UNIVERSIDADE DE SÃO PAULO
Faculdade de Zootecnia e Engenharia de Alimentos

ZAB1111 – Estatística Básica

Prof. César Gonçalves de Lima cegdlima@usp.br

Aula 10 – PROBABILIDADE (4)

MODELOS PARA VARIÁVEIS ALEATÓRIAS CONTÍNUAS

7. VARIÁVEIS ALEATÓRIAS CONTÍNUAS

Uma variável aleatória contínua é uma função que pode assumir infinitos valores num intervalo de números reais.

Associaremos a cada subintervalo do seu domínio uma probabilidade usando uma função densidade de probabilidade (f.d.p.).

Definição 1. Uma função $f(x)$ definida para $x \in [a, b]$, é chamada de **função densidade de probabilidade** (f.d.p.) se satisfaz as seguintes condições:

a) $f(x)$ é positiva, para todo $x \in [a, b]$;

b) $\int_a^b f(x)dx = 1$, ou seja, a área sob a curva representativa de $f(x)$, entre as abscissas a e b, é igual a um.

Observe que:

a) A função $f(x)$ não define uma probabilidade.

b) O que define uma probabilidade é o valor da integral de $f(x)$ no intervalo $[x_1, x_2]$, por exemplo, que coincide com a área da região sob a curva de $f(x)$, o eixo das abscissas e os limites de integração.

c) Para **calcular a probabilidade** da v.a. X assumir valores entre x_1 e x_2, com $x_1 < x_2$, precisamos resolver:

$$P(x_1 < \mathrm{X} < x_2) = \int_{x_1}^{x_2} f(x)dx$$

d) $P(X = k) = 0$ porque $\int_k^k f(x)dx = [F(x)]_k^k = F(k) - F(k) = 0$.

e) Definimos a **função distribuição acumulada** da variável contínua X, com f.d.p. $f(x)$, por:

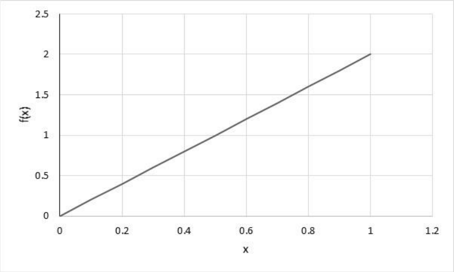
$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(z)dz$$

f) Se X é uma v.a. <u>contínua</u> definida no intervalo $[a; b]$ e $f(x)$ é sua f.d.p. definimos:

$$E(\mathrm{X}) = \int_a^b x\, f(x)\, dx \qquad var(\mathrm{X}) = \int_a^b [x - E(X)]^2 f(x)\, dx$$

Exemplo 1. Dada a função $f(x) = 2x$, $x \in [0, 1]$ pede-se:

$i)$ Verificar se $f(x)$ é uma função densidade de probabilidade.



$f(x)$ é positiva para $x \in [0, 1]$

$\int_0^1 2x\, dx = x^2|_0^1 = 1$

$ii)$ Calcular $P(0 < X < 0{,}5)$ e $P(0{,}2 < X < 0{,}7)$.

- $P(0 < X < 0{,}5) = x^2|_0^{0,5} = 0{,}25$
- $P(0{,}2 < X < 0{,}7) = x^2|_{0,2}^{0,7} = 0{,}49 - 0{,}04 = 0{,}45$

A seguir conheceremos os modelos Exponencial e Normal, que são úteis e usados em diversas áreas de pesquisa.

Em cada caso precisaremos conhecer sua f.d.p., seu gráfico, sua média e variância e saber calcular probabilidades.

8. ALGUNS MODELOS PROBABILÍSTICOS PARA V. A. CONTÍNUA

8.1. DISTRIBUIÇÃO UNIFORME

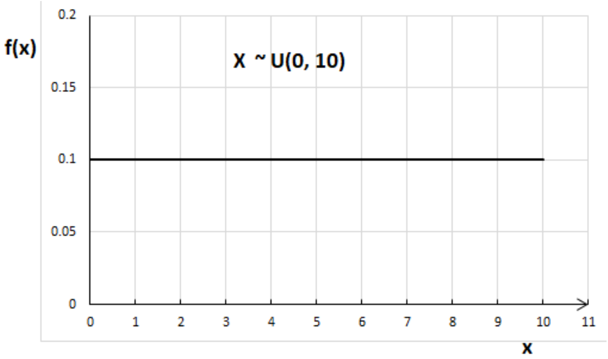
É a distribuição de probabilidades de v.a. contínuas mais simples: sua f.d.p. é constante e a probabilidade $P(x_1 < X < x_2)$ é proporcional ao tamanho do intervalo.

Definição 8.1. Dizemos que a v.a. contínua X tem distribuição uniforme no intervalo real $[a, b]$ se a sua f.d.p. for dada por:

$$f(x) = \frac{1}{b-a}$$

Pode-se provar que: $F(x) = \frac{x-a}{b-a}$, para todo $x \in [a, b]$, $E(X) = \frac{a+b}{2}$ e $var(X) = \frac{(b-a)^2}{12}$

Exemplo 1. Se $X \sim U(0,10)$ calcular $P(1 < X < 3)$ e $P(2 < X < 4)$



Para calcular as probabilidades vamos usar a função de distribuição acumulada:

$$F(x) = \frac{x-a}{b-a} = \frac{x-0}{10-0} = \frac{x}{10}$$

- $P(1 < X < 3) = F(3) - F(1) = \frac{3}{10} - \frac{1}{10} = 0{,}3 - 0{,}1 = 0{,}2$
- $P(2 < X < 4) = F(4) - F(2) = \frac{4}{10} - \frac{2}{10} = 0{,}4 - 0{,}2 = 0{,}2$

Note que as probabilidades são iguais porque os intervalos de cálculo de ambas as probabilidades são iguais

8.2. A DISTRIBUIÇÃO EXPONENCIAL

É usada para descrever o tempo de ocorrência de um evento. **Exemplo:** o tempo de vida de uma bateria de celular; o tempo exigido para um técnico executar certa tarefa; o tempo de chegada de um carro a um posto de pedágio *etc*.

Definição 8.2. Dizemos que a v.a. contínua X, definida para valores positivos, tem distribuição *exponencial* de parâmetro $\lambda > 0$, se a sua f.d.p. é dada por:

$$f(x) = \frac{1}{\lambda} e^{-x/\lambda} = \frac{1}{\lambda} exp(-x/\lambda)$$

Pode-se provar que: $E(X) = \lambda$ e $var(X) = \lambda^2$

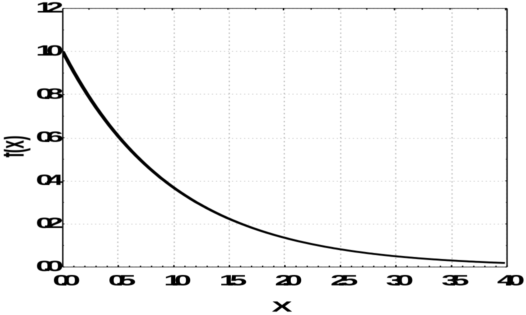


Figura 8.1. Função densidade de probabilidade de $X \sim exp(\lambda = 1)$

Note que:

$$P(0 < X < 1) < P(1 < X < 2) < P(2 < X < 3)$$

ou seja, as probabilidades de ocorrência de X em intervalos de mesmo tamanho serão menores, quanto mais distantes da origem estiverem os seus limites.

Para calcular probabilidades de uma variável com distribuição exponencial usamos a sua <u>função distribuição acumulada</u>:

$$F(x) = P(X \leq x) = 1 - e^{-x/\lambda}$$

Exemplo: Calcular a probabilidade da variável $X \sim exp\ (\lambda)$ assumir um valor entre x_1 e x_2.

$$\begin{aligned} P(x_1 < X < x_2) &= F(x_2) - F(x_1) \\ &= \left[1 - e^{-x_2/\lambda}\right] - \left[1 - e^{-x_1/\lambda}\right] \\ &= e^{-x_1/\lambda} - e^{-x_2/\lambda} \end{aligned}$$

Exemplo 2. O tempo de carga (em horas) de um determinado tipo de bateria de celular é uma v.a. T, contínua, com distribuição exponencial de média $\lambda = 50$h. Calcular a probabilidade de que o tempo de carga desta bateria dure entre 50 e 60 horas de uso.

Resolução: Se $\mathrm{T} \sim exp(50) \Rightarrow f(t) = \frac{1}{50}\, e^{-t/50} \Rightarrow F(t) = 1 - e^{-t/50}$, para qualquer $t > 0$. Então:

$$
\begin{aligned}
P(50 \leq \mathrm{T} \leq 60) &= \int_{50}^{60} \frac{1}{50} e^{-t/50}\, dt = F(60) - F(50) \\
&= e^{-50/50} - e^{-60/50} = e^{-1} - e^{-1,2} \\
&= 0,3679 - 0,3012 = 0,0667
\end{aligned}
$$

$\therefore$ A chance de a carga desta bateria de celular durar entre 50 e 60 horas é de apenas 0,0667.

8.3. O MODELO NORMAL (ou de Gauss)

A distribuição normal foi introduzida pelo matemático Abraham de Moivre (1733) e é uma das distribuições probabilísticas mais importantes da Estatística, pois é usada para descrever inúmeros fenômenos físicos, biológicos e financeiros.

Definição 8.3. Dizemos que a variável contínua X tem distribuição normal, com parâmetros μ e σ^2 se a sua f.d.p. é dada por:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad \text{para } -\infty < x < \infty$$

Pode-se provar que: $\mu = E(\mathrm{X})$ e $\sigma^2 = var(\mathrm{X})$.

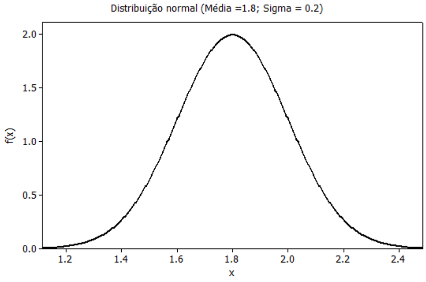
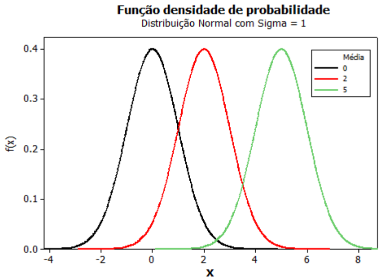


Figura 8.3. Distribuição normal com média $\mu = 1{,}8$ e variância $\sigma^2 = 0{,}04$.

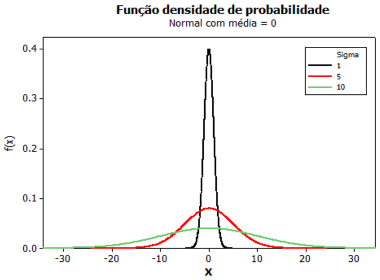
Características interessantes do gráfico da distribuição normal:

- Tem a forma de um sino.
- Tem uma assíntota horizontal: $f(x) = 0$.
- É simétrico em relação ao ponto de abscissa $x = \mu$.
- Como a curva é simétrica: $P(X > \mu) = P(X < \mu) = 0{,}5$

- $x = \mu$ é a abscissa do ponto de máximo absoluto da função e coincide com a mediana da distribuição.
- Os pontos de abscissas $x_1 = \mu - \sigma$ e $x_2 = \mu + \sigma$ são pontos de inflexão da função.



a) Médias diferentes e mesmo desvio padrão



b) Desvios padrões diferentes e mesma média

Figura 8.4 Distribuição normal com diferentes médias e variâncias

Cálculo de probabilidades:

- Calcular $P(x_1 < \mathrm{X} < x_2) = \int_{x_1}^{x_2} f(x)dx$ é muito difícil!
- Para facilitar o cálculo de probabilidades precisamos usar a variável **normal padronizada** ou **reduzida**,

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

que tem distribuição $N(0; 1)$.

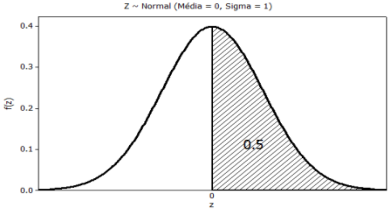
Exemplo: Para calcular $P(x_1 < X < x_2)$ devemos **padronizar** os limites de integração e usar a Tábua 1 para calcular a probabilidade

$$P(x_1 < X < x_2) = P(z_1 < Z < z_2)$$

Exemplo 4. Assumindo que o peso de frangos ao abate tem distribuição normal de média 1,80 kg e desvio padrão igual a 0,14 kg, calcular a probabilidade de encontrar um frango com peso

(a) superior a 1,80 kg
(b) inferior a 1,90 kg
(c) inferior a 1,70 kg
(d) entre 1,80 e 2,00 kg
(e) superior a 2,10 kg
(f) entre 1,60 e 1,70 kg.

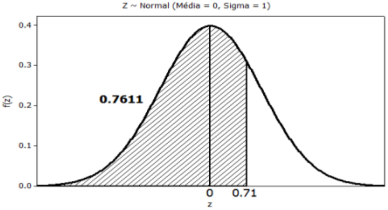
(a) $P(X > 1{,}80) = P\left(\frac{X-1{,}80}{0{,}14} \geq \frac{1{,}80-1{,}80}{0{,}14}\right)$
$= P(Z > 0) = 0{,}50$



(b) $P(X < 1{,}90) = P\left(Z < \frac{1{,}90-1{,}80}{0{,}14}\right) =$

$= P(Z < 0{,}71)$

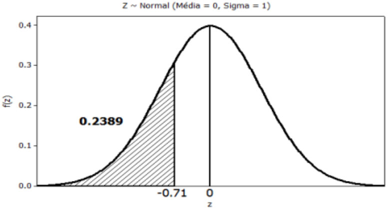
$= 0{,}50 + 0{,}2611 = 0{,}7611$



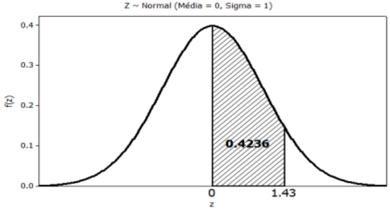
(c) $P(X < 1{,}70) = P\left(Z < \frac{1{,}70-1{,}80}{0{,}14}\right)$

$= P(Z < -0{,}71)$

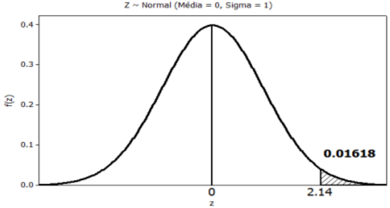
$= 0{,}50 - 0{,}2611 = 0{,}2389$



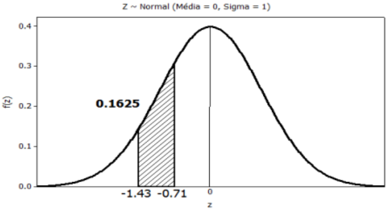
(d) $P(1{,}80 < X < 2{,}00)$
$= P(0 < Z < 1{,}43) = 0{,}4236$



(e) $P(X > 2{,}10) = P(Z > 2{,}14)$
$= 0{,}50 - 0{,}4838 = 0{,}0162$



(f) $P(1{,}60 < X < 1{,}70)$
$= P(-1{,}43 < Z < -0{,}71)$
$= 0{,}4236 - 0{,}2611 = 0{,}1625$



Exemplo 5. Admitindo que a altura dos alunos de Estatística tem distribuição normal com média $\mu = 1{,}67m$ e desvio padrão $\sigma = 0{,}08m$ calcule as seguintes probabilidades:

$a)\ P(X > 1{,}67)$ $b)\ P(X > 1{,}80)$ $c)\ P(1{,}60 < X < 1{,}75)$

$d)\ P(X < 1{,}50)$ $e)\ P(X > 1{,}85)$

8.4. APROXIMAÇÃO DA DISTRIBUIÇÃO BINOMIAL PELA NORMAL

Podemos usar a distribuição normal, que é associada a variáveis contínuas para calcular valores aproximados para as probabilidades da distribuição binomial, que está associada a variáveis aleatórias discretas.

Exemplo 6. Uma moeda é lançada 10 vezes. Seja X: número de caras, então a probabilidade exata de ocorrerem 7 caras ou mais é:

$$P(X \geq 7) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$
$$= 0,117 + 0{,}044 + 0{,}010 + 0{,}001 = 0{,}172.$$

Podemos aproximar a distribuição da v.a. X por uma distribuição normal de média

$$\mu = np = (10)(0{,}5) = 5$$

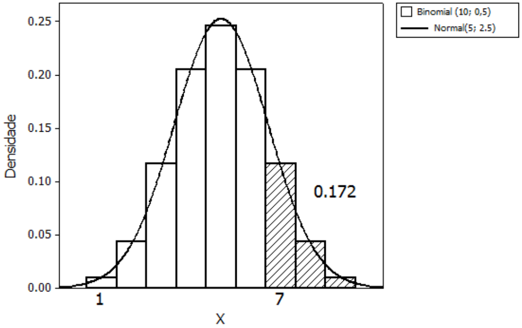
e variância

$$\sigma^2 = np(1-p) = (10)(0{,}5)(0{,}5) = 2{,}5$$

ou seja, vamos usar a nova variável W $\sim N(5;\, 2{,}5)$ e calcular a probabilidade:

$$P(X \geq 7) \cong P(W \geq 6{,}5) = P(Z \geq 0{,}949) = 0{,}1711$$

que é um valor bem próximo da probabilidade exata (calculada com a distribuição binomial)



k	$\mathrm{P}(\mathrm{X}=k)$
0	0,001
1	0,010
2	0,044
3	0,117
4	0,205
5	0,246
6	0,205
7	0,117
8	0,044
9	0,010
10	0,001

Figura 8.5. Aproximação da $Binomial(10;\, 0{,}5)$ pela $N(0{,}5;\, 2{,}5)$

Observações importantes:

- A aproximação de probabilidades da distribuição binomial pela distribuição normal será tanto melhor quanto maior for o valor de n e mais próximo de 0,5 for o valor de $p = P(sucesso)$.
- Quando o valor de n for grande ($n \longrightarrow \infty$) e o valor de p for muito pequeno ($p \rightarrow 0$), podemos obter melhores aproximações para probabilidades de uma $Binomial(n, p)$ utilizando uma distribuição Poisson($\lambda = np$).