5.4 Identificação da Distribuição Teórica de Probabilidades

A terceira etapa da metodologia visa à identificação de uma distribuição teórica de probabilidades que possa representar, da melhor maneira possível, o comportamento estocástico da variável sob análise. Neste caso, a construção de uma distribuição de freqüências e a utilização de histogramas, é fundamental para a identificação ou delineamento da distribuição teórica de probabilidades.

A construção de um histograma permite dar inicio ao processo de inferência sobre uma distribuição teórica de probabilidades. As hipóteses sobre qual distribuição adotar devem estar baseadas no contexto do assunto investigado e no perfil do histograma obtido. Por exemplo, se os dados tratam de tempos entre chegadas e o histograma possui um perfil semelhante ao de uma distribuição exponencial, a hipótese de que os dados são gerados de acordo com aquela distribuição é quase uma garantia, uma vez que a distribuição exponencial costuma estar associada a tempos decorridos entre chegadas de entidades em um sistema. Da mesma forma, se a variável coletada tratar, por exemplo, de pesos de pallets e o histograma apresentar um gráfico com certa simetria em torno da média, assemelhando-se a uma distribuição normal, não se estará longe da verdade ao se adotar esta distribuição.

No tópico que segue apresentam-se as principais distribuições teóricas de probabilidades, considerando suas características gerais, tais como, aplicações mais comuns, parâmetros, etc..

5.4.1 Principais Distribuições Teóricas de Probabilidades

As distribuições: Normal, Exponencial e Poisson são usualmente encontradas e não apresentam dificuldades em serem reconhecidas e analisadas. Apesar de apresentarem maiores dificuldades para a análise, outras distribuições como a Gama e Weibull fornecem uma grande variedade de formas e, portanto, não podem ser negligenciadas no processo de identificação.

No processo de identificação, é comum a ocorrência de um diagnóstico preliminar. Nesta fase, atribui-se a uma determinada distribuição teórica (geralmente as mais conhecidas) a responsabilidade pela geração de dados de um processo. Durante o processo de ajuste das curvas, no entanto, podem-se ter algumas dificuldades em demonstrar a aderência entre os dados empíricos e aqueles da curva teórica. É preciso examinar com a máxima cautela tais diferenças. Se as estas se encontram em um dos extremos da curva, talvez distribuições como a Gama ou Weibull possam aderir mais adequadamente aos dados.

Principais Distribuições Contínuas

Normal

Dentre as muitas distribuições contínuas citadas ao longo deste texto sem dúvida a *distribuição normal* é mais importante delas. Foi primeiramente estudada no início do século XVIII, quando alguns pesquisadores verificaram o incrível grau de regularidade associados com erros de medição. Eles descobriram que os padrões (distribuições) observados eram aproximados por uma distribuição contínua a qual se referiam como *curva normal de erros* e atribuída às leis do acaso [O'Keefe, 1989].

A conhecida curva na forma de sino descreve fenômenos simétricos em torno da média (Figuras 5.2 e 5.3). Possui dois parâmetros: a média μ e o desvio-padrão σ. É utilizada sempre que a aleatoriedade for causada por várias fontes independentes agindo de forma aditiva. Por exemplo, um cabo de aço trançado utilizado num elevador, ou alguma outra aplicação de tração, é formado por muitos fios de aço que adequadamente entrelaçados conferem uma forte resistência ao cabo. A força ou capacidade de resistência de cada um dos fios é uma variável aleatória com distribuição normal. A capacidade total do cabo X será equivalente à soma das capacidades individuais dos fios x_i e terá também distribuição normal. Isto é, X = Normal (μ, σ). Outros exemplos de aplicação, como erros de medição ou ainda a soma ou média de amostras de um grande número de observações independentes de uma distribuição qualquer (base do *Teorema Central do Limite*). Talvez este último exemplo reflita a grande importância desta distribuição na simulação.

O *Teorema Central do Limite* estabelece que a soma ou a média resultante de um grande número de valores aleatórios e independentes é aproximadamente normal, independente da distribuição dos valores individuais. Com base neste teorema, pode-se, por exemplo, agregar os tempos de inúmeros subprocessos independentes, somando-os e substituindo-os por um único valor (soma ou média) cujo resultado tende a normalidade na medida em que cresce o número de subprocessos. Segundo Pegden (1991), a menos que estes possuam distribuições extremamente assimétricas (como a Exponencial, por exemplo), tal aproximação continua válida mesmo para pequenas amostras. Como será visto mais adiante no capítulo seis, este teorema ajudará, também, nos processos de análise dos resultados de simulações, uma vez que a média de uma distribuição amostral tende a normalidade, independentemente da distribuição que rege o comportamento da população da qual foi retirada.

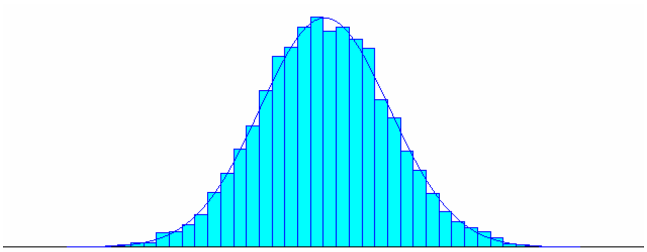

Figura 5.2: Exemplo de uma Normal com $\mu = 4$ e $\sigma = 0{,}5$

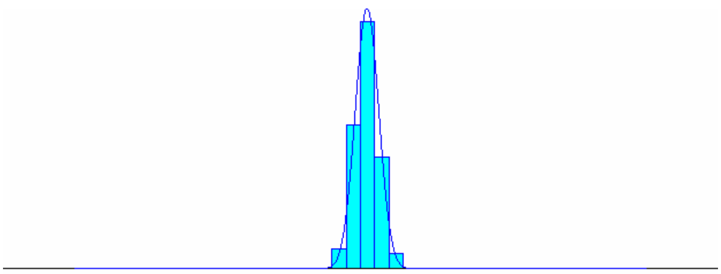

Figura 5.3: Exemplo de uma Normal com $\mu = 4$ e $\sigma = 0{,}01$

O fato desta distribuição admitir valores negativos requer que certos cuidados devam ser tomados no caso de seu emprego em modelos de simulação. Por exemplo, se for associada uma distribuição desta natureza a um tempo de processo, é preciso verificar a possibilidade de ocorrência de valores muito pequenos ou mesmo negativos, dependendo dos parâmetros adotados. No caso de uma variável aleatória normalmente distribuída com média = 3 e desvio padrão = 1, existe uma chance de ≈ 1% de ocorrerem valores muito próximos de zero e/ou negativos. Um descuido em aceitar tais valores num modelo de simulação pode levar à resultados indesejados e a erros na tomada de decisão.

Uniforme

A distribuição Uniforme é talvez àquela com o maior número de adjetivos. Seu emprego costuma estar associado a expressões como: a mais simples ou a que ilustra o maior desconhecimento do fenômeno aleatório sob análise. É tradicionalmente empregada quando a única informação disponível sobre a variável aleatória é que esta ocorre entre dois limites. Possui dois parâmetros: o seu valor mínimo a e seu valor máximo b.

Como exemplo de seu uso, pode-se citar: distância entre fonte e destino de uma entidade através de um sistema (uma mensagem em uma rede de comunicações ou uma matéria-prima em um sistema de manufatura). Sua grande importância está em seu emprego como fonte geradora de valores aleatórios no intervalo de 0 a 1, função indispensável para a aplicação dos inúmeros métodos numéricos destinados a geração de variáveis aleatórias. Na Figura 5.4 observa-se um histograma relativo a uma amostra com 10.000 pontos de uma variável aleatória com comportamento descrito por uma distribuição Uniforme [0; 1].

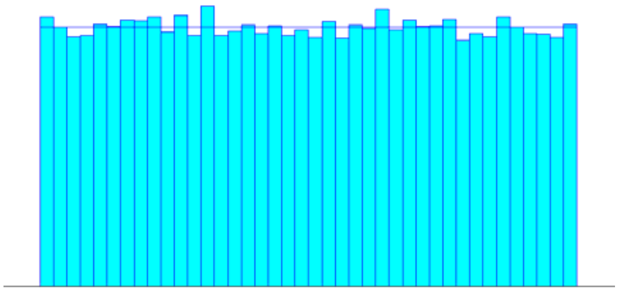

Figura 5.4: Exemplo de uma Uniforme com a=0 e b=1

Triangular

A semelhança da distribuição Uniforme, o emprego da distribuição Triangular ocorre, principalmente, quando se desconhece a curva associada a uma variável aleatória, mas têm-se boas estimativas dos seus limites: inferior (a) e superior (b), bem como, de seu valor mais provável (m). Por exemplo, é conhecido o valor mínimo, mais provável e máximo associados ao tempo necessário para a realização de um teste de qualidade de um produto. Pelo fato de se estar estimando o comportamento de variáveis com uma informação a mais (o valor mais provável), implica, em geral, em perspectivas de resultados mais aderentes à realidade do que aqueles com base apenas nos valores mínimo e máximo (com o emprego da distribuição uniforme). Na Figura 5.5 tem-se o histograma de uma distribuição Triangular (a, m ,b).

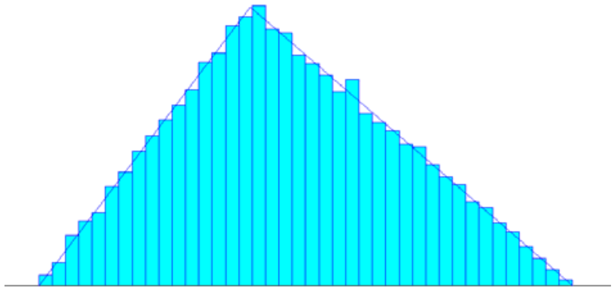

Figura 5.5: Triangular com parâmetros $a = 1$, $m = 3$ e $b= 6$

Exponencial

A principal característica da distribuição exponencial, e razão da sua grande aplicabilidade em sistemas de filas, é sua falta de *memória*. Todo fenômeno aleatório descrito por esta distribuição se caracteriza pela total imprevisibilidade, mesmo que se conheça seu passado.

Pode-se imaginar, por exemplo, uma variável cujo comportamento é delineado por uma distribuição exponencial com média β. Suponha agora, que ao se observar uma ocorrência desta variável, se passe a contar o tempo decorrido até uma próxima ocorrência. O tempo médio decorrido entre as duas observações será β. Suponha agora que não se observe ocorrências por um longo tempo $t = x$, ainda assim, a situação será idêntica, isto é, o tempo médio esperado até a próxima ocorrência continuará sendo β. Em outras palavras, isto significa dizer que o conhecimento prévio do tempo da ocorrência do último evento não ajuda na previsão do tempo de ocorrência do próximo evento. Na observação de um processo cujo tempo é exponencialmente distribuído, mesmo que se conheçam os resultados do mesmo processo em outros experimentos no passado e ainda que se saiba quanto tempo já passou desde que o processo foi iniciado, nada se poderá afirmar sobre quanto tempo ainda resta para o mesmo se encerrar.

A exponencial é muito utilizada na modelagem de tempos decorridos entre dois eventos, particularmente se estes forem causados por um grande número de fatores independentes. Por exemplo: tempo entre duas chegadas consecutivas de entidades em um sistema ou entre duas falhas de um equipamento. Embora não seja recomendável, devido a sua alta variabilidade, algumas vezes é empregada para caracterizar tempos de serviços, por exemplo, para modelagem de tempos de atendimento em centrais de atendimento telefônico (*Call Centers*).

Nas Figuras 5.6 e 5.7 pode-se observar histogramas relativos a duas amostras com 10.000 pontos cada uma delas. O primeiro refere-se a uma variável exponencialmente distribuída com média $\beta = 1$. Os valores mínimo e máximo da amostra foram 0,0 e 9,5; respectivamente. No segundo a variável possui média $\beta = 10$. Neste caso, os valores mínimo e máximo foram 0,0 e 95,5 respectivamente. Tais valores atestam a alta variabilidade associada a esta distribuição e os cuidados que devem ser tomados no seu emprego.

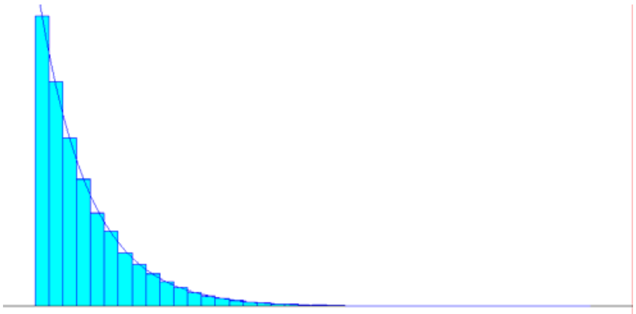

Figura 5.6: Exponencial com média $\beta = 1$

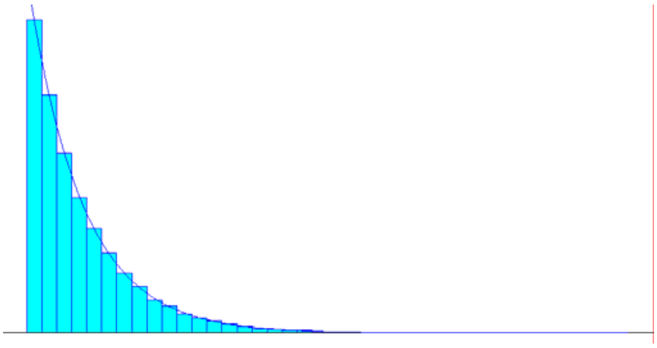

Figura 5.7: Exponencial com média $\beta = 10$

Lognormal

O logaritmo natural de uma variável que segue uma distribuição normal possui uma distribuição lognormal. Quando a variável sob análise é resultante do produto de um grande número de variáveis aleatórias positivas é comum que esta variável tenha uma tendência a uma distribuição Lognormal (Figuras 5.8 a 5.10). Um exemplo prático da multiplicação de variáveis aleatórias é a determinação da taxa de retorno de um ativo durante um mês, obtido como resultado da multiplicação da variação de preços diários desse ativo. Outro exemplo de emprego da distribuição Lognormal é em engenharia de confiabilidade para descrever falhas causadas por fadiga de material, incertezas e taxas de falhas, além de uma variedade de outros fenômenos. Uma importante propriedade é que se duas variáveis aleatórias têm distribuição Lognormal, a função gerada pelo produto dessas duas variáveis também terá distribuição Lognormal.

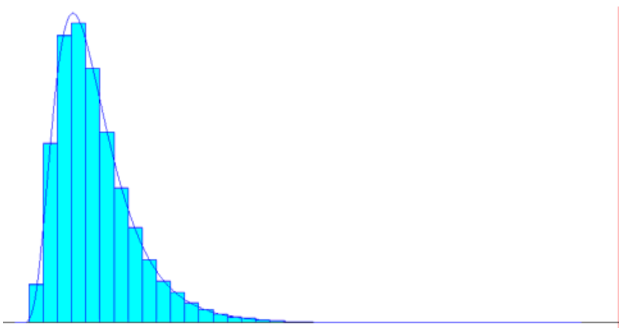

Figura 5.8: Lognormal com média =1 e desvio =0.5

Figura 5.9: Lognormal com média =1 e desvio =1

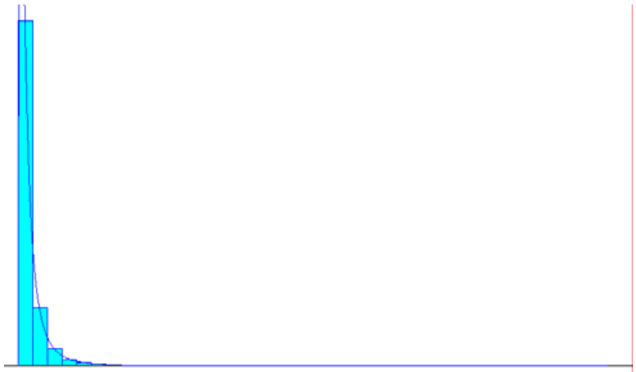

Figura 5.10: Lognormal com media =1 e desvio =1.5

Erlang

A distribuição Erlang é utilizada como uma extensão da exponencial, especialmente quando o fenômeno aleatório é observado ao longo de diversas etapas as quais podem ser descritas, de forma independente, com distribuições exponenciais. Desta forma, a soma destas *m* distribuições exponenciais de média β é uma distribuição Erlang com parâmetros β e *m*.

Pode ser empregada, por exemplo, para modelar um servidor que represente uma série de outros *m* servidores cujos tempos de processo ou serviço possam ser descritos por distribuições exponenciais com média β. Outro emprego desta distribuição é na modelagem de tempos de reparos (manutenção de sistemas) ou ainda na modelagem do tempo total de permanência em centrais de atendimento *(Call Centers*). Esta distribuição é um caso especial da distribuição Gama, onde o parâmetro de forma (α) é um inteiro (*k*).

Nas Figuras 5.11 a 5.13, apresenta-se distribuições Erlang com mesma média (β) e com valores crescentes para *m* (1, 2 e 3). O primeiro gráfico confunde-se com uma distribuição exponencial de mesma média.

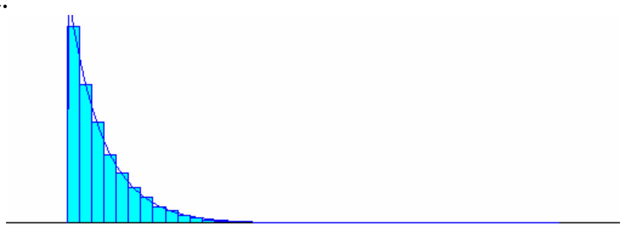

Figura 5.11: Erlang com β = 1 e *m*=1

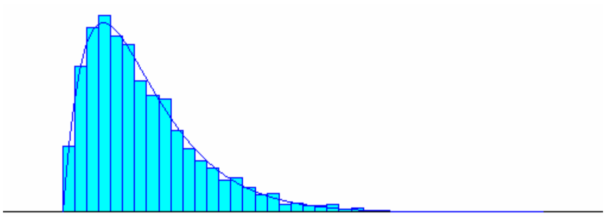

Figura 5.12: Erlang com $\beta = 1$ e m=2

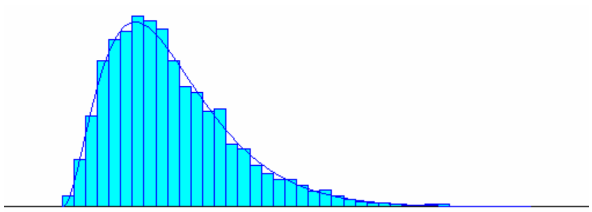

Figura 5.13: Erlang com $\beta = 1$ e m=3

Gama

A distribuição Gama (Figura 5.14) é uma generalização da distribuição Erlang. A diferença é que esta permite que m assuma valores não inteiros. Quando m é inteiro, as duas distribuições se confundem (Figura 5.15). É empregada em condições de modelagem semelhantes a da distribuição Erlang. A curva da distribuição Gama com média $\beta = 1$ e parâmetro $m = 1$, terá a mesma forma da Erlang, com mesmos parâmetros e da exponencial com média a.

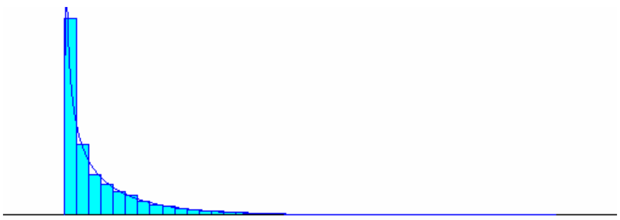

Figura 5.14: Gama com β =1 e m =½

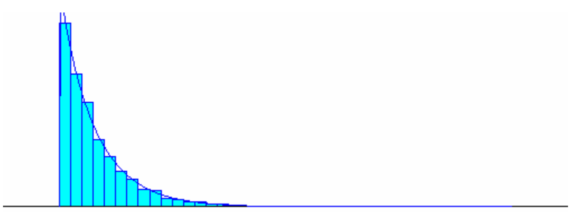

Figura 5.15: Gama com β =1 e m =1

Beta

A distribuição Beta (Figuras 5.16 a 5.20) é utilizada para caracterizar variáveis aleatórias cujos valores encontrem-se dentro do intervalo [0;1]. Desta maneira, uma de suas principais aplicações está na representação de proporções ou frações. Como exemplos podem-se citar: a proporção de defeituosos em lotes de produtos, a fração de pacotes que devem ser retransmitidos, ou ainda as partes de trechos de estradas que anualmente necessitam reparos. Outra característica desta distribuição é o grande número de formas que ela pode assumir, dependendo de seus dois parâmetros de forma a e escala b.

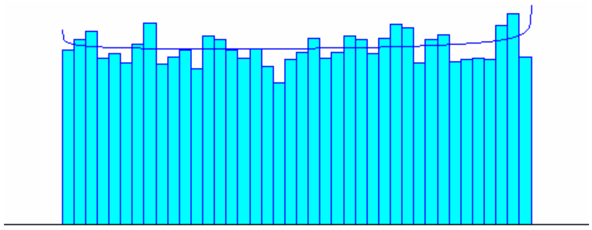

Figura 5.16: Beta com $a = 1$ e $b = 1$

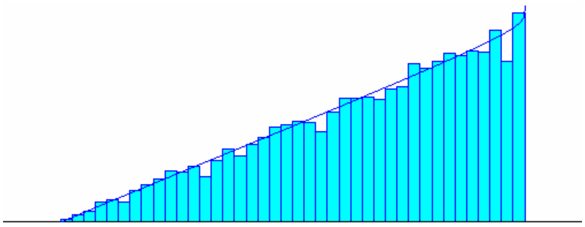

Figura 5.17: Beta com $a = 2$ e $b = 1$

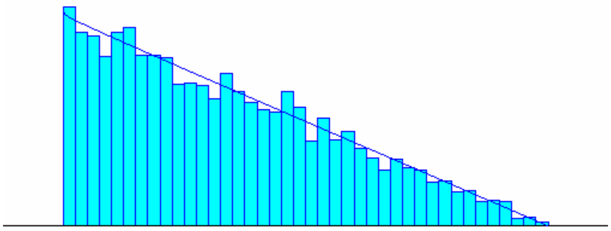

Figura 5.18: Beta com $a = 1$ e $b = 2$

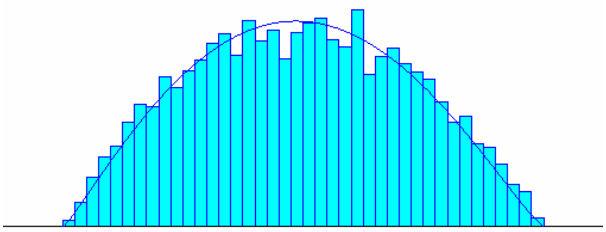

Figura 5.19: Beta com $a = 2$ e $b = 2$

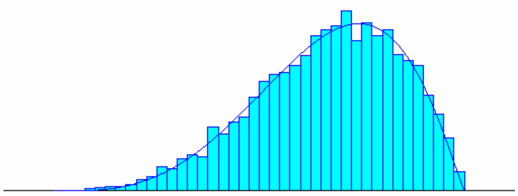

Figura 5.20: Beta com $a = 4$ e $b = 2$

Weibull

A principal utilização da distribuição Weibull é na representação de variáveis aleatórias que descrevam características de confiabilidade de sistemas ou equipamentos. Uma aplicação típica é na modelagem de falhas de componentes ou sistemas. Por exemplo, se um sistema é formado por inúmeros componentes que falham de forma independente e, se uma das partes falharem, o sistema inteiro falha, então, o tempo decorrido entre falhas do sistema pode ser modelado utilizando-se uma distribuição Weibull (Figuras 5.21 a 5.24). Assim como a Beta, esta é uma distribuição que pode assumir vários perfis dependendo de seus parâmetros, especialmente do *b* (parâmetro de forma). Por exemplo, com *b* = 3,602 a curva se assemelha a uma distribuição normal (Figura 5.24). Um perfil da distribuição Weibull poderá apresentar longas caudas à direita ou esquerda, formas de sino ou mesmo acentuados picos no valor modal.

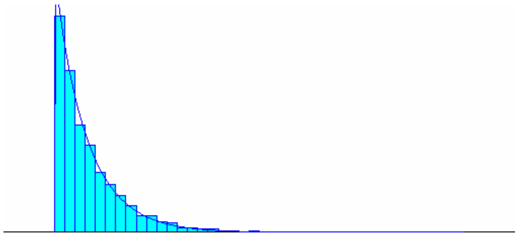

Figura 5.21: Weibull com *a*=1 e *b*=0,5

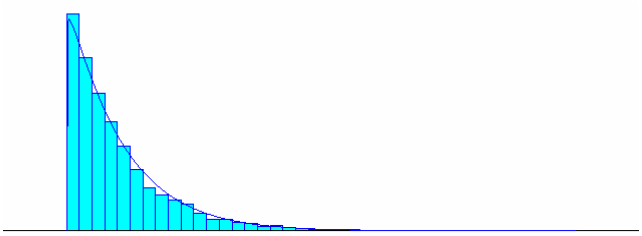

Figura 5.22: Weibull com a=1 e b=1

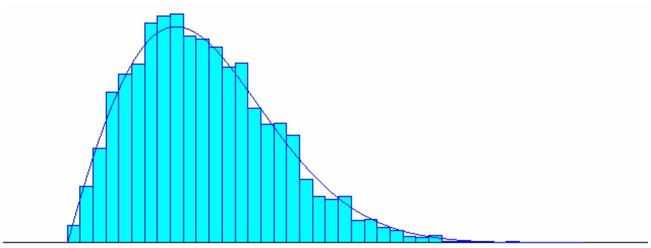

Figura 5.23: Weibull com a=1 e b=2

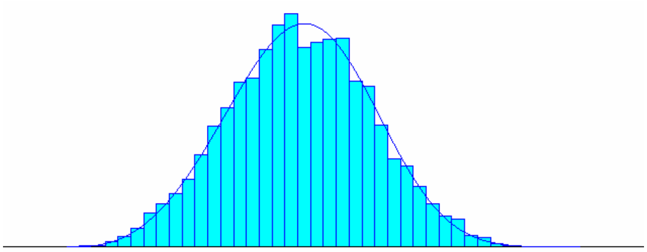

Figura 5.24: Weibull com a=1 e b=3,602

Principais Distribuições Discretas

Poisson

Na modelagem de sistemas, tanto via tratamento analítico (por meio da Teoria das Filas, por exemplo) ou tratamento por simulação, a distribuição de Poisson é, sem dúvida, uma das mais empregadas. Esta distribuição discreta é utilizada para modelar o número de ocorrências (valores discretos) que uma variável possa assumir, ao longo de um intervalo contínuo. Os exemplos de seu emprego são diversos: número de requisições feitas a um servidor em determinado intervalo de tempo, número de componentes de um sistema que falham num intervalo de tempo, número de chegadas de entidades a um sistema em determinado intervalo de tempo, etc.. Nem sempre o intervalo contínuo refere-se a intervalos de tempo. Por exemplo: número de falhas observadas a cada cem metros lineares na produção de bobinas de papel ou ainda, número de erros encontrados por lauda digitada, também são caracterizados como variáveis que se distribuem de acordo com um Poisson (Figura 5.25).

A distribuição de Poisson é particularmente apropriada quando se está modelando um sistema com inúmeras fontes independentes de chegadas. A este tipo particular de processo de chegadas, dá-se o nome de *processo de Poisson*. As distribuições Poisson e Exponencial possuem um relação importante. Se os tempos decorridos entre dois eventos podem ser descritos por uma distribuição exponencial com média β, então diz-se que o número de ocorrências deste mesmo evento num intervalo *a* de tempo segue uma distribuição de Poisson com média $1/\beta = \lambda$

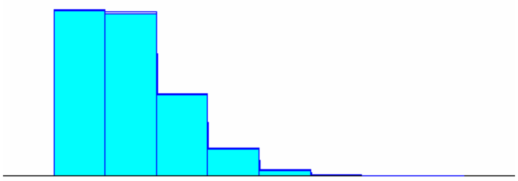

Figura 5.25: Poisson com média $\lambda = 1$

Uniforme Discreta

A distribuição uniforme discreta é empregada quando a variável aleatória que está sendo modelada pode assumir apenas valores inteiros, todos com igual probabilidade, limitados a um intervalo [mínimo, máximo]. Alguns exemplos de seu emprego são: número de destinos possíveis para uma mensagem em uma rede local de computadores, número de peças que se encontram no buffer limitado em um centro de usinagem, etc..

Além destas duas, pode-se citar outras distribuições discretas tais como a geométrica, a binomial e ainda a distribuição de Bernoulli, como importantes na modelagem de sistemas.