Esperanças e Probabilidades Condicionais

Ricardo Ehlers
ehlers@icmc.usp.br

Departamento de Matemática Aplicada e Estatística
Universidade de São Paulo

Caso Discreto

Probabilidade Condicional

Sejam X e Y variáveis aleatórias discretas. A função massa de probabilidade condicional de X dado que $Y = y$ é dada por,

$$\begin{aligned} p_{X|Y}(x|y) &= P(X = x|Y = y) \\ &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{p(x, y)}{p_Y(y)} \end{aligned}$$

para todos os valores de y tais que $P(Y = y) > 0$.

Equivalentemente,

$$p_{X|Y}(x|y) = \frac{p_{Y|X}(y|x)\ p_X(x)}{\sum_x p_{Y|X}(y|x) p_X(x)}$$

$$\propto\ p_{Y|X}(y|x)\ p_X(x).$$

Função de Distribuição Condicional

$$
\begin{aligned}
F_{X|Y}(x|y) &= P(X \leq x|Y = y) \\
&= \sum_{a \leq x} p_{X|Y}(a|y)
\end{aligned}
$$

definida para todos os valores de y tais que $P(Y = y) > 0$.

Esperança Condicional

$$E(h(X)|Y = y) = \sum_x h(x)P(X = x|Y = y).$$

Estude os exemplos da Seção 3.2 em Sheldon Ross.

Exemplo. Sejam X_1 e X_2 independentes tais que,

$$\begin{aligned} X_1 &\sim \text{Binomial}(n_1, p) \\ X_2 &\sim \text{Binomial}(n_2, p). \end{aligned}$$

Deseja-se obter a distribuição de X_1 dado que $X_1 + X_2 = m$.

$$\begin{aligned} P(X_1 = k | X_1 + X_2 = m) &= \frac{P(X_1 = k, X_1 + X_2 = m)}{P(X_1 + X_2 = m)} \\ &= \frac{P(X_1 = k, X_2 = m - k)}{P(X_1 + X_2 = m)} \\ &= \frac{P(X_1 = k)\ P(X_2 = m - k)}{P(X_1 + X_2 = m)} \end{aligned}$$

Como X_1 e X_2 são independentes e,

$$X_1 + X_2 \sim \text{Binomial}(n_1 + n_2, p),$$

segue que,

$$P(X_1 = k | X_1 + X_2 = m) =$$

$$\frac{\binom{n_1}{k} p^k (1-p)^{n_1-k} \binom{n_2}{m-k} p^{m-k} (1-p)^{n_2-m+k}}{\binom{n_1+n_2}{m} p^m (1-p)^{n_1+n_2-m}} =$$

$$\frac{\binom{n_1}{k}\binom{n_2}{m-k}}{\binom{n_1+n_2}{m}}$$

Conclusão: $(X_1 | X_1 + X_2 = m)$ tem distribuição *hipergeometrica*.

Caso Contínuo

Função de Densidade Condicional

Sejam X e Y variáveis aleatórias contínuas com função de densidade conjunta $f(x, y)$. A função densidade de probabilidade condicional de X dado que $Y = y$ é dada por,

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

definida para todos os valores de y tais que $f_Y(y) > 0$.

Equivalentemente,

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x)\ f_X(x)}{\int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) dx}$$

$$\propto \ f_{Y|X}(y|x)\ f_X(x)$$

Função de distribuição condicional

$$F(x|y) = \int_{-\infty}^{x} f(u|y) du.$$

Esperança Condicional

$$E(h(X)|Y = y) = \int_{-\infty}^{\infty} h(x) f_{X|Y}(x|y) dx.$$

Estude os exemplos da Seção 3.3 em Sheldon Ross.

Exemplo. Sejam X e Y variáveis aleatórias contínuas tais que,

$$\begin{aligned} X|Y=y &\sim N\left(\mu, \frac{\sigma^2}{y}\right) \\ Y &\sim Gama\left(\frac{\nu}{2}, \frac{\nu}{2}\right). \end{aligned}$$

Deseja-se obter a distribuição marginal de X, ou seja,

$$f(x) = \int_0^\infty f(x|y)f(y)dy.$$

Definição

Uma variável aleatória X tem distribuição Gama com parâmetros a e b, se sua função de densidade é dada por,

$$f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx}, \quad x > 0,$$

para $a > 0$ e $b > 0$.

Então,

$$\begin{aligned} f(x,y) &= f(x|y)f(y) \\ &= \frac{1}{\sqrt{2\pi(\sigma^2/y)}} \exp\left(-\frac{y}{2\sigma^2}(x-\mu)^2\right) \times \\ & \quad \frac{(\nu/2)^{\nu/2}}{\Gamma(\nu/2)} y^{\nu/2-1} \exp(-y\nu/2) \end{aligned}$$

Verifique que a densidade marginal de X é,

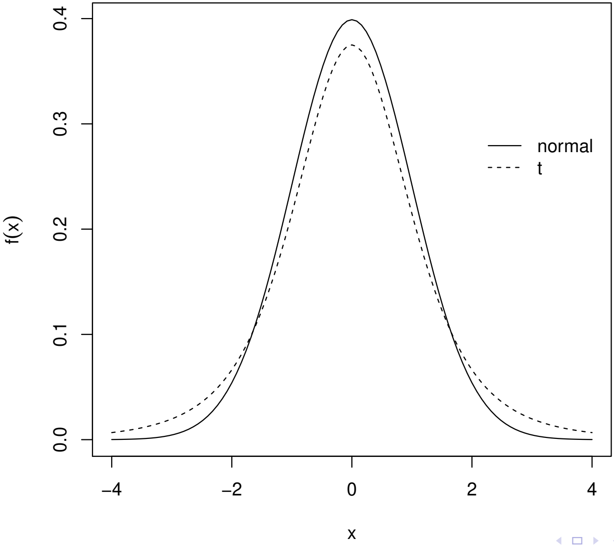
$$\begin{aligned} f(x) &= \int_0^\infty f(x|y)f(y)dy \\ &= \frac{\Gamma((\nu+1)/2)}{\sigma\Gamma(\nu/2)(\pi\nu)^{1/2}}\left[1+\frac{(x-\mu)^2}{\nu\sigma^2}\right]^{-(\nu+1)/2}. \end{aligned}$$

Dizemos que X tem distribuição t de Student com parâmetros μ, σ e ν,

$$X \sim t(\mu, \sigma, \nu).$$

Este é um resultado muito importante em aplicações pois esta distribuição tem caudas mais pesadas do que a distribuição normal.

Densidades N(0,1) e t de Student com $\mu = 0$, $\sigma = 1$ e $\nu = 4$



Esperança e Variância por Condicionamento

Teorema

Seja $E(X|Y)$ uma função da variável aleatória Y cujo valor em $Y = y$ é $E(X|Y = y)$. Então, para variáveis aleatórias quaisquer X e Y,

$$E(X) = E[E(X|Y)].$$

Se Y for discreta,

$$E[X] = \sum_y E[X|Y = y]P(Y = y).$$

Se Y for contínua,

$$E(X) = \int_{-\infty}^{\infty} E[X|Y = y]f_Y(y)dy.$$

Definição

Seja a variável aleatória,

$$Z = \sum_{i=1}^{N} X_i$$

a soma de um número aleatório N de variáveis aleatórias independentes e identicamente distribuidas e independentes de N. Então Z é uma *variável aleatória composta*.

Calcule $E(Z)$ por condicionamento.

Exemplo. Seja Z uma variável aleatória composta sendo,

$$Z = \sum_{i=1}^{N} X_i,$$

com $E(X_i) = E(X)$.

Por condicionamento temos que,

$$\begin{aligned} E(Z) &= E[E(Z|N=n)] \\ &= E\left[E\left(\sum_{i=1}^{n} X_i \middle| N=n\right)\right] \\ &= E\left[E\left(\sum_{i=1}^{n} X_i\right)\right]. \end{aligned}$$

Portanto, como $E(\sum_{i=1}^{n} X_i) = nE(X)$,

$$E(Z) = E[E(Z|N)] = E[N\ E(X)] = E(N)E(X).$$

Proposição

Seja $E(X|Y)$ uma função da variável aleatória Y cujo valor em $Y = y$ é $E(X|Y = y)$. Então, para variáveis aleatórias quaisquer X e Y,

$$Var(X) \quad = \quad E[Var(X|Y)] + Var[E(X|Y)].$$

Use este resultado para calcular a variância de uma variável aleatória composta.

Estude os exemplos da Seção 3.4 em Sheldon Ross.

Probabilidades por Condicionamento

Seja A um evento arbitrário e deseja-se calcular $P(A)$. Defina a variável aleatória X tal que,

$$X = 1 \text{ se } A \text{ ocorre, e } X = 0 \text{ se } A \text{ não ocorre.}$$

Então,

$$\begin{aligned} E[X] &= P(A), \\ E[X|Y=y] &= P(A|Y=y), \text{ para qualquer } Y. \end{aligned}$$

e portanto, $P(A) = E[P(A|Y=y)]$.

Se Y for discreta,

$$P(A) = \sum_y P(A|Y = y)P(Y = y).$$

Se Y for continua,

$$P(A) = \int_{-\infty}^{\infty} P(A|Y = y)f_Y(y)dy.$$

Estude os exemplos da Seção 3.5 em Sheldon Ross.

Variáveis aleatórias n-dimensionais

Seja o vetor aleatório $\mathbf{X} = (X_1, \ldots, X_n)'$. As definições univariadas de valor esperado e variância se estendem naturalmente ao caso multivariado, bem como suas propriedades.

- Vetor de esperanças,

$$E(\mathbf{X}) = \boldsymbol{\mu} = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

- Matriz de variâncias e covariâncias $n \times n$,

$$Var(\mathbf{X}) = E(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \ldots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \ldots & \sigma_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \ldots & \sigma_{nn} \end{bmatrix}$$

- Na matriz de variâncias e covariâncias cada elemento dado por,

$$\sigma_{ij} = Cov(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j) = E(X_i X_j) - \mu_i \mu_j.$$

- Há então n médias e $n(n+1)/2$ variâncias e covariâncias.

Densidades marginais e condicionais

Se $\mathbf{X} = (X_1, \ldots, X_n)'$ é um vetor aleatório contínuo,

$$f(x_i) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x})\ dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n.$$

Particionando o vetor $\mathbf{X}$ em 2 subvetores, $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ segue que,

$$\begin{aligned} f(\mathbf{x}_1) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x})\ d\mathbf{x}_2 \\ f(\mathbf{x}_2) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x})\ d\mathbf{x}_1 \\ f(\mathbf{x}_1 | \mathbf{x}_2) &= \frac{f(\mathbf{x})}{f(\mathbf{x}_2)} \\ f(\mathbf{x}_2 | \mathbf{x}_1) &= \frac{f(\mathbf{x})}{f(\mathbf{x}_1)} \end{aligned}$$

Exemplo. Seja o vetor aleatório $\mathbf{X} = (X_1, \ldots, X_n)'$. Algumas densidades condicionais são,

$$\begin{aligned}
f(x_1|x_2, \ldots, x_n) &= \frac{f(\mathbf{x})}{f(x_2, \ldots, x_n)} \\
f(x_n|x_1, \ldots, x_{n-1}) &= \frac{f(\mathbf{x})}{f(x_1, \ldots, x_{n-1})} \\
f(x_1, x_n|x_2, \ldots, x_{n-1}) &= \frac{f(\mathbf{x})}{f(x_2, \ldots, x_{n-1})}
\end{aligned}$$

Exemplo. Seja o vetor aleatório contínuo $\mathbf{X} = (X_1, X_2, X_3)'$. A função de densidade conjunta de $\mathbf{X}$ pode ser escrita como,

$$
\begin{aligned}
f(\mathbf{x}) &= f(x_1|x_2, x_3)f(x_2, x_3) \\
&= f(x_1|x_2, x_3)f(x_2|x_3)f(x_3)
\end{aligned}
$$

ou,

$$
\begin{aligned}
f(\mathbf{x}) &= f(x_3|x_2, x_1)f(x_2, x_1) \\
&= f(x_3|x_2, x_1)f(x_2|x_1)f(x_1).
\end{aligned}
$$

Exemplo. Seja o vetor aleatório $\mathbf{X} = (X_1, \ldots, X_n)'$. Uma possivel forma de escrever a densidade conjunta é,

$$
\begin{aligned}
f(\mathbf{x}) &= f(x_n|x_{n-1}, \ldots, x_1)\ f(x_{n-1}, \ldots, x_1) \\
&= f(x_n|x_{n-1}, \ldots, x_1)\ f(x_{n-1}|x_{n-2}, \ldots, x_1)\ f(x_{n-2}, \ldots, x_1) \\
&\ \vdots \\
&= f(x_n|x_{n-1}, \ldots, x_1)\ f(x_{n-1}|x_{n-2}, \ldots, x_1) \cdots f(x_2|x_1)\ f(x_1) \\
&= f(x_1) \prod_{t=2}^{n} f(x_t|x_{t-1}, \ldots, x_1)
\end{aligned}
$$

Exemplo. No exemplo anterior suponha que a distribuição de X_t depende somente de X_{t-1}. Ou seja, dado X_{t-1}, X_t é independente de $X_{t-2}, \ldots, X_1$. Então,

$$
\begin{aligned}
f(\mathbf{x}) &= f(x_n|x_{n-1})\; f(x_{n-1}|x_{n-2}) \cdots f(x_2|x_1)\; f(x_1) \\
&= f(x_1)\; \prod_{t=2}^{n} f(x_t|x_{t-1}).
\end{aligned}
$$

Então, as distribuições condicionais de $X_t|X_{t-1}$, $t = 2, \ldots, n$ e de X_1 definem a distribuição conjunta.

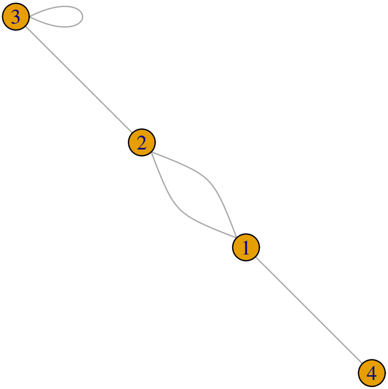
Grafos aleatórios

Definição

Um grafo consiste de um conjunto V de elementos chamados *nós* ou *vértices* e um conjunto A de pares de elementos de V chamados *arcos*.

Graficamente círculos representam nós e linhas representam arcos.

Exemplo. Grafo conectado com $V = \{1, 2, 3, 4\}$ e $A = \{(1, 2), (1, 4), (2, 3), (1, 2), (3, 3)\}$.



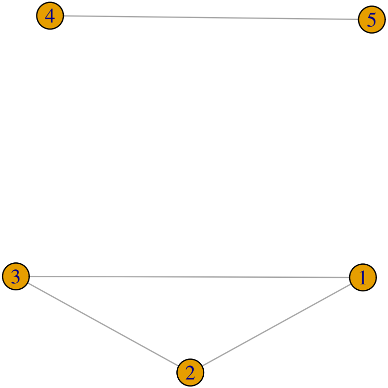
Definição

Dizemos que existe uma *trajetória* de i e j, $i \neq j$, se existir uma sequência de nós $i, i_1, \ldots, i_k, j$ tal que $(i, i_1), (i_1, i_2), \ldots, (i_k, j)$ são arcos.

Definição

Se existir uma trajetória entre todos os pares de nós o grafo é *conectado*.

Exemplo. Grafo não conectado com $V = \{1, 2, 3, 4, 5\}$ e $A = \{(1, 2), (1, 3), (2, 3), (4, 5)\}$.



Grafo Aleatório

Seja um grafo com $V = \{1, \ldots, n\}$ e $A = \{(i, X(i)), i = 1, \ldots, n\}$ sendo $X(i)$ variáveis aleatórias independentes com,

$$P(X(i) = j) = \frac{1}{n},\ j = 1, \ldots, n.$$

Este é chamado um *grafo aleatório uniforme*. Para cada nó i seleciona-se outro nó aleatoriamente com a mesma probabilidade e conecta-se os dois nós.

- O número de pares de nós em um grafo com $V = \{1, \ldots, n\}$ é o número de pares (i, j), $i \neq j$, que podem ser selecionados sendo (j, i) considerado o mesmo par.
- Ou seja, existem $\binom{n}{2}$ pares distintos de nós.

Ensaios de Bernoulli

Seja uma variável aleatória X que representa o número de sucessos em n (fixo) ensaios de Bernoulli independentes com $P(\text{sucesso}) = p$ constante e desconhecida.

Assume-se que p tem distribuição uniforme em $(0, 1)$.

Assim,

$$\begin{aligned} X|p &\sim \text{Binomial}(n, p) \\ p &\sim U(0, 1) \\ P(X = k|p) &= \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, \ldots, n, \\ f(p) &= 1,\; p \in (0, 1). \end{aligned}$$

Qual a probabilidade de se obter k sucessos independente do valor de p?

Ou seja, deseja-se calcular,

$$
\begin{aligned}
P(X = k) &= \int_0^1 P(X = k|p)\ f(p)dp \\
&= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp.
\end{aligned}
$$

Função Beta

Define-se a função Beta como

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \; a > 0, \; b > 0.$$

Em particular, se a e b são inteiros positivos,

$$\Gamma(a) = (a-1)!, \quad \Gamma(b) = (b-1)! \quad \text{e} \quad \Gamma(a+b) = (a+b-1)!$$

Portanto, temos que,

$$\int_0^1 p^k(1-p)^{n-k}dp = \frac{\Gamma(k+1)\Gamma(n-k+1)}{\Gamma(n+2)} = \frac{k!\,(n-k)!}{(n+1)!},$$

e finalmente,

$$P(X=k) = \binom{n}{k}\frac{k!\,(n-k)!}{(n+1)!} = \frac{1}{n+1}, k = 0, 1, \ldots, n.$$

ou seja, a distribuição marginal (ou preditiva) de X é uniforme em $\{0, 1, \ldots, n\}$.

Distribuição Multinomial

Neste modelo denota-se o número de ocorrências em cada uma de m categorias em n ensaios de Bernoulli independentes por $\mathbf{X} = (X_1, \ldots, X_m)$. As probabilidades associadas são denotadas por $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$.

Temos $m - 1$ parâmetros pois $\sum_{i=1}^{m} \theta_i = 1$.

A restrição $\sum_{i=1}^{m} X_i = n$ também se aplica.

Definição

Dizemos que o vetor **X** tem distribuição multinomial com parâmetros n e $\boldsymbol{\theta}$ se a função massa de probabilidade conjunta de **X** é dada por,

$$p(\mathbf{x}|\boldsymbol{\theta}) = \frac{n!}{\prod_{i=1}^{m} x_i!} \prod_{i=1}^{m} \theta_i^{x_i}, \quad x_i = 0, \ldots, n, \quad \sum_{i=1}^{m} x_i = n$$

com $0 < \theta_i < 1$ e $\sum_{i=1}^{m} \theta_i = 1$.

Verifique que a distribuição Binomial é um caso particular quando $m = 2$.

Assume-se que as probabilidades no vetor $\boldsymbol{\theta}$ são desconhecidas e seguem uma distribuição contínua cuja função de densidade conjunta é,

$$f(\boldsymbol{\theta}) = c, \; 0 < \theta_i < 1 \quad \text{e} \quad \sum_{i=1}^{m} \theta_i = 1,$$

para uma constante $c > 0$ e $f(\boldsymbol{\theta}) = 0$ caso contrário.

Qual a distribuição de $\mathbf{X}$ independente do valor de $\boldsymbol{\theta}$?

Ou seja, deseja-se obter a distribuição marginal de $\mathbf{X}$,

$$P(\mathbf{X} = \mathbf{x}) = \int \cdots \int p(\mathbf{x}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\theta_1 \dots d\theta_m.$$

Definição

Dizemos que o vetor aleatório $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ tem distribuição de Dirichlet com parâmetros $a_1, \ldots, a_m$ se a função densidade de probabilidade conjunta é,

$$f(\boldsymbol{\theta}) = \frac{\Gamma(a_1 + \cdots + a_m)}{\prod_{i=1}^{m} \Gamma(a_i)} \prod_{i=1}^{m} \theta_i^{a_i - 1},$$

com $0 < \theta_i < 1$, $\sum_{i=1}^{m} \theta_i = 1$ e $a_i > 0$.

Se $a_1 = \cdots = a_m = 1$ temos que,

$$f(\boldsymbol{\theta}) = \Gamma(m) = (m-1)!,$$

ou seja $c = (m-1)!$.

Portanto,

$$P(\mathbf{X} = \mathbf{x}) \quad = \quad \frac{n!(m-1)!}{\prod_{i=1}^{m} x_i!} \int \cdots \int \theta_1^{x_1} \ldots \theta_m^{x_m} \; d\theta_1 \ldots d\theta_m.$$

sujeito a $0 \leq \theta_i \leq 1$ e $\sum_{i=1}^{m} \theta_i = 1$.

Usando a definição da distribuição Dirichlet pode-se mostrar que,

$$\int \cdots \int \theta_1^{x_1} \ldots \theta_m^{x_m} \; d\theta_1 \ldots d\theta_m = \frac{\prod_{i=1}^{m} x_i!}{\sum_{i=1}^{m} (x_i + 1) - 1},$$

portanto,

$$P(\mathbf{X} = \mathbf{x}) = \frac{n!(m-1)!}{\prod_{i=1}^{m} x_i!} \frac{\prod_{i=1}^{m} x_i!}{\sum_{i=1}^{m} (x_i + 1) - 1} = \frac{n!(m-1)!}{(n+m-1)!}.$$