Resumo: Fundamentos Estatísticos para Ciência dos Dados

Ricardo Pagoto Marinho

12 de abril de 2018

1 Regra de Bayes

Inverte as probabilidades de interesse.

Exemplo:
$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$

2 Função distribuição acumulada de probabilidade

$\mathbb{F}(x)$ definida $\forall\ x\ \in\ \mathbb{R}$ é dada por:

$$\begin{aligned}\mathbb{F}:\mathbb{R} \to&\ [0,1]\\ x \to&\ \mathbb{F}(x) = \mathbb{P}(X \leq\ x)\end{aligned}$$

Caso geral de $\mathbb{F}(x)$

$$\mathbb{F}(x) = \mathbb{P}(X \leq\ x) = \sum_{x_i \leq\ x} p(x_i)$$

3 Esperança matemática ($\mathbb{E}(X)$)

3.1 V.A. Discreta

O valor esperado de uma V.A. discreta é a soma de seus valores possíveis ponderados pelas suas probabilidades respectivas.

$$\mathbb{E}(X) = \sum_i\ x_i p(x_i)$$

Suponha que numa amostra grande de instâncias, x_i apareceu N_i vezes. A probabilidade de x_i ocorrer na amostra é sua frequência relativa, *i.e.*:

$$p(x_i) = \mathbb{P}(X = x_i) \approx\ \frac{N_i}{N}$$

Logo:

$$\mathbb{E}(X) = \sum_i x_i p(x_i) \approx \sum_i x_i \frac{N_i}{N}$$

Se a amostra for grande, o número teórico $\mathbb{E}(X)$ é aproximadamente igual à média aritmética dos N elementos da amostra.

3.2 V.A. contínuas

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} y f(y) dy$$

3.3 Linearidade da esperança

Caso geral: Y=a+bX, onde a e b são constantes. Então $\mathbb{E}(X)$ e $\mathbb{E}(Y)$ estão relacionadas:

$$\mathbb{E}(Y) = \mathbb{E}(a + bX) = a + b\mathbb{E}(X)$$

Exemplo:

Medimos uma temperatura aleatória **C** em graus Celsius. Suponha que $\mathbb{E}(C) = 28$ graus. Seja F a V.A. que mede a temperatura em graus Fahrenheit. Temos que C e F estão relacionadas por: $F = 32 + \frac{9}{5}C$. Pelo caso geral da linearidade, a=32 e b=$\frac{9}{5}$. Logo

$$\begin{aligned} \mathbb{E}(F) &= \mathbb{E}(32 + \tfrac{9}{5}C) \\ &= 32 + \tfrac{9}{5}\mathbb{E}(C) \\ &= 32 + \tfrac{9}{5} \times 28 \\ &= 82.4 \end{aligned}$$

Caso duas variáveis aleatórias sejam DISJUNTAS:

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

Se independentes:

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$$

4 Variância

$$\mathbb{V}(X) = \mathbb{E}((X-\mu)^2)$$

Outra fórmula:

$$\begin{aligned}\mathbb{V}(X) = &\quad \mathbb{E}((X-\mu)^2)\\ = &\quad \mathbb{E}(X^2) - (\mu)^2\\ = &\quad \mathbb{E}(X^2) - (\mathbb{E}(X))^2\end{aligned}$$

Seja X uma v.a. com $\mu_x = \mathbb{E}(X)$ e $\sigma^2 = \mathbb{V}(X)$. Se Y=a+bX, então $\mu_y = \mathbb{E}(Y) = a + b\mu_x$ e

$$\sigma_y^2 = \mathbb{V}(Y) = \mathbb{V}(a+bX) = b^2\mathbb{V}(X) = b^2\sigma_x^2$$

Em termos de DP das v.a.'s:

$$DP_y = |b|DP_X$$

Se as v.a.'s são independentes, temos:

$$\mathbb{V}(X+Y) = \mathbb{V}(X) + \mathbb{V}(Y)$$

4.1 Caso discreto

Como $\mathbb{E}(g(X)) = \sum_i g(x_i)\mathbb{P}(X = x_i)$, e tomando $g(X) = (X-\mu)^2$, então:

$$\mathbb{V}(X) = E((X-\mu)^2) = \sum_i (x_i - \mu)^2 \mathbb{P}(X = x_i)$$

4.2 Caso contínuo

$$\mathbb{V}(X) = \mathbb{E}((X-\mu)^2) = \int (x-\mu)^2 f(x)dx$$

5 Distribuição de Bernoulli

É a distribuição mais simples: dois resultados possíveis $X(\omega) \in \{0,1\} \, \forall \, \omega \in \Omega$

Duas probabilidades são definidas:

- p(0)=$\mathbb{P}(X = 0) = \mathbb{P}(\omega \in \Omega : X(\omega) = 0)$

- p(1)=$\mathbb{P}(X=1) = \mathbb{P}(\omega \in \Omega : X(\omega)=1)$

$p(0)+p(1)=1 \rightarrow \ p(1)=1-p(0)$
É comum escrever $p(1)=p$ e $p(0)=q$. Daí, $\mathbb{E}(X)=1\times p+0\times(1-p)=p$
Como a média aritmética dessa distribuição é a proporção de 1's na amostra:

$$\mathbb{E}(X) \approx \hat{p} = \frac{1}{N}\sum_i \ x_i$$

6 Distribuição Binomial

Frequentemente utilizada quando um número máximo possível grande de n de repetições e θ muito pequeno.

n repetições independentes de um experimento de Bernoulli: sucesso ou fracasso. Probabilidade de sucesso é igual a $\theta \in [0,1]$

A V.A. X conta o número total de sucessos: $X \sim \ Bin(n,\theta)$. Os valores possíveis são: $0,1,2,\cdots,n$ e suas probabilidades, respectivamente são: $(1-\theta)^n, \ n\theta(1-\theta),\cdots,\theta^n$

Exemplo: n lançamentos de uma moeda não viciada.

$$Cara \rightarrow \ C$$
$$Coroa \rightarrow \ \tilde{C}$$

$P(X=0)=(1-\theta)^n$
$[X=0]=\{\omega\in\Omega : X(\omega)=0\}=\{\omega\in\Omega : \omega\in\{(\tilde{C},\tilde{C},\tilde{C},\cdots,\tilde{C})\}\}=P(\tilde{C}\ no\ 1^o)\times P(\tilde{C}\ no\ 2^o)\times\cdots=(1-\theta)\times(1-\theta)\cdots=(1-\theta)^n$

- Fórmula geral:

$$\mathbb{P}(Y=k)=\frac{n!}{k!(n-k)!}\theta^k(1-\theta)^{n-k}$$

- $\mathbb{E}(Y)=n\theta$ e DP=$\sqrt{\mathbb{V}(Y)}=\sqrt{n\theta(1-\theta)}$

7 Distribuição Multinomial

Mais de duas categorias de resultados nos experimentos, diferente da Binomial que são só duas (1 ou 0). Ao fim de n lançamentos, teremos um vetor aleatório multinomial que conta quantas vezes cada categoria apareceu no experimento. Temos k categorias:

$$(N_1,N_2,\cdots,N_k)\sim \ \mathbb{M}(n;\theta_1,\cdots,\theta_k)$$

Sendo que $\theta_1,\cdots,\theta_k$ são as probabilidades de cada categoria.

Exemplo: lançamento de um dado. $k = 6$

$$\begin{aligned} N_1 &= n^o \ de \ lanamentos \ na \ categoria \ 1 \\ N_2 &= n^o \ de \ lanamentos \ na \ categoria \ 2 \\ N_3 &= n^o \ de \ lanamentos \ na \ categoria \ 3 \\ &\vdots \\ N_6 &= n^o \ de \ lanamentos \ na \ categoria \ 6 \end{aligned}$$

$$(N_1, N_2, \cdots, N_6) \sim \ \mathbb{M}(n; \theta_1, \cdots, \theta_6)$$

Podemos escrever a Binomial como uma Multinomial de duas categorias: sucesso e fracasso. X é o número de sucessos em n repetições.

$$(X, n - X) \sim \ \mathbb{M}(n; \theta, 1 - \theta)$$

A probabilidade de ocorrer uma configuração do vetor aleatório é:

$$\mathbb{P}(\mathbf{N} = (n_1, n_2, \cdots, n_k)) = \frac{n!}{n_1! n_2! \cdots n_k!} \theta_1^{n_1} \theta_2^{n_2} \cdots \theta_k^{n_k} \tag{1}$$

Exemplo: 8 lançamentos de um dado. A probabilidade de:

$$\mathbb{P}(\mathbf{N} = (2, 0, 2, 1, 0, 3))$$

Existem várias configurações de ω as quais 8 lançamentos levam ao resultado acima. Uma é $\omega = (3, 1, 6, 6, 1, 4, 6, 3)$. Logo:

$$\mathbf{N}(\omega) = (N_1(\omega), N_2(\omega), \cdots, N_6(\omega)) = (2, 0, 2, 1, 0, 3)$$

A probabilidade de sair essa configuração, levando em conta que os lançamentos são independentes é:

$$\begin{aligned} \mathbb{P}(\omega = (3, 1, 6, 6, 1, 4, 6, 3)) &= \mathbb{P}(sair \ 3 \ no \ 1^o \ E \ sair \ 1 \ no \ 2^o \ E \cdots \ sair \ 3 \ no \ 8^o \\ &= \mathbb{P}(sair \ 3 \ no \ 1^o) \mathbb{P}(sair \ 1 \ no \ 2^o) \cdots \mathbb{P}(sair \ 3 \ no \ 8^o) \\ &= \theta_3 \ \theta_1 \ \theta_6 \ \theta_6 \ \theta_1 \ \theta_4 \ \theta_6 \ \theta_3 \\ &= \theta_1^2 \ \theta_2^0 \ \theta_3^2 \ \theta_4^1 \ \theta_5^0 \ \theta_6^3 \end{aligned}$$

Se a sequência ω tiver n lançamentos:

$$\begin{aligned} n_1 &\quad aparies \ da \ face 1 \\ n_2 &\quad aparies \ da \ face 2 \\ &\vdots \\ n_6 &\quad aparies \ da \ face 6 \end{aligned}$$

Teremos:

$$\mathbb{P}(\omega) = \theta_1^{n_1}\ \theta_2^{n_2}\ \theta_3^{n_3}\ \theta_4^{n_4}\ \theta_5^{n_5}\ \theta_6^{n_6}$$

Dessa forma, seja A o evento formado por todos os ω tais que $\mathbb{P}(\mathbf{N} = (2,0,2,1,0,3))$

$\mathbb{P}(\mathbf{N} = (2,0,2,1,0,3)) = \mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega) = c \times \theta_1^2\ \theta_2^0\ \theta_3^2\ \theta_4^1\ \theta_5^0\ \theta_6^3$ Onde c é o número de sequências de tamanho 8 tais que $\mathbb{P}(\mathbf{N} = (2,0,2,1,0,3))$ c é o número de permutações do vetor $\omega = (3,1,6,6,1,4,6,3)$. Generalizando para k categorias, temos:

$$\mathbf{N} = (N_1, N_2, \cdots, N_k) \sim \mathbb{M}(n; \theta_1, \cdots, \theta_n)$$

Então, chegamos na Equação **??**.

8 Distribuição de Poisson

Frequentemente utilizada em situações onde o número de ocorrências não tem um limite claro para o limite.

$\mathbb{P}(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}$

$\mathbb{E}(Y) = \lambda$

9 Distribuição geométrica

Y é o número de **fracassos** em uma sequência de ensaios independentes de Bernoulli até que um sucesso (probabilidade θ) seja observado. Logo, Y=0 significa que no primeiro ensaio houve um sucesso e $\mathbb{P}(Y = 0) = \mathbb{P}(S) = \theta$. Y=1 significa que o primeiro ensaio foi um fracasso e o segundo sucesso: $\mathbb{P}(Y = 1) = \mathbb{P}(FS) = (1 - \theta)\theta$.

De forma geral, $\mathbb{P}(Y = k) = (1 - \theta)^k \theta$

Para uma geométrica com probabilidade de sucesso θ:

$$\mathbb{E}(Y) = \frac{1}{\theta}$$

Uma distribuição geométrica com θ alto significa que a probabilidade de sucesso é grande. Logo, a função de distribuição de probabilidade se concentrará mais nos números iniciais.

10 Distribuição de Zipf ou de Pareto

$X \in 1, 2, 3, \cdots, N$, sendo que N pode ser infinito.

$$\mathbb{P}(X = k) = \frac{C}{k^(1 + \alpha)},\ com\ \alpha > 0$$

C é uma constante que garante que as probabilidades somem 1:

$$\begin{aligned} 1 = & \quad \mathbb{P}(X=1)+\mathbb{P}(X=2)+\mathbb{P}(X=3)+\dots \\ = & \quad C(\tfrac{1}{1^{1+\alpha}}+\tfrac{1}{2^{1+\alpha}}+\tfrac{1}{3^{1+\alpha}}+\dots \\ = & \quad C\textstyle\sum_{k=1}^{N}\tfrac{1}{k^{1+\alpha}} \end{aligned}$$

O que implica que:

$$C = \frac{1}{\sum_{k=1}^{N}\frac{1}{k^{1+\alpha}}}$$

IMPORTANTE:

$$\mathbb{P}(X=k) \propto \frac{1}{k^{1+\alpha}}$$

i.e., inversamente proporcional a uma potência de k.

Com $\alpha = 1.0$, a probabilidade de 0 é maior (≈ 0.6). Com $\alpha = 0.5$, a probabilidade de 0 diminui.

Como $\mathbb{P}(Y=k) \propto \frac{1}{k}$:

$$\mathbb{P}(Y=1) \propto 1$$
$$\mathbb{P}(Y=2) \propto \frac{1}{2}$$
$$\mathbb{P}(Y=2) \propto \frac{1}{3}, etc$$

11 Desigualdade de Tchebyshev

$$\mathbb{P}(|Y-\mu| > k\sigma) \leq \frac{1}{k^2}$$