Inferência Bayesiana

Henrique Nunes de Oliveira

Depto. Zootecnia

UNESP - Jaboticabal

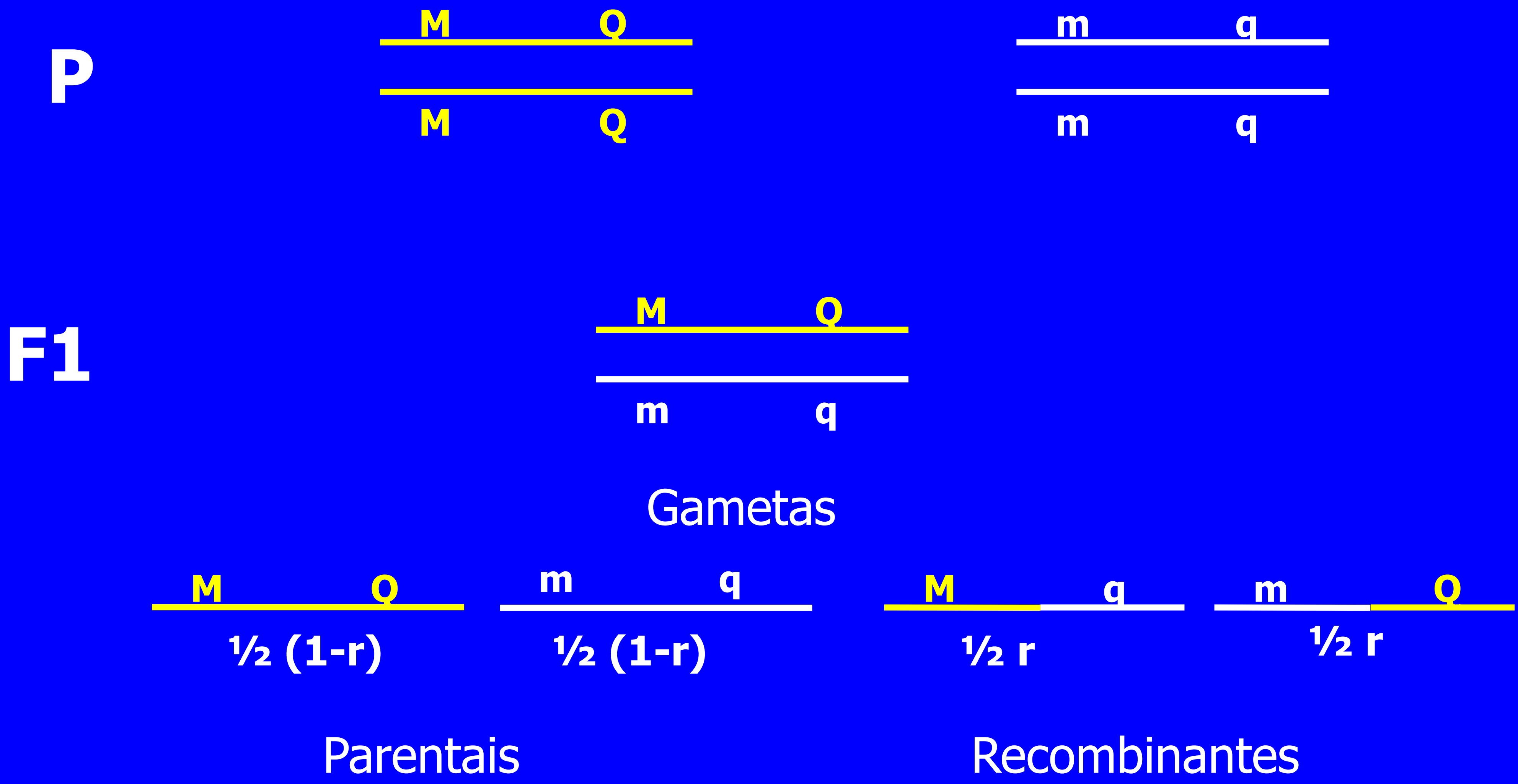
Estatística Fundamental

- Probabilidade e Teorema de Bayes
- Variáveis Aleatórias
- Distribuições de Probabilidade
- Distribuições Conjuntas
- Amostras Aleatórias
- Distribuição condicional e marginal

Probabilidades

- Espaço amostral = conjunto de resultados de um experimento.
- Evento = Possível resultado de um experimento (subconjunto do espaço amostral).
- P(A) = Probabilidade de ocorrência de um evento A
- Probabilidades podem ser conhecidas a priori ou calculadas com base em freqüências observadas

Exemplo: Marcadores Moleculares



Exemplo: Marcadores Moleculares

	Pais	MMQQ	x	mmqq	
	F1	MmQq			
	Gametas F1	**MQ**	**Mq**	**mQ**	**mq**
Não Ligados (inútil)	Freq(%)	**25**	**25**	**25**	**25**
Muito Ligados	Freq(%)	**48,6**	**1,4**	**1,4**	**48,6**

- Espaço amostral marcador: {M,m}
- Espaço amostral QTL: {Q,q}
- Espaço amostral conjunto: {MQ,Mq,mQ,mq}
- Eventos : A={M}; B={Q}; H={MQ}; ...
- P(A) = P(B) = ½ (Conhecidas a priori)
- P(H) = ? (Desconhecida: depende de r)

Distribuição Hipotética dos Gametas

Marcador **QTL**	**M**	**m**	TOTAL
Q	48,6%	1,4%	50%
q	1,4%	48,6%	50%
TOTAL	50%	50%	100

Eventos nos Gametas

- Evento A = Marcador (M)
- Evento C = QTL (Q)
- Evento B = Marcador (m)
- Evento D = QTL (q)

Eventos	A	B	TOTAL
C	48,6%	1,4%	50%
D	1,4%	48,6%	50%
TOTAL	50%	50%	100

Eventos Mutuamente Exclusivos

- A ocorrência de um implica na não ocorrência do outro:
- A e B são mutuamente exclusivos.
- A e C são **Não** mutuamente exclusivos.

Probabilidade Marginal

- P(A) ; P(B) ; P(C) ; P(D)

- Probabilidade de ocorrência de um determinado evento, ignorando-se eventos associados.

Probabilidade Marginal

	A	B	TOTAL
C	48,6%	1,4%	50%
D	1,4%	48,6%	50%
TOTAL	50%	50%	100

Probabilidade Conjunta

- Eventos **Não** Mutuamente Exclusivos
- P(A e C); P(A e D); P(B e C); P(B e D)

Probabilidade Conjunta

	A	B	TOTAL
Q	48,6%	1,4%	50%
q	1,4%	48,6%	50%
TOTAL	50%	50%	100

Eventos Independentes

- Eventos independentes
  - A ocorrência de um evento não interfere na probabilidade de ocorrência do outro.
  - **P(F e G) = P(F) * P(G)**

- Eventos Não Independentes
  - **P(A e C) <> P(A)*P(C)**

Probabilidade Condicional

- Probabilidade de ocorrência de um evento, considerando-se que outro evento já tenha ocorrido.

  – $P(C|A)$ probabilidade de ocorrência do evento C, dado que ocorreu o evento A.

Eventos	A	B	TOTAL
C	48,6%	1,4%	50%
D	1,4%	48,6%	50%
TOTAL	50%	50%	100

Probabilidade Condicional

- $P(C|A) = P(AC)/P(A)$
- $P(A)=P(B)=P(C)=P(D) = 50\%$
- $P(AC)=48,6\%$
- $P(C|A)=0,486/0,5 =0,972$
- Caso o gameta contenha o Marcador M, a probabilidade de conter o QTL Q é de 97,2%

Probabilidade Condicional

- $P(C|B) = 0,014/0,50 = 0,028$
- Caso o gameta contenha o marcador m, a probabilidade de conter o QTL Q é 2,8%

Teorema de Bayes

- P(B|A) = P(AB)/P(A)
- P(AB)=P(B|A)P(A)
- P(A|B) = P(AB)/P(B)
- P(A|B)=P(B|A)P(A)/P(B)

$$P(\theta \mid y) = \frac{P(y \mid \theta) * P(\theta)}{P(y)}$$

$$P(\theta \mid y) \propto P(y \mid \theta) * P(\theta)$$

Exemplo

Genótipo	QQ	Qq	qq
Freq(genótipo)	0.5	0.3	0.2
Pr(Altura >70 | genotipo)	0.3	0.6	0.9

Pr(altura > 70) = 0.3*0.5 +0.6*0.3 + 0.9*0.2 = 0.51

$$\Pr(QQ \mid altura > 70) = \frac{\Pr(QQ) * \Pr(altura > 70 \mid QQ)}{\Pr(altura > 70)}$$

= 0.5*0.3 / 0.51 = 0.294

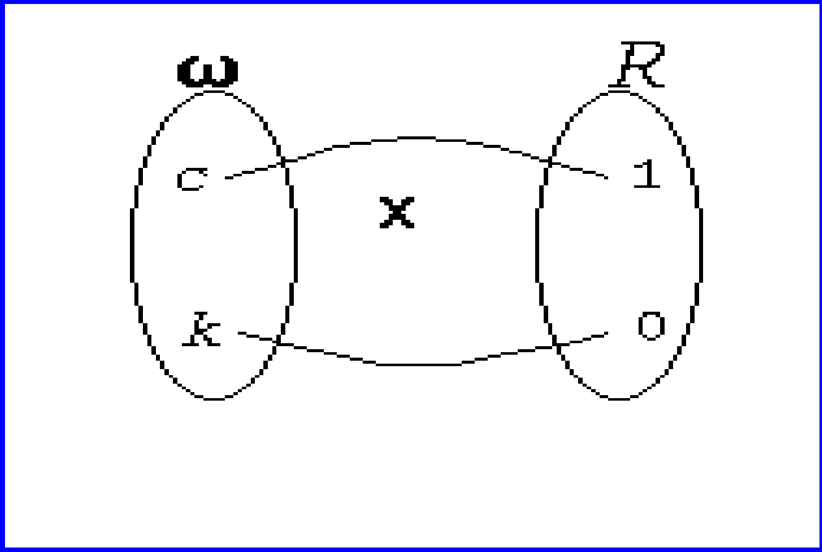
Estatística Fundamental

- Variáveis Aleatórias
- Distribuições de Probabilidade
- Distribuições Conjuntas
- Amostras Aleatórias
- Distribuição condicional e marginal

Variável Aleatória

- É uma função com domínio no espaço amostral de um experimento e contradomínio no conjunto dos Números Reais.

$$X(\omega) = R$$



Variável Aleatória

- Função definida no espaço amostral e definida no conjunto dos Reais

  Dado um experimento aleatório com espaço amostral w, uma variável aleatória (v. a.) é uma função que associa a cada elemento amostral um número real.

Variável Aleatória

- Portanto: sempre, a qualquer variável aleatória vão estar associados, um experimento (espaço amostral); uma função de probabilidade e valores no conjunto dos reais.

Variável Aleatória

- X é discreta se pode assumir um número contável de valores em um intervalo
- X é continua se puder assumir qualquer valor em um intervalo

Distribuições

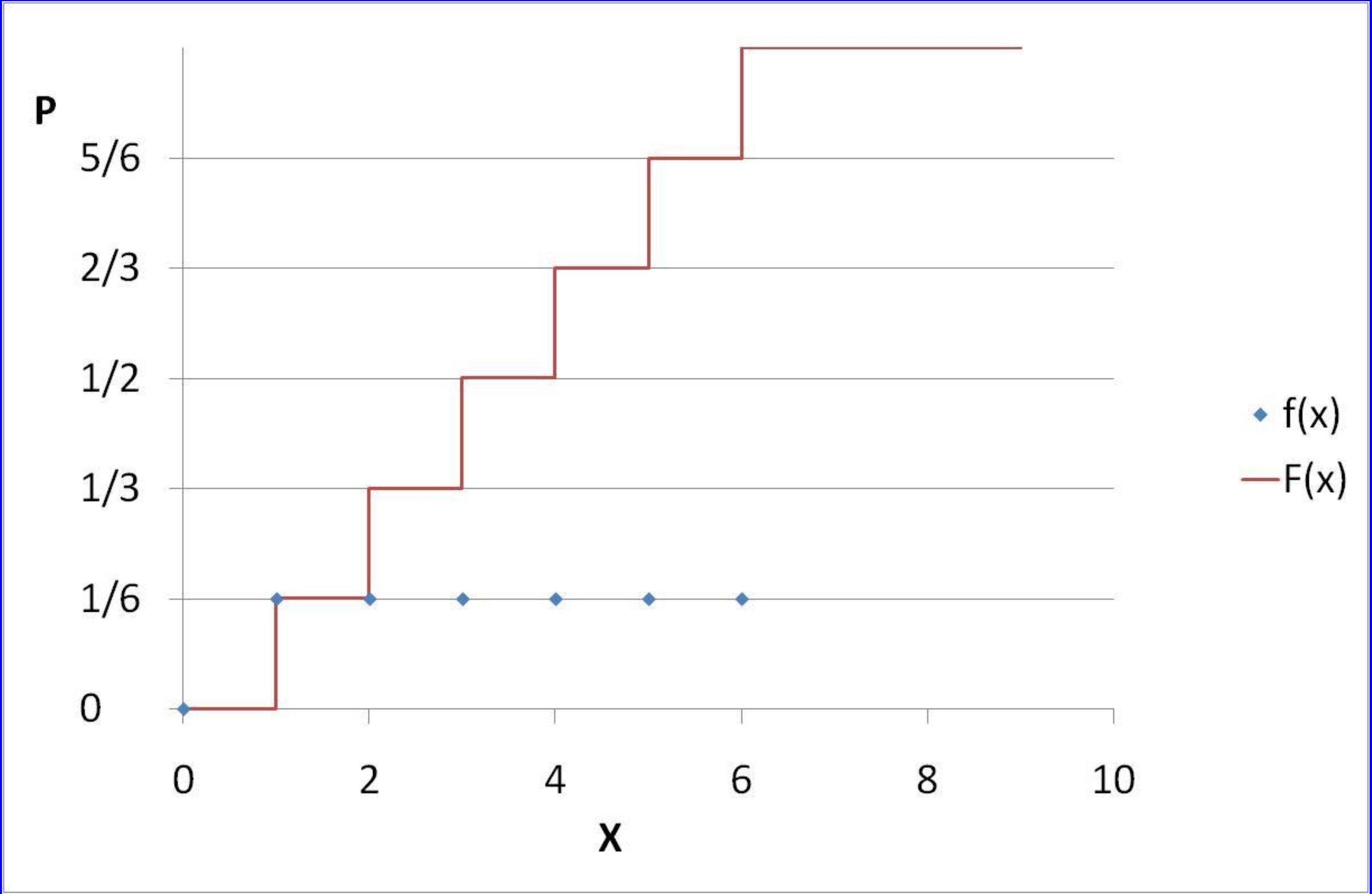
- Funções de distribuição de probabilidade f(x)

Distribuições

- Funções de distribuição acumulada: $F_x(.)$
- Função com domínio em R e contradomíno em [0,1]:

$$F(x) = P(X \leq x)$$

Distribuição Uniforme (n=6)



Distribuições Acumuladas

Propriedades:

$F_x(-\infty) = \lim_{x \to -\infty} = 0$

$F_x(+\infty) = \lim_{x \to +\infty} = 1$

Distribuições Acumuladas

Propriedades:

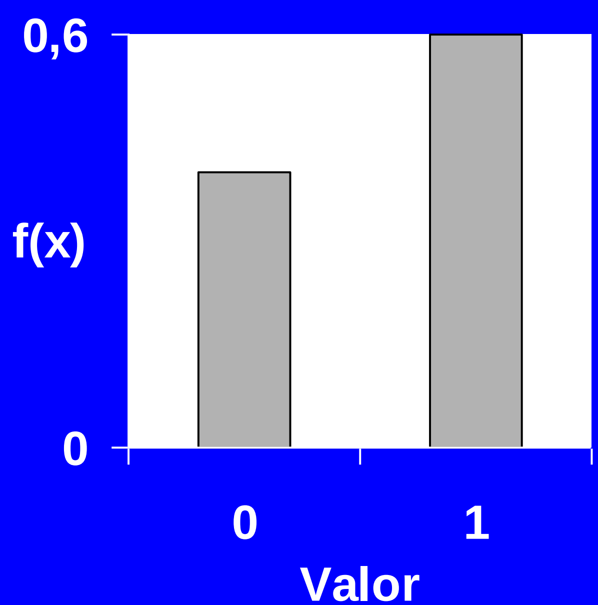
Monótona, não decrescente

$F_X(a) \leq F_X(b) \quad$ se $a < b$

Variáveis aleatórias discretas

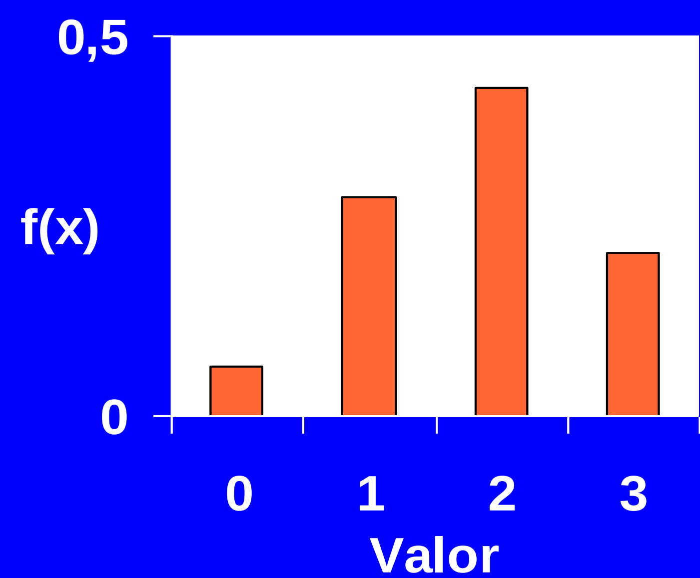
Bernoulli

$f(x)=p^{x}(1-p)^{1-x}I_{\{0,1\}}(x)$



Binomial

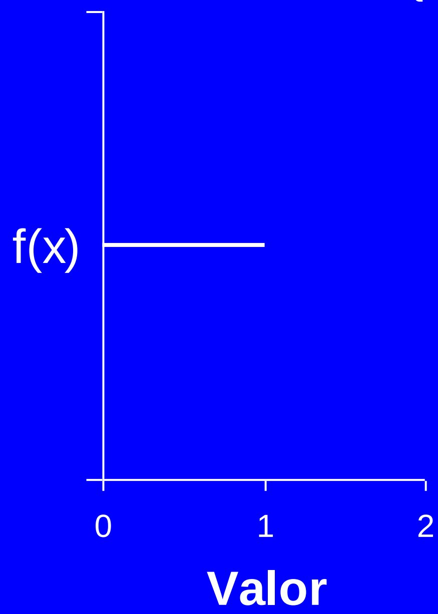
$f(x)=\binom{n}{k}p^{x}(1-p)^{n-x}\,I_{\{0,n\}}(x)$



Variáveis contínuas

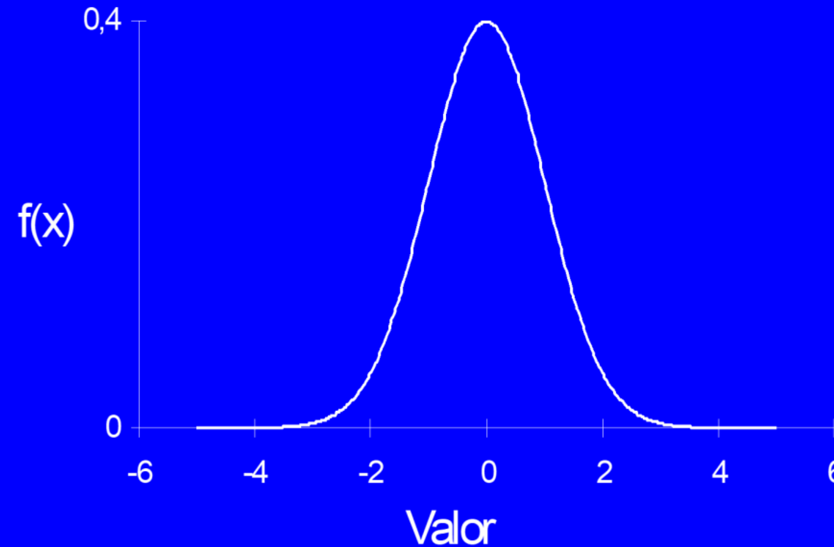
Uniforme Contínua

$f(x)=1/(b-a)I_{\{a,b\}}(x)$



Normal Padrão

$f(x)=(2\pi)^{-1/2}e^{-x/2}$



Distribuição Acumulada

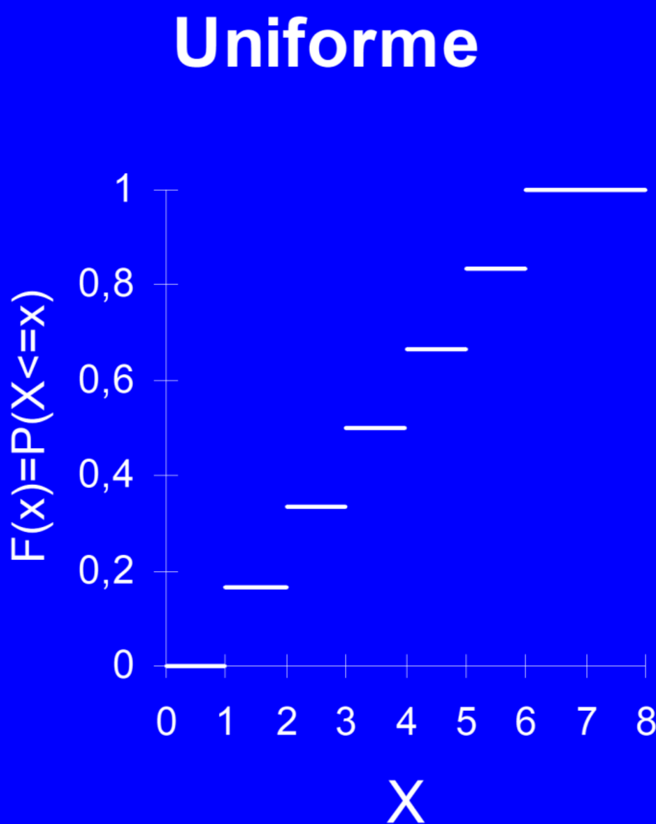
- Variáveis Discretas

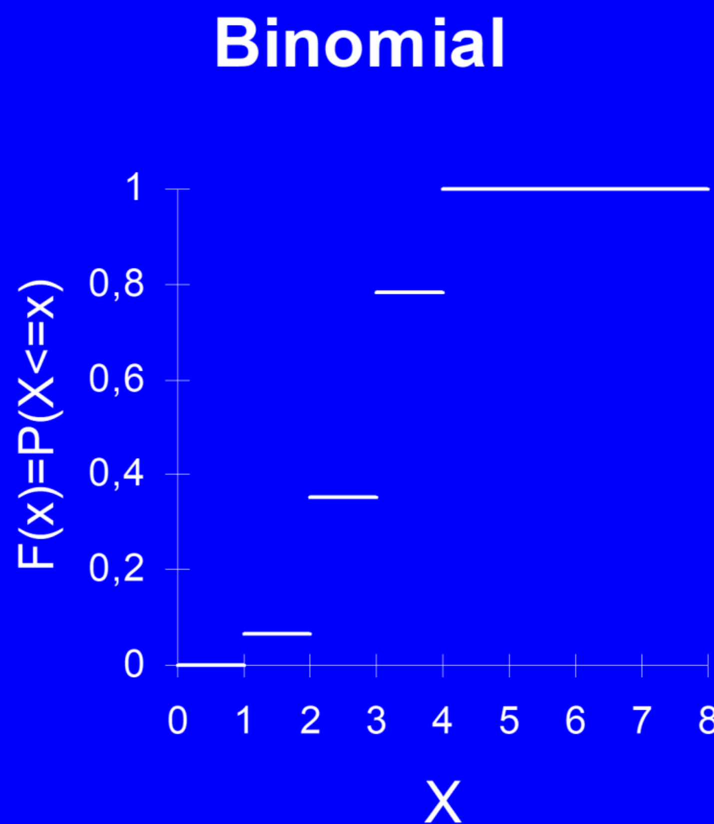
$$F(x) = P(X \leq x) = \sum_{-\infty}^{x} f(x)$$

- Variáveis Contínuas

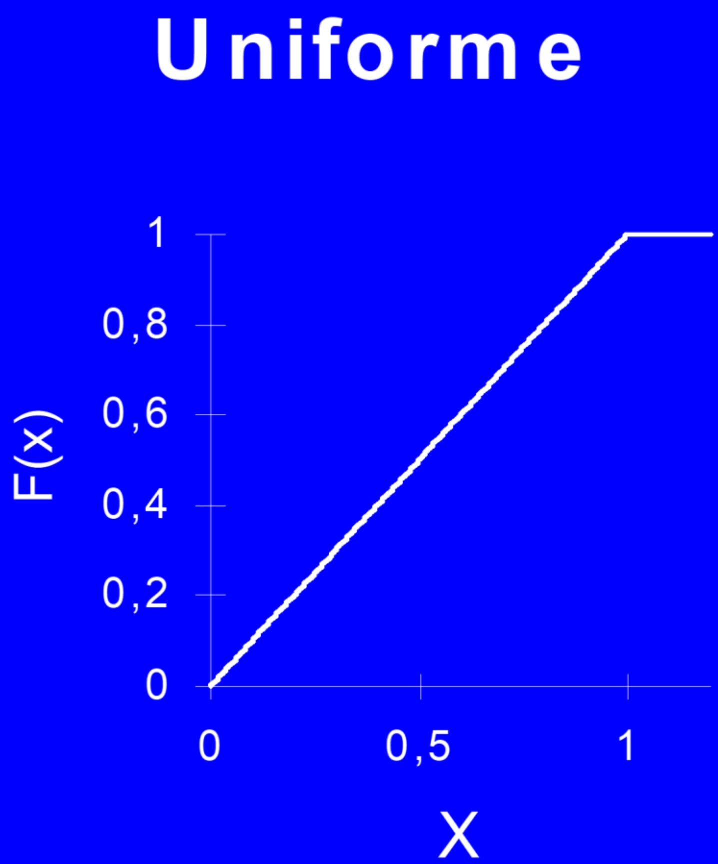
$$F(x) = \int_{-\infty}^{x} f(x)dx \Leftrightarrow f(x) = \frac{\partial F(X)}{\partial x}$$

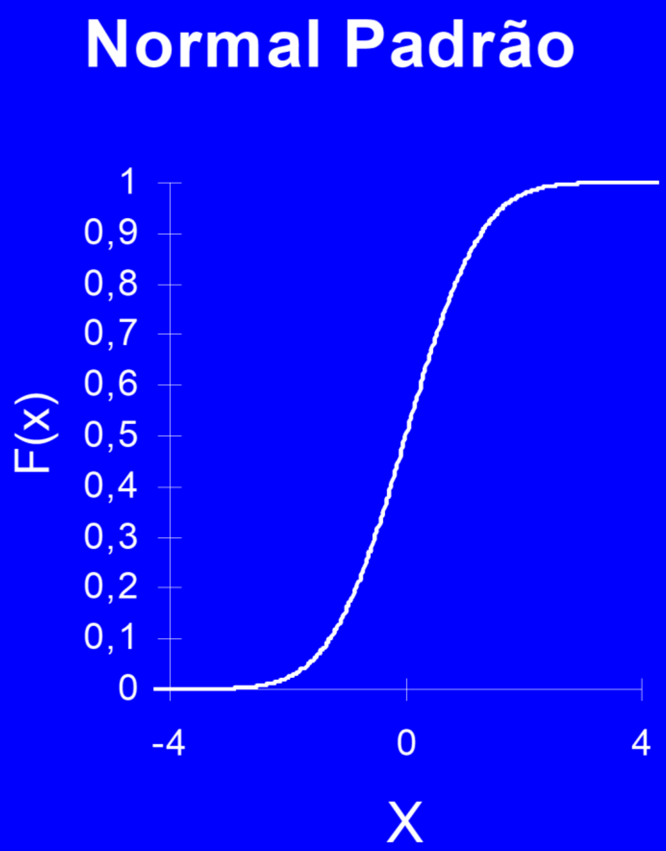
Distribuição Acumulada (discretas)





Distribuição Acumulada (contínuas)





Amostra aleatória

- Sejam $X_1, X_2, X_3, \ldots X_n$ variáveis aleatórias tais que $f(x_1, x_2, x_3, \ldots x_n) = f(x_1)f(x_2)f(x_3)\ldots f(x_n)$ sendo, $f(x_i)$ uma distribuição comum a cada X_i, então $X_1, X_2, X_3, \ldots X_n$ é uma amostra aleatória de tamanho n, de uma população com distribuição $f(x)$.
- Amostra aleatórias são conj. de variáveis aleatórias independentes e identicamente distribuídas (iid~$f(x)$)

Distribuições conjuntas

P(x,y)

X\Y	40	50	60
0	0,300	0,125	0,100
1	0,200	0,125	0,150

Distribuições Marginais

P(x,y)

X\Y	40	50	60	P(x)
0	0,300	0,125	0,100	*0,525*
1	0,200	0,125	0,150	*0,475*
P(y)	*0,50*	*0,25*	*0,25*	*1,00*

Marginais Discretas

$$P(x) = P(X = x) = \sum_{y} P(x, y)$$

$$P(y) = P(Y = y) = \sum_{x} P(x, y)$$

Exemplo

$$P(0) = P(X = 0) = \sum_{y} P(0, y) =$$

$$= P(0{,}40) + P(0{,}50) + P(0{,}60) =$$

$$0{,}300 + 0{,}125 + 0{,}100 = 0{,}425$$

Distribuições condicionais

P(x|y)

X\Y	40	50	60	P(x)
0	0,60	0,50	0,40	*0,525*
1	0,40	0,50	0,60	*0,475*
P(y)	*0,50*	*0,25*	*0,25*	*1,00*

Condicionais Discretas

- Condicional

$$P(x|y) = P(X = x|Y = y) = \frac{P(x,y)}{P(y)}$$

$$Ex: P(0\,|\,40) = P(X = 0\,|\,Y = 40) =$$

$$= \frac{P(0{,}40)}{P(40)} = \frac{0{,}300}{0{,}50} = 0{,}60$$

Distribuições condicionais

P(y|x)

X\Y	40	50	60	P(x)
0	0,57	0,24	0,19	0,525
1	0,42	0,26	0,32	0,475
P(y)	0,50	0,25	0,25	1,00

Marginais e Condicionais Continuas

- Condicional

$$f(x \mid y) = \frac{f(x, y)}{f(y)}$$

- Marginal

$$f(x) = \int_{R_y} f(x, y)dy = \int_{R_y} f(x \mid y) f(y)dy$$

Marginais e Condicionais Discretas

$$\sum_x P(x \mid y) = 1$$

$$\sum_y P(x \mid y) P(y) = P(x)$$

$$\sum_y P(x \mid y) P(y) = \sum_y P(x, y)$$

Teorema de Bayes

$$P(x|y) = \frac{P(x,y)}{P(y)} \Leftrightarrow P(x,y) = P(x|y)P(y)$$

$$P(y|x)P(x) = P(x,y) = P(x|y)P(y)$$

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}$$