilidade!

Conseguimos calcular probabilidades. (Aula Passada)

Matematicamente: $B \subset \mathbb{R}$, $\underbrace{[X \in B]}_{\text{Evento}} = \{\omega \in \Omega \mid X(\omega) \in B\} \subseteq \Omega$ | $P([X \in B]) = P\{\omega \in \Omega \mid X(\omega) \in B\}$

(letra grega Chi)

Classificação: $\underbrace{X(\Omega)}_{\text{Imagem de } X} = \mathcal{X}$ é chamado de suporte de X.



- $\mathcal{X} \subseteq \mathbb{Z} \Rightarrow$ X é uma variável aleatória (quantitativa) discreta; → Sobre toda população!
- $\mathcal{X}$ é um intervalo (R) $\Rightarrow$ X é uma variável aleatória (quantitativa) contínua.

Na prática: Temos vários modelos (Distribuição) de Variável Aleatória. Observamos e analisamos a variável quantitativa usando estatística descritiva, e escolhemos um modelo adequado para o nosso estudo ou problema.

Variável Aleatória Discreta.

Relembre a tabela de Distribuição de Frequências:

Tabela de distribuição de frequência para uma variável quantitativa discreta

(na amostra)

X	frequência	frequência relativa	porcentagem
x_1	n_1	$f_1 = n_1/n$	$100 \cdot f_1\%$
x_2	n_2	$f_2 = n_2/n$	$100 \cdot f_2\%$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_k	$f_k = n_k/n$	$100 \cdot f_k\%$
Total	n	1	100

Variável aleatória Discreta

na População!

X	frequência relativa
x_1	$f(x_1)$
x_2	$f(x_2)$
$\vdots$	
x_k	$f(x_k)$

$f(\cdot)$ — Função de Probabilidade

$f(x) = P([X = x]) = P(\{\omega \in \Omega \mid X(\omega) = x\})$

Seja $f(x) = P([X = x]) = P\{\omega \in \Omega \mid X(\omega) = x\}$.

Interpretação: $f(x)$ é a frequência relativa na população.

$f(x)$ é chamada de função de probabilidade.

$x \notin \mathcal{X}$

$\{\omega \in \Omega \mid X(\omega) = x\} = \emptyset$

$P(\emptyset) = 0$

$\mathcal{X} = \{x_1, \dots, x_k\}$

X	Função de Probabilidade
x_1	$f(x_1)$
x_2	$f(x_2)$
$\vdots$	$\vdots$
x_k	$f(x_k)$

Por conveniência:

$$f(x) = \begin{cases} P([X = x]), & \text{se } x \in \mathcal{X} \\ 0, & \text{se } x \notin \mathcal{X} \end{cases}$$

(Por conveniência)

Observação: Note que $\mathcal{X} = \{x \in \mathbb{R} \mid f(x) > 0\}$.

É suficiente dar a função de probabilidade! E a gente esquece quem é Ω.

Função de Distribuição Acumulada.

X variável aleatória discreta com suporte $\mathcal{X}$ e função de probabilidade $f(\cdot)$.

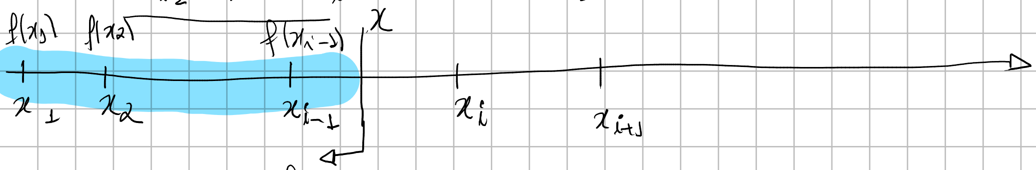
Especificação de X:

Jeito 1) Fç de prob.

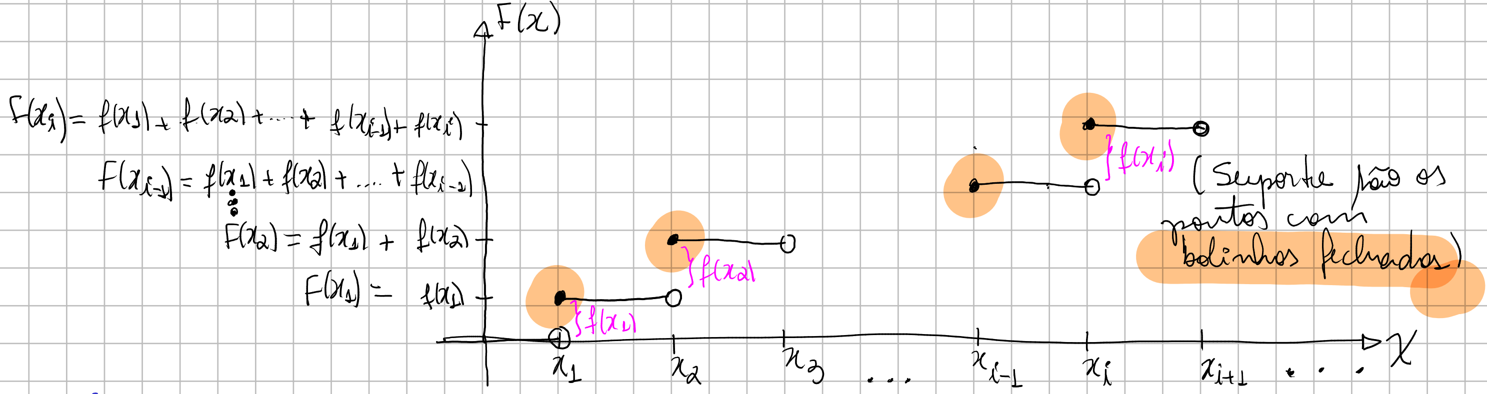
Jeito 2) FDA

- $F: \mathbb{R} \to \mathbb{R}$
- $\mathcal{X} = \{x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots\}$
- Se $x_{i-1} \leq x < x_i$

$\Rightarrow F(x) = f(x_1) + f(x_2) + \dots + f(x_{i-1})$



FDA ⇒ É acumulação da função de probabilidade!



Note que:

$F(x_1) = f(x_1)$

$F(x_2) = \overbrace{f(x_1)}^{F(x_1)} + f(x_2)$

$F(x_3) = \overbrace{f(x_1) + f(x_2)}^{F(x_2)} + f(x_3)$

$\vdots$

$F(x_k)\ f(x_1) + f(x_2) + \dots + f(x_k)$

$f(x_1) = F(x_1)$

$f(x_2) = F(x_2) - F(x_1)$

$f(x_3) = F(x_3) - F(x_2)$

$\vdots$

$f(x_k) - F(x_k)\ \ F(x_{k-1})$

$P([X \in B]) = f(a) + f(b) + f(c)$

$B = \{a, b, c\}$

Medidas de Resumo / $\bar{x} = x_1 \cdot f_1 + \dots + x_k f_k$ / $s^2 = (x_1 - \bar{x})^2 f_1 + \dots + (x_k - \bar{x})^2 f_k$

Média: $\mu = E(X) = x_1 \cdot f(x_1) + \dots + x_k\, f(x_k) + \dots +$

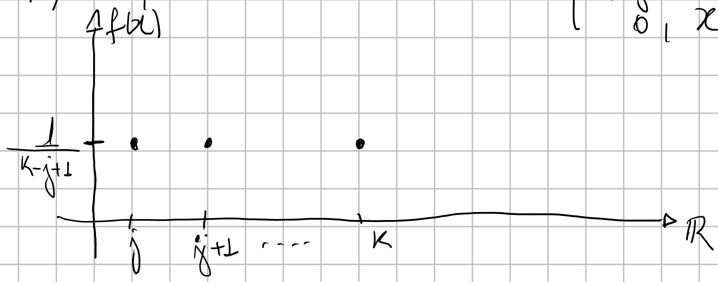
Variância: $\sigma^2 = (x_1 - \mu)^2 f(x_1) + \dots + (x_k - \mu)^2 f(x_k) + \dots$ / Desvio Padrão: $\sigma = \sqrt{\sigma^2}$

Quantil: $0 \leq p \leq 1$. $Q(p)$ tal que $\begin{cases} P(X \geq Q(p)) \geq 0{,}5 \\ P(X \leq Q(p)) \geq 0{,}5 \end{cases}$ → Trocamos "exatamente" por "pelo menos" 50%

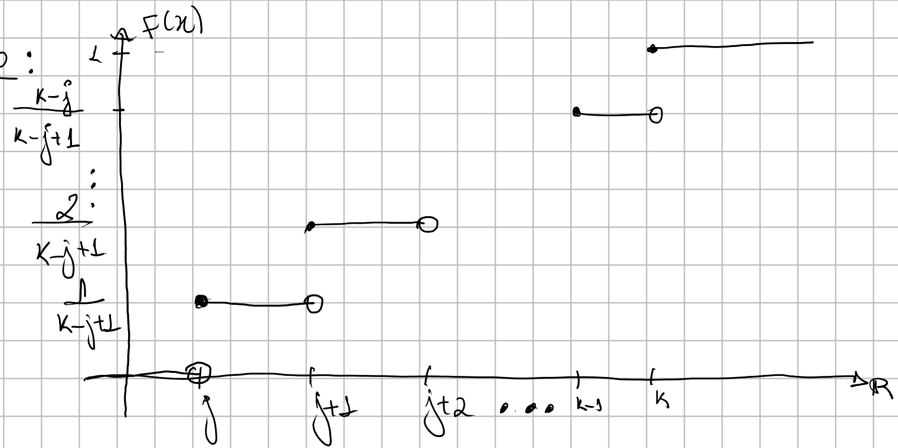
Distribuição Uniforme: $U_D[j, k]$. Notação: $X \sim U_D[j, k]$

Quando usar: Dados são valores inteiros entre j and k, e cada valor é igualmente provável de ser coletado ou observado.

Suporte: $\mathcal{X} = \{j, j+1, \dots, k\}$; Função de Probabilidade: $f(x) = \begin{cases} \frac{1}{(k-j+1)}, & x \in \mathcal{X} \\ 0, & x \notin \mathcal{X} \end{cases}$



Função de Distribuição Acumulada:



Exemplo: Uma rifa tem 100 bilhetes numerados de 1 a 100. Tenho 5 bilhetes consecutivos numerados de 21 a 25 e meu colega tem outros 5 bilhetes com os números 1, 11, 29, 68 e 93. Quem tem mais chance de ganhar?

Solução: $\mathcal{X}=\{1,\dots,100\}$; $f(x)=\begin{cases}\frac{1}{100}, & x\in\mathcal{X}\\ 0, & x\notin\mathcal{X}\end{cases}$.

- Eu: $B=\{21,22,23,24,25\}$. $P[X\in B]=f(21)+f(22)+f(23)+f(24)+f(25)=\frac{5}{100}=0{,}05$
- Meu amigo: $B=\{1,11,29,68,93\}$. $P[X\in B]=f(1)+f(11)+f(29)+f(68)+f(93)=\frac{5}{100}=0{,}05$

Média: $\mu=1.f(1)+2.f(2)+3.f(3)+\dots+100.f(100)$
$$=\frac{(1+100).\cancel{100}}{2}.\frac{1}{\cancel{100}}=50{,}5$$

Variância: $\sigma^2=(50{,}5-1)^2f(1)+(50{,}5-2)^2f(2)+\dots+(50{,}5-100)^2f(100)$
$$=833{,}25$$

Desvio Padrão: $\sigma=\sqrt{833{,}25}\cong 28{,}87$

Mediana: Quantil de ordem 50%: $Q(0{,}5)=50$

$$P(X\leq 50)=f(1)+\dots+f(50)=\frac{50}{100}=0{,}5$$

$$P(X\geq 50)=f(50)+f(51)+\dots+f(100)=\frac{51}{100}=0{,}51.$$

Distribuição Bernoulli: $X\sim \text{Bernoulli}(p)$.

Quando usar: O experimento/estudo tem apenas 2 resultados possíveis. (*)

Um resultado chamamos de sucesso, e o outro chamamos de fracasso.

Sucesso $\to 1$	Fracasso $\to 0$
Ganhar Copa	Perder copa
Sim	Não
Infectado	Não Infectado
Brasileiro	Estrangeiro

(*) Experimento Bernoulli.

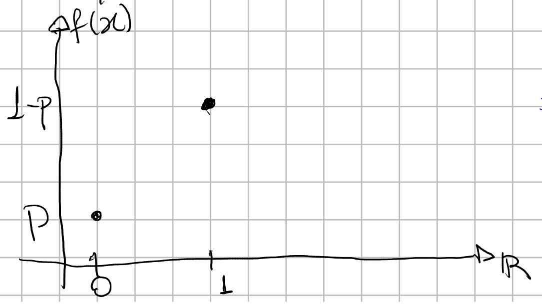
$P(\text{Sucesso})=p$ $[X=1]$

$P(\text{Fracasso})=1-p$ $[X=0]$

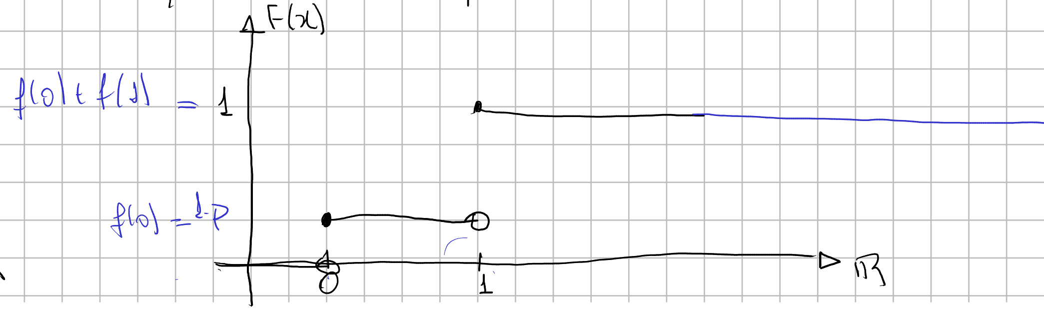
Sucesso codificamos com 1 e o fracasso com 0.

Suporte: $\mathcal{X}=\{0,1\}$; **Função de Probabilidade:** $f(x)=\begin{cases}p, & x=1\\ 1-p, & x=0\\ 0, & x\notin\mathcal{X}\end{cases}$

Função de Probabilidade



Função de Distribuição Acumulada



Exemplo:

Assuma que a prevalência de infecção pelo vírus HIV em país da África Subsariana seja 30%. Considere o fenômeno aleatório que consiste de prever se um novo paciente está infectado. Qual o modelo probabilístico adequado neste contexto? Qual a função de probabilidade? Qual a função de distribuição acumulada?

Solution: Sucesso: Pessoa infectada com o vírus HIV

$\mathcal{X}=\{0,1\}$; $f(x)=\begin{cases}0{,}3, & x=1\\ 0{,}7, & x=0\\ 0, & x\notin\mathcal{X}\end{cases}$; $F(x)=\begin{cases}0, & x<0\\ 0{,}7, & 0\leq x<1\\ 1, & x\geq 1\end{cases}$

$\mu = x_1 f(x_1) + x_2 f(x_2) + \dots + x_k f(x_k)$

Média: $\mu = 1\cdot 0{,}3 + 0\cdot 0{,}7 = 0{,}3$ | $\mu = 0\cdot f(0) + 1\cdot f(1) = p$

$f(1)\cdot(1-\mu)^2 + f(0)\cdot(0-\mu)^2$

Variância: $\sigma^2 = (1-p)^2\cdot p + (0-p)^2\cdot(1-p) = (1-p)\cdot p\,(1-p+p) = p(1-p)$

$\sigma^2 = 0{,}3\cdot 0{,}7 = 0{,}21$

$Q(0{,}5)$ ① $= 0{,}3+0{,}7$

x	~~0~~	1
$f(x)$	0,3	0,7

Desvio Padrão: $\sigma = \sqrt{0{,}21} = 0{,}46$

Ex: $Q(0{,}5)$ $\neq Q(0{,}5)$

x	2	7	9	11
$f(x)$	0,2	0,2	0,2	0,4

~~0,2+0,2~~ 0,4 ~~X~~ 0,2+0,2+0,2 0,6 Bateu p! Achei $Q(p)$!

Mediana: $Q(0{,}5) = 1$, $P(X\geq 1) = 0{,}7 = f(1) \geq 0{,}5$

$P(X\leq 1) = 0{,}3+0{,}7 = 1 \geq 0{,}5$

Distribuição Binomial. $X \sim b(n,p)$.

Quando usar: Temos n experimentos Bernoulli, e queremos contar o nº de sucessos!

X: nº de sucessos / em n casos / em n tentativas,

a) nº casos

b) cada caso semelhante

c) cada caso pode ser sucesso ou fracasso

$\mathcal{X}=\{0,1,2,\dots,n\}$; $f(x)=\begin{cases}\binom{n}{x}p^x(1-p)^{n-x}, & x\in\mathcal{X},\\ 0, & x\notin\mathcal{X}.\end{cases}$

Em que

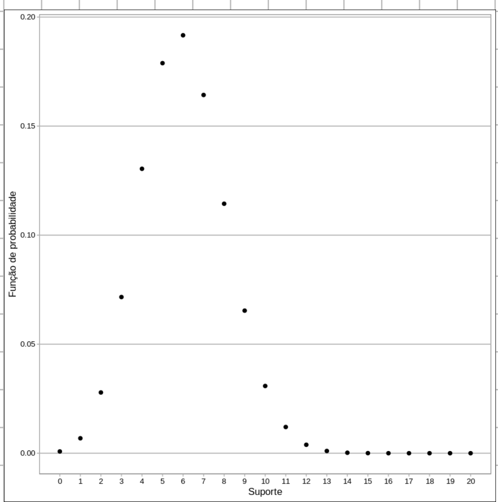
- $\binom{n}{x} = \dfrac{n!}{x!\,(n-x)!}$ Coeficiente binomial ou binômio de Newton

$4! = 4\cdot 3\cdot 2\cdot 1 = 24$

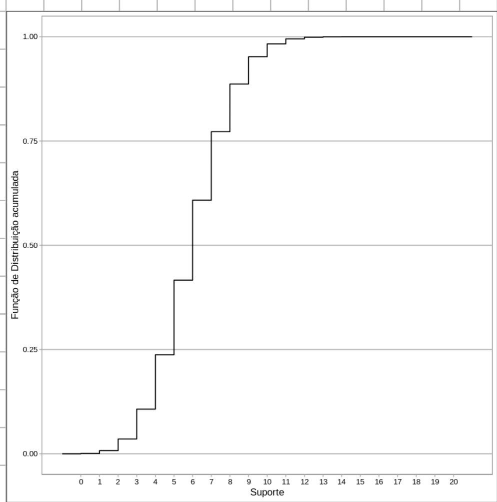
$5! = 5\cdot 4\cdot 3\cdot 2\cdot 1 = 120$

- $\begin{cases} k! = k\cdot(k-1)\cdot\ldots\cdot 2\cdot 1\\ 0! = 1\\ 1! = 1\end{cases}$ Por conveniência

Função de Probabilidade



Função de Distribuição Acumulada



Média: $\mu = 0\cdot f(0) + 1\cdot f(1) + \dots + n\cdot f(n) = n\cdot p$

Variância: $\sigma^2 = n\cdot p\cdot(1-p)$

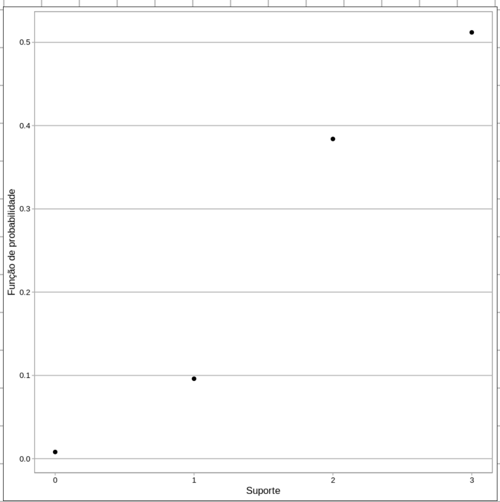
Mediana: $\lfloor n\cdot p \rfloor \leq Q(0,5) \leq \lceil n\cdot p \rceil$

Exemplo: Sabe-se que a eficiência de uma vacina é de 80%. Um grupo de três indivíduos é sorteado dentre a população vacinada e submetido a testes para averiguar se a imunização foi efetivada. Qual o modelo adequado para este fenômeno aleatório? Encontre a função de probabilidade e a fução de distribuição acumulada.

Solução: Sucesso: Pessoa está imunizada; $p = 0,8$; $n = 3$

$\mathcal{X} = \{0, 1, 2, 3\}$;

$f(0) = \binom{3}{0} 0,8^0 \cdot 0,2^3 = 0,008$; $f(1) = \binom{3}{1} 0,8^1 \cdot 0,2^2 = 0,096$; $f(2) = \binom{3}{2} 0,8^2 \cdot 0,2^1 = 0,384$; $f(3) = \binom{3}{3} 0,8^3 \cdot 0,2^0 = 0,512$



Média: $\mu = 0\cdot f(0) + 1\cdot f(1) + 2\cdot f(2) + 3\cdot f(3) = 3\cdot 0,8 = 2,4$

Variância: $\sigma^2 = 3\cdot 0,8 \cdot 0,2 = 0,48$; $\sigma = \sqrt{0,48} = 0,692820323$

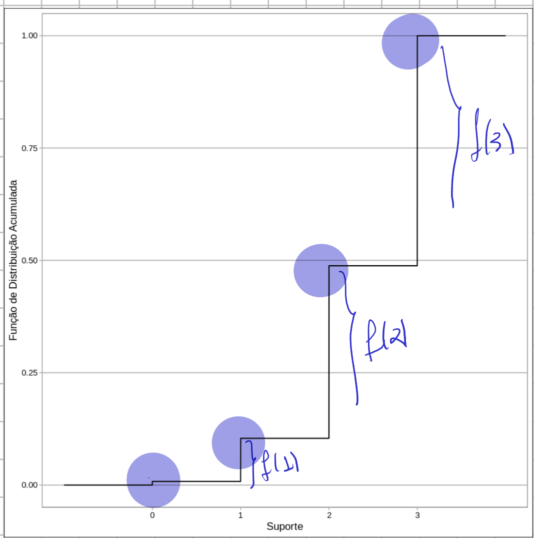
Mediana: $2 \leq Q(0,5) \leq 3$

$P(X \leq 2) = f(0) + f(1) + f(2) = 0,488$

$\Rightarrow \begin{cases} P(X \leq 3) = 1 \geq 0,5 = f(0) + f(1) + f(2) + f(3) \geq 0,5 \\ P(X \geq 3) = f(3) = 0,512 \geq 0,5 \end{cases}$

$Q(0,5) = 3$

Função de Distribuição Acumulada



Distribuição Poisson: $X \sim \text{Poisson}(\lambda)$.

Quando usar: Temos contagem de ocorrências num intervalo delimitado de tempo.

Ocorrências:
- Chamada no call center em um dia
- Pessoa entrar numa loja em uma hora
- Partícula emitida em um minuto

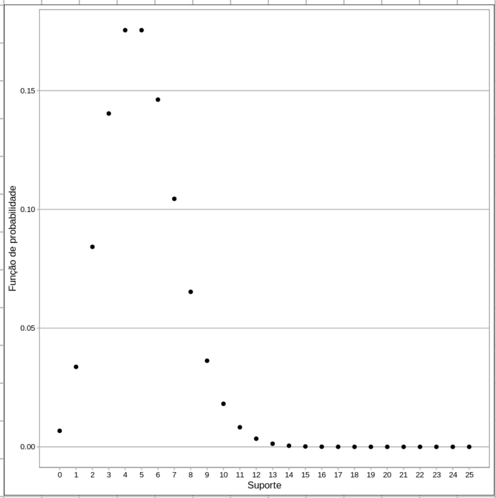
λ: Média na população de ocorrências no intervalo de tempo.

Contagem!

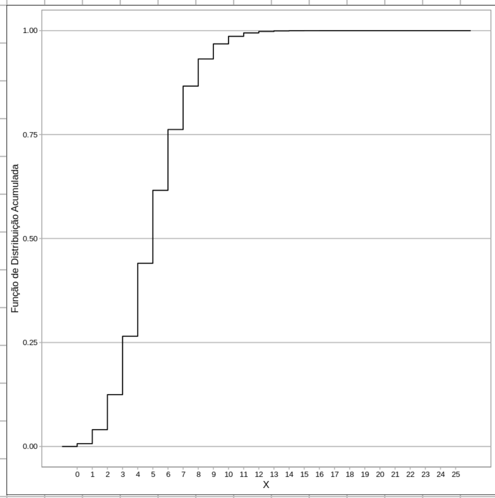
Suporte: $\mathcal{X} = \mathbb{Z}^+$; $f(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^x}{x!}, & x \in \mathcal{X} \\ 0, & x \notin \mathcal{X} \end{cases}$

$\mathcal{X} = \{0, 1, 2, 3, 4, \dots\}$

Função de Probabilidade



Função de Distribuição Acumulada



Média: $\mu = \lambda$; Variância: $\sigma^2 = \lambda$

Exemplo: Suponha que uma unidade básica de saúde de um bairro de classe média realiza em média 10 atendimentos em dias de segunda-feira. Qual a probabilidade desta UBS atender, na próxima segunda-feira, no máximo 5 cidadãos?

Solução: $\lambda = 10$; $\mathcal{X} = \mathbb{Z}^+$; $f(x) = \begin{cases} \dfrac{e^{-10}\, 10^x}{x!}, & x \in \mathbb{Z}^+, \\ 0, & x \notin \mathbb{Z}^+. \end{cases}$

$\mathcal{X} = \{0, 1, 2, 3, 4, 5, 6, \dots\}$

$P(X \leq 5) = f(0) + f(1) + f(2) + f(3) + f(4) + f(5) \cong 0{,}07$

Média: $\mu = 10$; Variância: $\sigma^2 = 10$; Mediana: $Q(0{,}5) = 10$