Distribuições de Probabilidade
Prof. Igor Nascimento

26 de maio de 2021

1 Introdução

Para facilitar o cálculo de probabilidades, existem modelos de probabilidade que nos auxiliam a mensurar a incerteza relacionados aos experimentos aleatórios.

Com isso, vamos continuar utilizando os conceitos de variáveis aleatórias, comumente representadas por letras maiúsculas X,Y e Z. As letras minúsculas x,y e z serão utilizadas para indicar um valor possível para essa variável.

É importante lembrar que os valores que as variáveis aleatórias assumem são sempre números. Isto é, a variável aleatória transforma o espaço amostral de um experimento aleatório em um espaço amostral números, podendo ser discreto ou contínuo.

2 Distribuições contínuas

2.1 Distribuições Gaussiana ou normal

Se um conjunto de dados tem distribuição de probabilidade normal, temos as seguintes características:

$$f(X = x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(X-\mu)^2}{2\sigma}\right)} \tag{2.1}$$

- média (μ) e desvio padrão (σ) definem as probabilidades utilizando a tabela normal.
- X assume valores entre $-\infty$ e ∞, mas é mais provável próximos à média e concentramos em torno do desvio padrão.
- Assimetria 0 e curtose 0.2631. Essa distribuição serve de referência para as demais (Normal e não Normal).
- área total abaixo da curva é 1 ou 100%
- exemplos de comportamento normal: “pressão arterial”, “altura”, erros astronômicos Caire (2012).

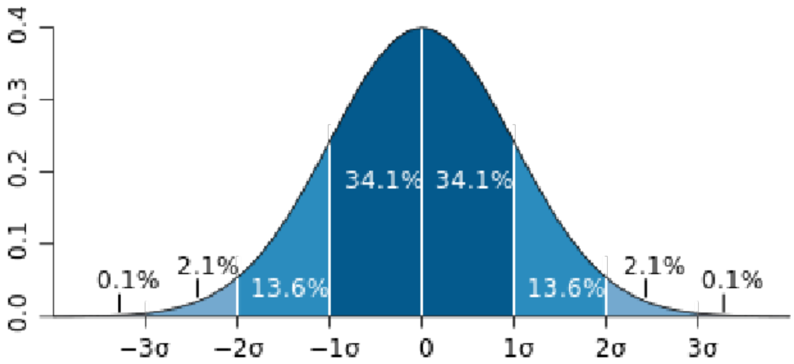

Figura 1 – Distribuição Normal. Fonte: google

- é usada para quase todos os testes estatísticos
- qualquer distribuição normal pode ser transformada em uma distribuição normal com $\mu = 0$ e $\sigma = 1$, fazendo:

$$Z = \frac{X - \mu}{\sigma}$$

- os pontos de σ acima ou abaixo da média são pontos de inflexão

A distribuição normal padrão, $\mu = 0$ e $\sigma = 1$, a probabilidade $P(Z < 0) = 50\%$. Isto é, o valor de 0 também é a mediana, distribuição simétrica. A probabilidade $P(-1 < Z < 0) = 34,13\%$ e por complemento, a probabilidade $P(Z < -1) = 50\% - 34,13\% = 13.6\%$. A probabilidade $P(Z < -2) = 2.1\%$ e $P(Z < -3) = 0.1\%$. A probabilidade $P(0 < Z < 1) = 34,13\%$ e por complemento, a probabilidade $P(Z > 1) = 50\% - 34,13\% = 13.6\%$. A probabilidade $P(Z > 2) = 2.1\%$ e $P(Z > 3) = 0.1\%$.

Também é possível obter probabilidades em intervalos maiores. Por exemplo, a probabilidade $P(-1 < Z < 3)$ é igual a duas partes $P(-1 < Z < 3) = P(-1 < Z < 0) + P(0 < Z < 3)$, isto é, duas regiões.

Veja o aplicativo desenvolvido e veja esses resultados para a Normal $Z \sim N(\mu = 0, \sigma = 1)$.

Com isso, podemos obter qualquer proabilidade. Considere X uma variável aleatória com o padrão normal $X \sim N(\mu = 500, \sigma = 100)$. Desejamos obter a probabilidade $P(400 < X < 800)$. Vamos descobrir quais sãos os valores 400 e 800 na normal $Z \sim N(\mu = 0, \sigma = 1)$, por meio da transformação:

$$z_1 = \frac{400 - \mu}{\sigma} = \frac{400 - 500}{100} = \frac{-100}{100} = -1$$

Além disso:

$$z_2 = \frac{800 - \mu}{\sigma} = \frac{800 - 500}{100} = \frac{300}{100} = +3$$

Portanto, $P(400 < X < 800) = P(-1 < Z < 3) = P(-1 < Z < 0) + P(0 < Z < 3) = 0,8399 = 83,99\%$.

Além disso, também representamos as distribuições de probabilidade pela distribuição acumulada $F(X)$.

$$F(X = x) = \int_{-\infty}^{x} f(X = x) \tag{2.2}$$

Com isso, podemos verifiquer se os dados observados se ajustam à distribuição normal por meio dos percentis. O gráfico que compara os percentis da distribuição observada e da distribuição normal é chamado de qq-plot, isto é, o gráfico dos percentis.

2.2 Distribuições Exponencial

Outra distribuição de probabilidade para variáveis aleatórias contínuas é a expoencial, que tem aplicações em confiabilidade de sistemas. Se X tem distribuição exponencial com parâmetro λ, então X é estritamente positiva e a função de distribuição de probabilidade é expressa por:

$$f(X=x)=\frac{e^{-\frac{x}{\lambda}}}{\lambda}, x\geq 0 \tag{2.3}$$

No caso discreto, somávamos cada valor que a variável aleatória poderia assumir multiplicado pela probabilidade. Mas isso só era possível em razão dos valores assumidos serem discretos, isto é, pontos específicos.

Agora a variável aleatória X, no caso da exponecial, pode assuimr qualquer valor $\{x|x\leq 0\}$. Com isso, vamos utilizar a integração para realizar tal "soma".

Vamos iniciar comprovando que a soma da probabilidades da exponecial é 1. Isto é, a probabilidade de ocorrência de um valor aleatório de X maior ou igual a zero é igual a 1.

$$\int_0^{\infty}\frac{e^{-\frac{x}{\lambda}}}{\lambda}dx=\left[-e^{-\frac{x}{\lambda}}\right]_0^{\infty}=0-(-1)=1 \tag{2.4}$$

Com isso, podemos calcular o valor médio de uma variável aleatória X que possui distribuição exponencial com parâmetro λ, representado por $X\sim Exp(\lambda)$.

$$E(x)=\int_0^{\infty}xf(X=x)dx=\int_0^{\infty}x\frac{e^{-\frac{x}{\lambda}}}{\lambda}dx \tag{2.5}$$

Uma das forma de resolver essa integral é por meio de integrais parciais. Lembre-se:

$$\int g(x)f^{'}(x)dx=g(x)f(x)-\int g^{'}(x)f(x)dx \tag{2.6}$$

sendo $g^{'}(x)=\frac{\partial g(x)}{\partial x}$. Vamos considerar $g(x)=x$ e $f^{'}(x)=\frac{e^{-\frac{x}{\lambda}}}{\lambda}$. Então $g^{'}(x)=1$ e $f(x)=-e^{-\frac{x}{\lambda}}$. Com isso, temos:

$$\int g(x)f^{'}(x)dx=g(x)f(x)-\int g^{'}(x)f(x)dx \tag{2.7}$$

$$E(X)=\int_0^{\infty}x\frac{e^{-\frac{x}{\lambda}}}{\lambda}dx=\left[-xe^{-\frac{x}{\lambda}}\right]_0^{\infty}-\int_0^{\infty}\left[-e^{-\frac{x}{\lambda}}\right]dx \tag{2.8}$$

Resolvendo a integral e substituindos os limites, temos:

$$E(X)=\int_0^{\infty}x\frac{e^{-\frac{x}{\lambda}}}{\lambda}dx=0+\left[-\lambda e^{-\frac{x}{\lambda}}\right]_0^{\infty} \tag{2.9}$$

$$E(X)=\int_0^{\infty}x\frac{e^{-\frac{x}{\lambda}}}{\lambda}dx=\lambda \tag{2.10}$$

Com isso, temos que a média de uma distribuição expoencial com parâmetro λ é $E(X)=\lambda$

Também podemos obter a variância da distribuição:

$$V(X)=\int_0^{\infty}(x-E(X))^2\frac{e^{-\frac{x}{\lambda}}}{\lambda}dx \tag{2.11}$$

$$V(X)=\int_0^{\infty}(x-\lambda)^2\frac{e^{-\frac{x}{\lambda}}}{\lambda}dx \tag{2.12}$$

Resolvendo a equação acima, temos que:

$$V(X) = \lambda^2 \tag{2.13}$$

A prova fica como exercício. Dúvidas consulte Magalhães (2006).

A distribuição acumulada é

$$F(X = x) = \int_0^x f(X = x) = 1 - e^{-\lambda x} \tag{2.14}$$

Então $F(X = x)$ representa a probabilidade de que a variável aleatória X assuma valores menores do que x, isto é, $F(X = x) = P(X < x)$. De maneira análoga, podemos obter a probabilidade de que a variável aleatória X assuma valores maiores ou iguais a x, isto é, $P(X > x) = 1 - F(X = x)$.

Na maioria dos casos, se X representa o tempo de vida de equipamentos, podemos definir a confiabilidade R(X):

$$R(X) = 1 - F(X = x) = e^{-\frac{x}{\lambda}} \tag{2.15}$$

O tempo de vida (em horas) de um transistor pode ser considerado uma variável aleatória com distribuição exponencial com média $E(X) = \lambda = 500h$. Qual a probabilidade do transistor durar mais do que o tempo médio? $P(X > 500)$

Por meio da função acumulada, temos que

$$P(X < 500) = F(X = 500) = 1 - e^{-\frac{x}{500}} = 0.6321206 \tag{2.16}$$

Essa é a probabilidade de falha até $x = 500h$, então a probabilidade do transistor durar mais de $x = 500h$ é:

$$R(X) = 1 - 0.63212 = 0.36788 \tag{2.17}$$

2.3 Distribuições Weibull

2.3.1 Distribuições Weibull dois parâmetros

A distribuição Weibull é, principalmente, utilizada para modelar o tempo de vida e resistência de equipamentos e materiais, entre eles fadiga de aço ST-37 e forças das fibras de algodão Batista (1989).

Se X tem distribuição weibull com parâmetro α (parâmetro de forma) e β (parâmetro de escala), então X é estritamente positiva e a função de distribuição de probabilidade é expressa por:

$$f(X = x) = \alpha\beta^{-\alpha}x^{\alpha-1}e^{-\left(\frac{x}{\beta}\right)^{\alpha}}, x \geq 0 \tag{2.18}$$

Diante da relação entre os parâmetros e a forma da distribuição, essa distribuição é muito utilizada. O parâmetro de escala β no ajuda a compreender a dimensão que a curva assume Batista (1989). Utilizando o aplicativo desenvolvido para fins pedagógicos, escolhendo $\beta = 2$ e $\beta = 3$ e $\alpha = 3.6$, temos a figura 2:

Perceba que com $\alpha = 3.6$ temos uma distribuição em formato de sino, semelhante à distribuição gaussiana. Caso esse valor seja menor do 3.6, temos uma distribuição mais assimétrica, conforme mostra a figura 3, com $\beta = 3$ e $\alpha = 1.7$ (vermelho) e $\alpha = 3.6$ (azul).

Perceba que se o parâmetro de forma $\alpha = 1$, temos a distribuição Exponencial, com parâmetro $\lambda = \beta$.

$$f(X = x|\alpha = 1) = \beta^{-1}e^{-\left(\frac{x}{\beta}\right)}, x \geq 0 \tag{2.19}$$

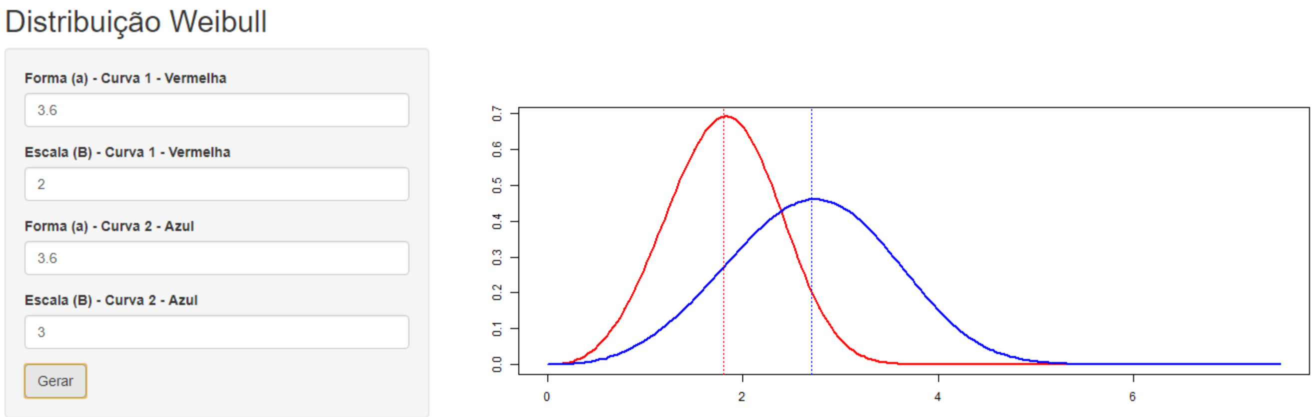

Figura 2 – Distribuição Weibull com $\beta = 2$ (vermelho) e $\beta = 3$ (azul) e $\alpha = 3.6$

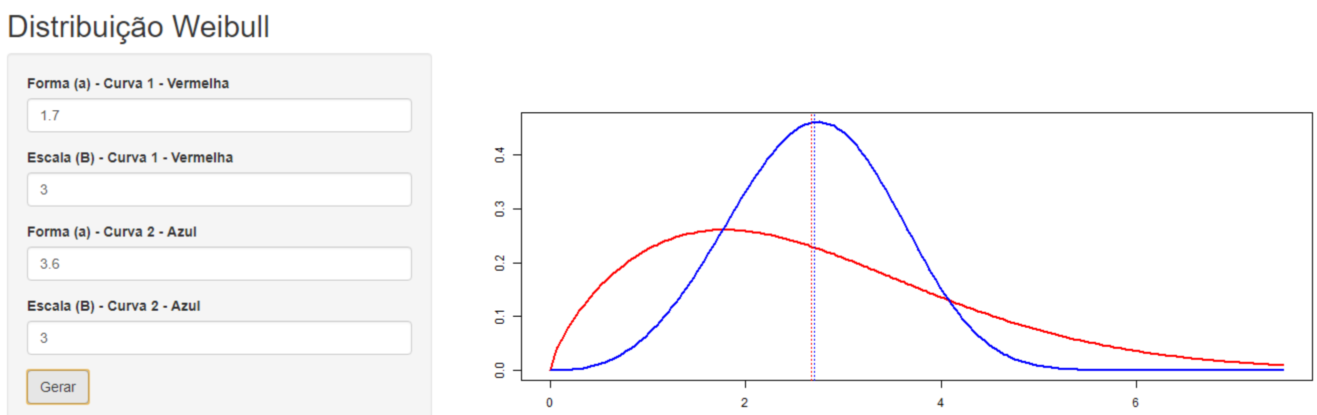

Figura 3 – Distribuição Weibull com $\alpha = 1.7$ (vermelho) e $\alpha = 3.6$ (azul) e $\beta = 3$

Dessa forma, dizemos que a distribuição Expoencial é um caso particular da Weibull, quando o parâmetro de forma $\alpha = 1$ é igual a 1, conforme mostra a figura 4

Cada uma dessas configurações de distribuições ajustam melhor alguns tipos de dados.

Se $\alpha < 1$ então há uma elevada probabilidade de falha no início da utilização, isto é, falhas precoces Dias (2014). Se α for próximo de 3.6, a distribuição é semalhante a distribuição Normal. Já para valores de α muito grandes, a distribuição se concentra em determinado instante da vida Dias (2014).

A média e variância de uma variável aleatória com distribuição Weibull é:

$$E(X) = \beta\Gamma\left(1 + \frac{1}{\alpha}\right) \tag{2.20}$$

Sendo Γ a função Gama.

A distribuição acumulada é

$$F(X = x) = \int_0^x f(X = x) = 1 - e^{-\left(\frac{x}{\beta}\right)^\alpha} \tag{2.21}$$

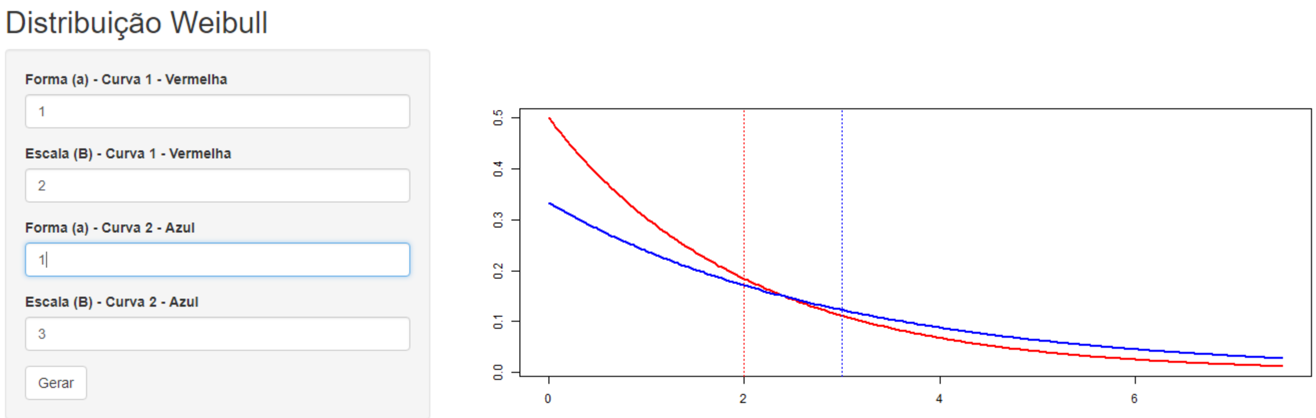

Figura 4 – Distribuição Exponencial como caso particular da Weibull com $\beta = 2$ (vermelho) e $\beta = 3$ (azul) e $\alpha = 1$

Na maioria dos casos, se X representa o tempo de vida de equipamentos, podemos definir a confiabilidade R(X):

$$R(X) = 1 - F(X = x) = e^{-\left(\frac{x}{\beta}\right)^{\alpha}} \tag{2.22}$$

O tempo de vida (em horas) de um motor pode ser considerado uma variável aleatória com distribuição weibull com $\alpha = 2.3$ e $\beta = 4.2$. Qual é o tempo médio? Qual a probabilidade do motor durar mais do que 5 anos?

A média é:

$$E(X) = \beta\Gamma\left(1 + \frac{1}{\alpha}\right) = 4.2\Gamma\left(1 + \frac{1}{2.3}\right) = 3.7208 \tag{2.23}$$

Por meio da função acumulada, temos que

$$P(X < 5) = F(X = 5) = 1 - e^{-\left(\frac{5}{4.2}\right)^{2.3}} = 0.775378 \tag{2.24}$$

Essa é a probabilidade de falha até $x = 5$, então a probabilidade do motor durar mais de $x = 5$ é:

$$R(X) = 1 - 0.775378 = 0.224622 \tag{2.25}$$

Na próxima etapa, aprenderemos como, baseado em dados observados, podemos modelar uma distribuição de probabilidade estimando os parâmetros. Procure sobre métodos para estimação de tais parâmetros Rocha et al. (2015). Veja ainda uma aplicação do modelo Weibull para o estado do Piauí Moreira et al. (2019).

Referências

BATISTA, J. L. F. *Função Weibull como modelo para a distribuição de diâmetros de espécies arbóreas tropicais.* Tese (Doutorado) — Universidade de São Paulo, 1989.

CAIRE, E. A história da origem da curva normal. Universidade Estadual Paulista (UNESP), 2012.

DIAS, A. *Confiabilidade na manutenção industrial.* [S.l.], 2014.

MAGALHÃES, M. N. *Probabilidade e variáveis aleatórias.* São Paulo: Edusp, 2006.

MOREIRA, T. B. et al. Proposta de aplicação da manutenção centrada na confiabilidade no desenvolvimento do plano estratégico da manutenção: um estudo de caso/proposal for implementation of reliable maintenance maintenance in developing the maintenance strategic plan: a case study. *Brazilian Journal of Business*, v. 1, n. 3, p. 842–856, 2019.

ROCHA, R. J. et al. Estudo sobre métodos de estimação de f (t) para o cálculo dos parâmetros de distribuições de weibull com a aplicação em uma análise de falhas de garrafas pet. Joinville, SC, 2015.