Variáveis Aleatórias

Distribuições Discretas e Contínuas

Prof.: Wagner Pinheiro
wagner2235@gmail.com

Sumário

Distribuição Binomial

Considere um experimento aleatório consistindo em n tentativas independentes e a probabilidade de ocorrer sucesso em casa uma das n tentativas é sempre igual a p e de fracasso é q, onde $p + q = 1$. A probabilidade de sucesso e fracasso são as mesmas para cada tentativa.

Definição

Seja X o número de sucesso em n tentativas, então X pode assumir os valores $0, 1, 2, \ldots, n$. Nesta condição a $v.a$ X tem distribuição Binomial com parâmetro n e p, isto é, $X \sim B(n, p)$.

Considere que se $X \sim B(n, p)$, então seus parâmetros média e variância são definidos por:

a) Média de X é $E(X) = np$.
b) Variância de X é $V(X) = npq$, onde $q = 1 - p$.

- A função de probabilidade $(f.p.)$ da variável aleatória $X \sim B(n, p)$ é dada por

$$P(X = x) = C_n^x \; p^x \; q^{n-x}$$

$$= \frac{n!}{x!(n-x)!} \; p^x \; q^{n-x}$$

Distribuição Poisson

Este tipo de distribuição é útil para descrever as probabilidades do nº de ocorrências num campo ou intervalo contínuo (em geral tem ou espaço).

É também usada quando o número de observações é muito grande (ex.: $n \geq 50$) e a probabilidade de sucesso é muito pequena (ex.: $p \leq 0,1$) e o termo np permanece constante.

Distribuição Poisson

A média é igual a variância na distribuição Poisson, ou seja, $E(X) = V(X) = \lambda$.

- A função de probabilidade $(f.p.)$ da variável aleatória $X \sim P(\lambda)$ é dada por

$$P(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

onde,

X: v.a.d.

x: Número de sucessos ou fracassos

e: Base dos logarítmos neperianos ($e \approx 2,718$)

λ: média da distribuição

Exemplos

1) Se 20% dos parafusos produzidos por uma maquina são defeituosos, determine qual a probabilidade de, entre 4 parafusos selecionados ao acaso, no máximo 2 deles serem defeituosos. (R:0,9728)

2) O número de mortes por afogamento em fins de semana, numa cidade praiana, é de 2 para cada 50.000 habitantes. Qual a probabilidade de que:
   a) Em 200.000 habitantes ocorram 5 afogamentos? (R:0,091603)
   b) Em 112.500 habitantes ocorram pelo menos 3 afogamentos? (R:0,826422)

Introdução

A distribuição normal conhecida também como distribuição gaussiana é sem dúvida a mais importante distribuição contínua.

Sua importância se deve a vários fatores, entre eles podemos citar o **Teorema Central do Limite**, o qual é um resultado fundamental em aplicações práticas e teóricas, pois ele garante que mesmo que os dados não sejam distribuídos segundo uma normal a média dos dados converge para uma distribuição normal conforme o número de dados aumenta.

Além disso diversos estudos práticos tem como resultado uma distribuição normal. Podemos citar como exemplo a altura de uma determinada população em geral segue uma distribuição normal. Entre outras características físicas e sociais tem um comportamento gaussiano, ou seja, segue uma distribuição normal.

Definição

Uma variável aleatória contínua X tem uma distribuição normal (ou Gaussiana) se sua função densidade de probabilidade $f.d.p.$ for do tipo:

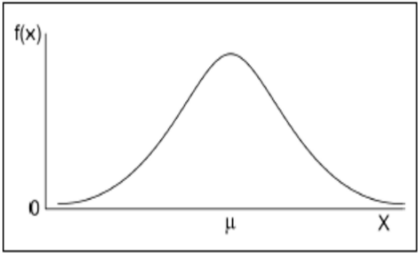
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

Onde $-\infty < \mu < \infty; \quad -\infty < x < \infty \quad$ e $\quad \sigma^2 > 0$.

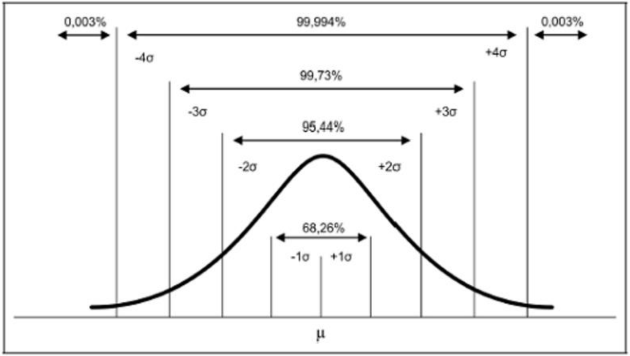
Usamos a notação $X \sim N(\mu, \sigma^2)$.

A variação natural de muitos processos, por exemplo, processos industriais é realmente aleatória. Embora as distribuições de muitos processos possam assumir uma variedade de formas, muitas variáveis observadas possuem uma distribuição de frequências que é, aproximadamente, uma distribuição de probabilidade Normal.

A distribuição é normal quando tem a forma de "sino":



Para achar a área sob a curva normal devemos conhecer dois valores numéricos, a média μ e o desvio padrão σ. A Figura a seguir mostra algumas áreas importantes:



Quando μ e σ são desconhecidos (caso mais comum), estes valores serão estimados por $\overline{X}$ e s, respectivamente, a partir da amostra.

Propriedades

P$_1$) O ponto de máximo de $f(x)$ é o ponto onde $E(X) = \mu$. Isto é, o parâmetro μ é a média da distribuição normal.

P$_2$) Os pontos de inflexão da função são: $X = \mu + \sigma$ e $X = \mu - \sigma$. E a variância da distribuição normal é o parâmetro σ^2

P$_3$) A curva é simética com relação a μ.

OBSERVAÇÃO: A integral da $f.d.p.$ normal, não pode ser avaliada pelo método tradicional (teorema fundamental do cálculo). Ela só pode ser calculada por métodos numéricos. E por isso ela é encontrada tabelada em qualquer livro texto de Probabilidade ou Estatística.

Variável normal padronizada (Z)

Para cada valor de μ e/ou σ temos uma curva de distribuição de probabilidade. Porém, para se calcular áreas específicas, faz-se uso de uma distribuição particular: a "distribuição normal padronizada", também chamada de Standartizada ou reduzida, o qual é a distribuição normal com $\mu = 0$ e $\sigma = 1$.

Para obter tal distribuição, isto é, quando se tem uma variável X com distribuição normal com média μ diferente de 0 (zero) e/ou desvio padrão σ diferente de 1 (um), devemos reduzi-la a uma variável Z, efetuando o seguinte cálculo

$$z = \frac{x - \mu}{\sigma}$$

Onde, Z: é o valor da variável normal padronizada Z; x: é o valor da $v.a$ X; μ: é a média da $v.a$ X e σ: é o desvio padrão da $v.a$ X

Média e Variância de Z

i) Média

$$E(Z) = E\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma}E(x-\mu) = \frac{1}{\sigma}E(x)-E(\mu) = \frac{1}{\sigma}(\mu-\mu) = 0$$

ii) Variância

$$\begin{aligned} V(Z) &= V\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma^2}V(x-\mu) \\ &= \frac{1}{\sigma^2}[V(x)+V(\mu)] = \frac{1}{\sigma^2}[\sigma^2+0] \\ &= \frac{1}{\sigma^2}\sigma^2 \\ &= 1 \end{aligned}$$

CONCLUSÃO:

Se $X \sim N(\mu, \sigma^2)$, e $z = \dfrac{x - \mu}{\sigma}$ então $Z \sim N(0,1)$, para quaisquer valores de μ e σ^2.

Portanto, será possível tabelar as probabilidades $P(X \leq x) = P(Z \leq z)$ em função dos valores padronizados z. E a $f.d.p$ de Z é dada por

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right]$$

e $-\infty < Z < \infty$.

Tabela da Distribuição Normal

Há vários tipos de tabelas que nos fornecem as probabilidades sob a curva normal. A tabela que vamos utilizar é aquela que fornece a probabilidade da variável Z assumir um valor entre zero e um particular valor que denotaremos por z_c, e sua probabilidade é calculada pela $f.d.p.$ dada por

$$P(0 \leq Z \leq z_c) = \int_0^{z_c} \varphi(z)dz = \frac{1}{\sqrt{2\pi}} \int_0^{z_c} \exp\left[-\frac{1}{2}z^2\right] dz$$

é a área hachurada sob a curva normal $[\varphi(z)]$, ou seja

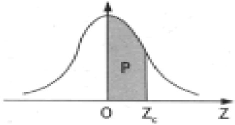
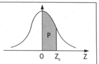


Tabela da Distribuição Normal

Tabela III — Distribuição Normal Padrão

$Z \sim N(0, 1)$

Corpo da tabela dá a probabilidade p, tal que $p = P(0 < Z < Z_c)$



parte inteira e primeira decimal de Z_c	Segunda decimal de Z_c										parte inteira e primeira decimal de Z_c
	0	1	2	3	4	5	6	7	8	9	
	p = 0										
0,0	00000	00399	00798	01197	01595	01994	02392	02790	03188	03586	0,0
0,1	03983	04380	04776	05172	05567	05962	06356	06749	07142	07535	0,1
0,2	07926	08317	08706	09095	09483	09871	10257	10642	11026	11409	0,2
0,3	11791	12172	12552	12930	13307	13683	14058	14431	14803	15173	0,3
0,4	15542	15910	16276	16640	17003	17364	17724	18082	18439	18793	0,4
0,5	19146	19497	19847	20194	20540	20884	21226	21566	21904	22240	0,5
0,6	22575	22907	23237	23565	23891	24215	24537	24857	25175	25490	0,6
0,7	25804	26115	26424	26730	27035	27337	27637	27935	28230	28524	0,7
0,8	28814	29103	29389	29673	29955	30234	30511	30785	31057	31327	0,8
0,9	31594	31859	32121	32381	32639	32894	33147	33398	33646	33891	0,9
1,0	34134	34375	34614	34850	35083	35314	35543	35769	35993	36214	1,0
1,1	36433	36650	36864	37076	37286	37493	37698	37900	38100	38298	1,1
1,2	38493	38686	38877	39065	39251	39435	39617	39796	39973	40147	1,2
1,3	40320	40490	40658	40824	40988	41149	41309	41466	41621	41774	1,3
1,4	41924	42073	42220	42364	42507	42647	42786	42922	43056	43189	1,4
1,5	43319	43448	43574	43699	43822	43943	44062	44179	44295	44408	1,5
1,6	44520	44630	44738	44845	44950	45053	45154	45254	45352	45449	1,6
1,7	45543	45637	45728	45818	45907	45994	46080	46164	46246	46327	1,7
1,8	46407	46485	46562	46638	46712	46784	46856	46926	46995	47062	1,8
1,9	47128	47193	47257	47320	47381	47441	47500	47558	47615	47670	1,9
2,0	47725	47778	47831	47882	47932	47982	48030	48077	48124	48169	2,0
2,1	48214	48257	48300	48341	48382	48422	48461	48500	48537	48574	2,1
2,2	48610	48645	48679	48713	48745	48778	48809	48840	48870	48899	2,2
2,3	48928	48956	48983	49010	49036	49061	49086	49111	49134	49158	2,3
2,4	49180	49202	49224	49245	49266	49286	49305	49324	49343	49361	2,4
2,5	49379	49396	49413	49430	49446	49461	49477	49492	49506	49520	2,5
2,6	49534	49547	49560	49573	49585	49598	49609	49621	49632	49643	2,6
2,7	49653	49664	49674	49683	49693	49702	49711	49720	49728	49736	2,7
2,8	49744	49752	49760	49767	49774	49781	49788	49795	49801	49807	2,8
2,9	49813	49819	49825	49831	49836	49841	49846	49851	49856	49861	2,9
3,0	49865	49869	49874	49878	49882	49886	49889	49893	49897	49900	3,0
3,1	49903	49906	49910	49913	49916	49918	49921	49924	49926	49929	3,1
3,2	49931	49934	49936	49938	49940	49942	49944	49946	49948	49950	3,2
3,3	49952	49953	49955	49957	49958	49960	49961	49962	49964	49965	3,3
3,4	49966	49968	49969	49970	49971	49972	49973	49974	49975	49976	3,4
3,5	49977	49978	49978	49979	49980	49981	49981	49982	49983	49983	3,5
3,6	49984	49985	49985	49986	49986	49987	49987	49988	49988	49989	3,6
3,7	49989	49990	49990	49990	49991	49991	49992	49992	49992	49992	3,7
3,8	49993	49993	49993	49994	49994	49994	49994	49995	49995	49995	3,8
3,9	49995	49995	49996	49996	49996	49996	49996	49996	49997	49997	3,9
4,0	49997	49997	49997	49997	49997	49997	49998	49998	49998	49998	4,0
4,5	49999	50000	50000	50000	50000	50000	50000	50000	50000	50000	4,5

Exemplo 1

Uma fábrica de carros sabe que os motores de sua fábrica têm duração normal com média de 150.000km e desvio padrão 5.000km. Qual a probabilidade de que um carro, escolhido ao acaso, dos fabricados por essa firma, tenha um motor que dure:

a) menos de 170.000km?

b) entre 140.000km e 165.000km?

c) Se a fábrica substitui o motor que apresenta duração inferior à garantia, qual deve ser esta garantia para que a porcentagem de motores substituídos seja inferior a 0,2%?

Exemplo 2

Um fabricante de baterias sabe, por experiência passada, que as baterias de sua fabricação tem vida média de 600 dias e desvio padrão de 100 dias, sendo que a duração tem aproximadamente distribuição normal. Oferece uma garantia de 312 dias, isto é, troca as baterias que apresentam falhas nesse período. Mensalmente são fabricadas 10.000 baterias. Quantas o fabricante deverá trocar pelo uso da garantia, mensalmente?

Gráfico das distribuições

http://www.vusoft.eu/apps/distributions/index.html

```
plot(dnorm, -3, 3, xlab = "valores de X", ylab = "densidade de probabilidade")
title("Distribuicão Normal\nX ~ N(100, 64)")
plot(function(x) dnorm(x, 100, 8), 60, 140, ylab = "f(x)")
plot(function(x) dnorm(x, 90, 8), 60, 140, add = T, col = 2)
plot(function(x) dnorm(x, 100, 15), 60, 140, add = T, col = 3)
legend(110, 0.05, c("N(100,64)", "N(90,64)", "N(100,225)"), fill = 1:3)
```