7600054 — Sistemas Complexos

Gonzalo Travieso

2020-06-03

Outline

Estatísticas

- Ao avaliar estatisticamente um fenômeno, consideramos que ele se constitui em uma variável aleatória. Dessa variável devemos extrair algumas amostras, $x_1, x_2, \ldots, x_n$.
- n é o tamanho da amostragem.
- A amostragem deve ser representativa (isto é, cada elemento da população de eventos amostrada tem que ter a mesma probabilidade de ser incluído na amostra).
- Em geral, não conhecemos as características da distribuição da variável aleatória, então precisamos avaliá-las através dos dados amostrados.

Média e desvio padrão

- A média m_1 pode ser estimada através da média amostral:

$$\bar{x} = \frac{\sum_{j=1}^{n} x_j}{n}.$$

- O desvio padrão amostral pode ser estimado como:

$$s = \sqrt{\frac{\sum_{j=1}^{n} (x_j - \bar{x})^2}{n-1}}.$$

- Note que o denominador é $n-1$, e não n. A explicação é que a média $\bar{x}$ foi calculada usando os n valores de x_j, e portanto entre os n valores $(x_j - \bar{x})$ apenas $n-1$ são independentes.
- No NumPy, para a função `numpy.std` calcular o desvio padrão amostral corretamente, devemos incluir o parâmetro opcional adicional `ddof=1`, como em `np.std(x, ddof=1)`.

Assimetria, curtose

- A assimetria é estimada por:

$$\frac{\sum_{j=1}^{n}\left(x_j-\bar{x}\right)^3}{ns^3}.$$

- A curtose em excesso é estimada por:

$$\frac{\sum_{j=1}^{n}\left(x_j-\bar{x}\right)^4}{ns^4}-3.$$

- Note que nesses casos não precisamos usar $n-1$, pois isso já foi considerado no cálculo de s, que aparece no denominador na potência apropriada.

Mediana

- A mediana $\tilde{x}$ pode ser estimada ordenando os valores e se n é impar, escolhendo o elemento central; se n é par, escolhe-se o valor médio entre os dois elementos centrais $(n-1)/2$ e $(n+1)/2$.
- Outra medida interessante é:

$$\frac{3(\bar{x}-\tilde{x})}{s},$$

que também serve para avaliar a assimetria da distribuição.

Variável binomial

- Seja uma sequência de variáveis aleatórias (X_j) com valores 0 ou 1 e

$$P(X_j = 1) = p, \quad P(X_j = 0) = 1 - p.$$

- Esse tipo de variável é denominada variável aleatória de Bernoulli.
- Agora considere a soma de n valores dessas variáveis. Essa soma terá uma distribuição dada por:

$$B_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k},$$

  onde k é o número de variáveis com valor 1.
- Uma variável aleatória com essa probabilidade é denominada uma variável aleatória binomial.

Características

- A média de uma variável de Bernoulli é

$$\langle X \rangle = p \cdot 1 + (1-p) \cdot 0 = p.$$

- A variância é

$$\langle X^2 \rangle - \langle X \rangle^2 = p \cdot 1^2 + (1-p) \cdot 0^2 - p^2 = p(1-p).$$

- Como uma variável binomial é a soma de n variáveis de Bernoulli, a média da binomial é

$$\langle B_{n,p} \rangle = np.$$

- Da mesma forma, a variância é

$$np(1-p).$$

Limite

- A assimetria vale
$$\frac{1-2p}{\sqrt{np(1-p)}}.$$
- A curtose em excesso vale
$$\frac{1-6p(1-p)}{np(1-p)}.$$
- Note que tanto a assimetria quanto a curtose tendem a 0 para n grande, o que indica que a binomial pode tender a uma gaussiana.
- Como uma variável binomial é a soma de diversas variáveis de Bernoulli, se o número de variáveis somadas for alto, podemos fazer uso da teorema central do limite e deduzir que, para n grande, a distribuição binomial tende para uma gaussiana com média np e variância $np(1-p)$:
$$B_{n,p} \xrightarrow{d} G(pn, np(1-p)).$$

Distribuição de Poisson

- Suponha que num sistema de tempo discreto temos um evento que pode ou não ocorrer em cada instante de tempo. A probabilidade do evento ocorrer em cada instante é p, e a ocorrência em instantes distintos é independente.
- Este é um denominado processo de Poisson.
- A distribuição de intervalos entre ocorrência dos eventos é dada por

$$\Pi_\lambda(k) = \frac{\lambda^k}{k!}e^{-\lambda},$$

  onde aqui k é o intervalo entre duas ocorrências do evento.
- Esta é denominada a distribuição de Poisson.

Características

- Tanto a média quando a variância da distribuição de Poisson Π_λ valem λ.
- A assimetria vale
$$\frac{1}{\sqrt{\lambda}}.$$
- A curtose em excesso vale
$$\frac{1}{\lambda}.$$

Binomial e Poisson

- Considere na distribuição binomial
$$B_{n,p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$
que fazemos n tender para infinito, mas de tal forma que np, que é a média, fique fixa $np = \alpha$, onde α é alguma constante.
- Primeiro notamos que, para n grande
$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \approx \frac{n^k}{k!}$$
e
$$(1-p)^{n-k} \approx (1-p)^n.$$
- Substituindo essas aproximações e $p = \frac{\alpha}{n}$ temos
$$B_{n,p} \approx \frac{n^k}{k!}\frac{\alpha^k}{n^k}\left(1 - \frac{\alpha}{n}\right)^n = \frac{\alpha^k}{k!}\left(1 - \frac{\alpha}{n}\right)^n.$$

Binomial e Poisson

- Sabemos que, para x pequeno,

$$e^x \approx 1 + x.$$

Portanto substituimos $1 - \frac{\alpha}{n}$ por $e^{-\frac{\alpha}{n}}$ e conseguimos

$$B_{n,p} \approx \frac{\alpha^k}{k!} e^{-\alpha},$$

isto é, no limite de n grande mantendo a média np, a distribuição binomial tende a uma distribuição de Poisson:

$$B_{n,p} \xrightarrow{d} \Pi_{np}.$$

Distribuição uniforme

- A distribuição uniforme no intervalo $[a, b]$ corresponde a uma distribuição com densidade de probabilidade dada por

$$u_{a,b}(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{c.c.} \end{cases} \tag{1}$$

- A média vale $\frac{a+b}{2}$.
- A variância vale $\frac{(b-a)^2}{12}$.
- A assimetria vale 0.
- A curtose em excesso vale $-\frac{6}{5}$.

Distribuição lognormal

- Uma função tem distribuição lognormal se o seu logaritmo tem distribuição normal (gaussiana). Isto é, se x é tem distribuição lognormal então $y = \ln x$ tem distribuição normal, ou se y tem distribuição normal, então $x = e^y$ tem distribuição lognormal.
- Designamos por L_{μ,σ^2} a distribuição lognormal que corresponde a uma distribuição gaussiana $G(\mu, \sigma^2)$.
- A distribuição lognormal surge quando uma variável aleatória é uma composição multiplicativa de diversas variáveis aleatórias independentes, pois

$$\ln \prod_{i=1}^{n} X_i = \sum_{i=1}^{n} \ln X_i.$$

Características

- A média vale
$$e^{\mu+\frac{\sigma^2}{2}}.$$
- A variância vale
$$e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right).$$
- A assimetria vale
$$\left(e^{\sigma^2}+2\right)\sqrt{e^{\sigma^2}-1}.$$
- A curtose em excesso vale:
$$e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-6.$$

Lei de potência

- Uma variável aleatória (positiva) é distribuída por uma lei de potência se sua densidade de probabilidade para $x > x_0 > 0$ é da forma:

  $$\Lambda_\gamma(x) \propto x^{-\gamma},$$

  onde $\gamma > 1$ é uma constante. Para $x < x_0$ devemos ter $P(x) = 0$.

- A constante de proporcionalidade é encontrada pela relação

  $$\int_{x_0}^{\infty} \Lambda_\gamma(x)\, dx = 1,$$

  e resulta em

  $$\Lambda_\gamma(x) = (\gamma - 1) x_0^{\gamma-1} x^{-\gamma}.$$

  As condição $\gamma > 1$ e $P(x) = 0, x < x_0$ garantem que a integral converge.

Características

- A média vale
$$\frac{\gamma - 1}{\gamma - 2} x_0,$$
apenas para $\gamma > 2$!
- A variância vale
$$\frac{\gamma - 1}{(\gamma - 2)^2 (\gamma - 3)} x_0^2,$$
para $\gamma > 3$!

Características (cont)

- A assimetria vale
$$\frac{2\gamma}{\gamma-4}\sqrt{\gamma\frac{\gamma-3}{\gamma-1}},$$
para $\gamma > 4$.
- A curtose em excesso é:
$$6\frac{\gamma^3-2\gamma^2-5\gamma+4}{(\gamma-1)(\gamma-4)(\gamma-5)},$$
para $\gamma > 5$.

Invariância a escala

- Uma característica das distribuições em lei de potência é que elas são invariantes a escala.
- Isso significa que uma mudança na escala da variável apenas ocasiona uma mudança na escala das probabilidades, sem mudança na estrutura funcional.
- Seja $f(x) = \alpha x^{-\gamma}$ uma distribuição. Se mudamos a escala de x para $y = \beta x$ temos:

$$f(\tfrac{y}{\beta}) = \alpha(\tfrac{y}{\beta})^{-\gamma} = \beta^{\gamma}\alpha y^{-\gamma} = \beta^{\gamma} f(y).$$

- Por essa razão, essas distribuições são chamadas livres de escala.

Aplicações

- Diversos sistemas apresentam distribuições livre de escala no ponto crítico de transição de fase, como vimos em exemplos das aulas anteriores.
- Entretanto, nos sistemas reais, a distribuição raramente é puramente livre de escala, mas apresenta uma região mais ou menos extensa de comportamento livre de escala, com um corte inferior determinado por características do sistema e um corte superior muitas vezes dominado por efeitos de tamanho finito (isto é, como o sistema não é infinito, o comportamento livre de escala tem que ser limitado a partir de algum ponto).
- Esses fatores precisam ser levados em consideração quando lidamos com possíveis fenômenos com distribuição livre de escala em sistemas reais ou simulações.