Probabilidades, Esperânça, Variância, etc.

April 24, 1998

1 Interpretação Frequentista

Em muitas circunstâncias estamos interessados em estudar situações que produzem resultados imprevisíveis. Chamamos estas situações de *experimentos*. Por exemplo, ao lançarmos um dado, consideramos seis resultados possíveis, cada resultado correspondendo a uma determinada face do dado voltada para cima. Chamamos de *Espaço Amostral*, A, ao conjunto de todos os possíveis resultados. Neste capítulo restringir-nos-emos a espaços amostrais *discretos*, i.e. que contêm um número finito (ou enumerável) de resultados. Chamamos *evento* um subconjunto do espaço amostral, $E \subseteq A$. Assim, no exemplo do dado, uma representação do espaço amostral é o conjunto $\{F1, F2, F3, F4, F5, F6\}$; nesta representação, o evento "obter uma face ímpar" corresponde ao subconjunto
$\{F1, F3, F5\}$.

Nos experimentos que estudaremos é útil atribuirmos a cada um dos eventos um valor numérico. Esta atribuição, ou função, chama-se *variável aleatória*. No exemplo do dado é usual atribuir valores, entre 1 e 6, a cada um dos eventos, i.e. atribuir $X(Fk) = k$.

Consideremos outro exemplo: lançamos dois dados, um verde e outro amarelo, cada qual tendo suas faces numeradas de 1 a 6. No exemplo dos dados verde e amarelo, entre muitas outras, poderíamos considerar as seguintes variáveis aleatórias:

1. O número decimal de dois dígitos cujo primeiro dígito corresponde a face de cima do dado verde, e o segundo dígito corresponde a face de cima do dado amarelo.

2. A soma dos números na face de cima dos dois dados.

Note que construímos duas variáveis aleatórias distintas, sobre um mesmo experimento: o lançamento dos dados verde e amarelo. Note também que pelo valor da primeira variável aleatória podemos saber exatamente o resultado obtido no experimento; o mesmo já não é verdade para a segunda variável aleatória. Em geral, quando lidamos com um experimento, temos uma dada variável aleatória em mente e, não havendo ambigüidade, quando mencionamos o experimento já subentendemos a variável aleatória apropriada.

Definimos a imagem de um evento $E \subseteq A$ por X, $X(E)$, como o conjunto de todos os valores assumidos por X dentro de E, i.e., $X(E) \equiv \{x = X(e), e \in E\}$. Analogamente definimos a pré-imagem de C por X, $X^{-1}(C)$, como o evento formado pelos resultados do experimento onde a variável X assume valores em C, i.e., $X^{-1}(C) = \{a \in A \mid X(a) \in C\}$.

Dado um experimento que possa ser repetido, interpretamos a *probabilidade* de um evento como a freqüência com que o obtemos como resultado do experimento. A probabilidade de que uma variável aleatória X assuma um valor dentro de um conjunto C, é simplesmente a probabilidade da pré-imagem de C por X. A *distribuição de probabilidade* de uma variável aleatória (discreta) é uma tabela especificando a probabilidade de que a variável assuma cada um dos seus possíveis valores.

Exemplos:

- Um dado é dito honesto se a freqüência com que gera cada um dos valores $\{1, \ldots, 6\}$ é a mesma. A probabilidade de obtermos um dado valor com um dado honesto é portanto de $1/6$.

- No experimento descrito no exemplo dos dados verde e amarelo, assumindo que ambos os dados são honestos, a probabilidade de cada variável aleatória assumir cada um dos valores possíveis é dada pelas tabelas seguintes:

  1. $Pr(X = x) = 1/36, \forall x \in \{11, 12, \ldots, 16, 21, \ldots, 26, \ldots, 61, \ldots, 66\}$.
  2. $X^{-1}(2) = \{(1,1)\} \Rightarrow Pr(X = 2) = 1/36$, $X^{-1}(3) = \{(1,2),(2,1)\} \Rightarrow Pr(X = 2) = 2/36$, $\ldots$, $X^{-1}(12) = \{(6,6)\} \Rightarrow Pr(X = 12) = 1/36$.

Propriedades da Probabilidade

Para assegurar a possibilidade de interpretar probabilidade de um evento como a de sua freqüência relativa, exigiremos que uma (medida de) probabilidade sobre o espaço amostral de um experimento aleatório, satisfaça as três propriedades (axiomas) seguintes. Notaremos por $\emptyset$ o conjunto vazio, e por $\bar{B} = A - B$ o complementar de um evento B no espaço amostral A do experimento.

1. $\forall B \subseteq A, \quad Pr(B) \geq 0$,
2. $Pr(A) = 1$,
3. Se $B \cap C = \emptyset$, então $Pr(B \cup C) = Pr(B) + Pr(C)$.

Exercício: Prove, a partir dos axiomas, as seguintes propriedades da probabilidade:

4. $Pr(\emptyset) = 0$,
5. $Pr(\bar{B}) = 1 - Pr(B)$,
6. $Pr(B \cup C) = Pr(B) + Pr(C) - Pr(B \cap C)$

2 Inferência

Em muitas situações temos que lidar com experimentos cujas especificações não conhecemos, ou conhecemos incompletamente. Por exemplo: Suponha sabermos que o dado verde foi construído tomando todos os cuidados para garantir que, exceto por pequenas marcas que distinguem suas faces, ele é absolutamente regular. Suponha também que o dado amarelo foi trazido por um desconhecido. Sabemos de antemão que o dado verde é honesto, mas não sabemos a distribuição de probabilidade para o segundo dado.

Para obter alguma informação sobre o dado amarelo podemos lançá-lo repetidas vezes, anotando a seqüência de valores. Este histórico de valores é o que se chama em estatística de uma *amostra*.

A partir da amostra poderíamos induzir alguma informação sobre a distribuição de probabilidade do dado amarelo: poderíamos por exemplo calcular a freqüência relativa de cada um dos valores na amostra; e então usar estas freqüências relativas como *estimativa* da distribuição de probabilidade do experimento. Este é um exemplo do que chamamos *inferência estatística*.

Note que a probabilidade da variável aleatória X assumir o valor x, $\Pr(X = x)$, e a estimativa utilizada, freqüência relativa na amostra, são conceitos distintos: A distribuição de probabilidade é um atributo do experimento em si, enquanto a estimativa baseia-se numa particular amostra de resultados do experimento. No uso prático, geralmente, a amostra resume nossas observações passadas, queremos basear alguma decisão futura na distribuição de probabilidade do experimento (que todavia desconhecemos), e usamos a estimativa como elo de ligação para tomar esta decisão.

Máxima Verossimilhança

Como fazer e justificar estimativas e inferências que sejam ótimas, segundo critérios pré-estabelecidos, é o objeto de estudo da estatística matemática. No caso do dado amarelo, o critério usado para fazer a estimativa foi o da *máxima verossimilhança*: A melhor estimativa da distribuição, segundo o critério da máxima verossimilhança, é a que maximiza a probabilidade de obter, num experimento com a distribuição estimada, uma amostra idêntica a amostra original.

Se a amostra original foi uma seqüência de n resultados, com $n_1,\ n_2, \ldots, n_6$ ocorrências dos valores $1,\ 2, \ldots, 6$, a probabilidade de obter uma amostra com as mesmas feqüências, independentemente da ordem, é

$$Pr(X = [n_1,\ n_2, \ldots n_6]) = \frac{n!}{n_1!\, n_2! \ldots n_6!} p_1^{n_1} p_2^{n_2} \ldots p_6^{n_6}$$

É um exercício de cálculo mostrar que a distribuição que maximiza esta probabilidade é a das freqüências relativas, $p_i = f_i$, onde $f_i = n_i/n$. Prove este resultado no caso de uma moeda.

3 Esperança

Dado um experimento, o valor esperado de uma variável aleatória X, ou sua *esperança*, é a média aritmética sobre o conjunto dos valores possíveis, $x \in A$, ponderados pela distribuição de probabilidade:

$$E(X) = \sum_{x \in X(A)} x * \Pr(X = x) \quad .$$

Dada uma amostra, i.e. uma seqüência de valores,

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad ,$$

definimos a sua **média** aritmética como sendo

$$\mu_x = \frac{1}{n}\sum_{i=1}^{n} x_i \quad .$$

Dada uma amostra, usaremos a sua média como estimativa da esperança. Esta estimativa é consistente com a estimativa da distribuição de probabilidade idêntica às freqüências relativas na amostra, que por sua vez tinha justificativa no critério da máxima verossimilhança.

Propriedades de Transformação

Estaremos sempre interessados em estudar transformações lineares de variáveis aleatórias, $Z = \alpha X + \beta Y + \gamma$ e como estas transformações se refletem nas quantidades estatísticas a serem definidas, por exemplo:

$$E(\alpha X + \beta Y + \gamma) = \alpha E(X) + \beta E(Y) + \gamma$$

A prova pode ser dividida em duas partes:

$$\begin{aligned} E(\alpha X + \gamma) &= \sum_x (\alpha x + \gamma)\Pr(X = x) \\ &= \gamma + \alpha \sum_x x\Pr(X = x) \\ &= \gamma + \alpha E(X) \end{aligned}$$

$$\begin{aligned} E(X + Y) &= \sum_{x,y} (x + y)\Pr(X = x \wedge Y = y) \\ &= \sum_{x,y} x\Pr(X = x \wedge Y = y) \quad + \sum_{x,y} y\Pr(X = x \wedge Y = y) \\ &= \sum_x x\Pr(X = x \wedge Y \in Y(A)) + \sum_y y\Pr(Y = y \wedge X \in X(A)) \\ &= E(X) + E(Y) \end{aligned}$$

4 Variância

A variância de uma variável aleatória é uma medida de erro ou dispersão desta variável em torno da sua esperança:

$$Var(X) = E((X - E(X))^2) \ .$$

É fácil ver que também podemos calcular a variância como

$$\begin{aligned} Var(X) &= E((X - E(X))^2) \\ &= E(X^2 - 2XE(X) + E(X)^2) \\ &= E(X^2) - E(X)^2 \end{aligned}$$

Desvio Padrão

Desvio padrão, $\sigma_x = \sqrt{var(x)}$ tem a mesma dimensão, ou unidade de medida, que $E(x)$ ou x, i.e., é uma medida de desvio comensurável com a média ou os valores assumidos pela variável aleatória, sendo portanto de interpretação mais natural.

Justificativa da Norma Quadrática

Existem outras alternativas para medir o tamanho do erro, x, além da norma $L_2 = (x'x)^{1/2}$; Por exemplo as normas $L_1 = \mathbf{1}'\text{abs}(x)$, ou $L_\infty = \max(\,\text{abs}(x)\,)$. Tanto L_1 como L_∞ são usadas na estatística em certas situações especiais, todavia, L_2 é usada na maioria das situações. Além de ser computacionalmente mais simples, é a única que tem a propriedade de ser invariante por uma (pelo grupo de) rotação.

5 Covariância

A covariância entre duas variáveis aleatórias, X e Y, é definida como

$$Cov(X, Y) = E((X - E(X)) * (Y - E(Y))) \ .$$

A covariância é uma medida da "interdependência" dos desvios de ambas as variáveis em relação a média. Adiaremos uma interpretação intuitiva mais detalhada para o conceito de correlação, discutido adiante. Por hora, podemos verificar as seguintes propriedades:

$$\begin{aligned} Cov(X,Y) &= E(XY - YE(X) - XE(Y) + E(X)E(Y)) \\ &= E(XY) - E(Y)E(X) \end{aligned}$$

donde

$$Cov(X,X) = E(X^2) - E(X)^2 = Var(X)$$

Propriedades de Transformação

Lema :

$$\begin{aligned} & Cov(\alpha X + \beta Y + \gamma, Z) \\ &= E((\alpha X + \beta Y + \gamma)Z) - E(\alpha X + \beta Y + \gamma)E(Z) \\ &= \alpha E(XZ) + \beta E(YZ) + \gamma E(Z) - \alpha E(X)E(Z) - \beta E(Y)E(Z) - \gamma E(Z) \\ &= \alpha Cov(X,Z) + \beta Cov(Y,Z) \end{aligned}$$

Do lema segue que:

$$Var(\alpha X + \beta Y + \gamma) = \alpha^2 Var(X) + \beta^2 Var(y) + 2\alpha\beta Cov(X,Y)$$

Notação Matricial

$$Cov(X)_{i,j} \equiv Cov(X_i, X_j)$$

Dado X um vetor de n variáveis aleatórias, A uma matriz real $m \times n$, e b um vetor real $m \times 1$, os resultados precedentes implicam na forma genérica da esperança e variância de uma transformação linear, que é dada por:

$$E(AX + b) = AE(X) + b$$

$$Cov(AX + b) = A\,Cov(X)\,A'$$

Em estatística é usual a notação $Cov(X)_{i,j} = \sigma_{i,j}$. Nesta notação o desvio padrão é $\sigma_i = \sqrt{\sigma_{i,i}}$.

Consideremos realizar n vezes e independentemente, um experimento cujos resultados são medidos pelo vetor de variáveis aleatórias $X = [X_1, \ldots, X_n]$, a variância da média é dada por

$$E(\bar{X}) = E(\ (1/n)\ \mathbf{1}'X\ ) = (1/n)\ \mathbf{1}'E(X) = E(X_1)$$

na ultima equação denotamos a somatória $\sum X$ por $\mathbf{1}'X$, onde $\mathbf{1}' = [1, 1, \ldots 1]$.

$$\begin{aligned}
Var(\bar{X}) &\equiv Var(\ (1/n)\ \mathbf{1}'X\ ) \\
&= (1/n^2)\ \mathbf{1}'\ \mathrm{diag}([\sigma_{1,1}, \ldots \sigma_{n,n}])\ \mathbf{1} \\
&= (1/n^2)(\ \sigma_{1,1} + \ldots + \sigma_{n,n}\ ) \\
&= (1/n)\sigma_{1,1}
\end{aligned}$$

Correlação

A correlação entre duas variáveis aleatórias, $Cor(X, Y)$ ou $\rho_{i,j} \equiv Cor(X_i, X_j)$, é a covariância "normalizada" pelo desvio padrão:

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} \quad \text{ou} \quad \rho_{i,j} = \frac{\sigma_{i,j}}{\sigma_i \sigma_j}$$

A correlação é uma medida incompleta de independência entre variáveis aleatórias, como mostraremos a seguir:

Dois eventos, B e C, são *independentes* sse $Pr(B \cap C) = Pr(B) * Pr(C)$. Duas variáveis aleatórias, X e Y, são independentes sse, para qualquer conjunto de valores, C e D, $Pr(X \in C e Y \in D) = Pr(X \in C) * Pr(Y \in D)$. Temos então as seguintes propriedades:

- Se duas variáveis aleatórias (de variância limitada), X e Y, são independentes, então $Cov(X, Y) = 0$, pois $E(XY) = E(X)E(Y)$.
- Todavia, correlação nula não garante independência! Considere duas variáveis aleatórias definidas sobre os resultados de um dado honesto: X assumindo valor -1 em $F1$, 1 em $F6$, e valor 0 em todas as outras faces; Y assumindo valor 1 em $\{F1, F6\}$, e valor 0 em todas as outras faces. As variáveis aleatórias X e Y não são independentes, embora tenham correlação nula (verifique).

– No caso de uma dependência linear, $Y = \alpha X + \gamma$, temos que:

$$Cor(Y,X) = \frac{Cov(\alpha X + \gamma, X)}{\sigma(\alpha X + \gamma)\sigma(X)} = \frac{\alpha Var(X)}{\sqrt{\alpha^2}\sigma_X \sigma_X} = sign(\alpha)$$

onde $sign(x) \equiv \ 1$ se $x > 0$, $\ -1$ se $x < 0$, e $\ 0$ se $x = 0$.

Provemos finalmente que $-1 \leq \rho_{i,j} \leq 1$:
Tomemos $X = [X_1, X_2]'$, e $Y = [a_1, a_2]X$.

$$\begin{aligned}
Var(Y) &= \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} \\ \sigma_{2,1} & \sigma_{2,2} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \\
&= \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} 1 & \rho_{1,2} \\ \rho_{2,1} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \\
&= \begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\
&= b_1^2 + 2\rho b_1 b_2 + b_2^2
\end{aligned}$$

Observemos agora que, pela definição de variância, $Var(Y) \geq 0$. Observemos ainda que se $-1 \leq \rho \leq 1$, podemos sempre "completar os quadrados" de modo a reescrever $Var(Y) = (b_1 \pm b_2)^2 + \text{abs}((1 \pm \rho) b_1 b_2)$, quantidade obviamente positiva. Todavia se $\text{abs}(\rho) > 1$, podemos sempre escolher valores para b_1 e b_2 em $\{-1, 1\}$ que tornam $b_1^2 + 2\rho b_1 b_2 + b_2^2 < 0$, uma contradição.

6 Utilidade e Decisões

Consideremos a seguinte aposta: Um banqueiro lança um dado honesto e paga o resultado, X o número na face de cima, em gramas de ouro. Um apostador, para apostar, deve pagar 4 gramas de ouro. Consideraremos nesta seção um modelo matemático, a *teoria estatística da decisão*, que nos dá critérios para tomar a decisão de participar ou não da aposta.

Além da descrição da aposta em si, precisamos de uma descrição das preferências do apostador. Esta descrição é dada por uma *função de ganho*, $G(x)$. $G(x)$ nos dá o ganho atribuído pelo apostador a cada um dos possíveis valores assumidos pela variável aleatória que caracteriza a aposta.

Consideremos um apostador, A, que se importe apenas com a quantidade de ouro ganha ou perdida no final da aposta. As preferências de A são adequadamente descritas pela função de ganho $G_A(x) = x - 4$.

Considere um segundo apostador, B, que possui 4 gramas de ouro, mas que só poderá participar de um negócio que com certeza lhe triplicará o capital se dispuser de 6 gramas. A função de ganho de B é: $G_B(x) = x - 4$ se $x \leq 5$, e $G(6) = 3 * 6 - 4 = 14$.

Um critério para medir a utilidade de um apostador, com função de ganho $G()$, participar de uma aposta caracterizada pela variável aleatória X, é o *ganho esperado*:

$$U(X) = E(G(X)) = \sum_x G(x) * \Pr(X = x)$$

Nos dois exemplos anteriores teríamos

$$\begin{aligned} U_A(X) &= E(G_A(X)) = \sum_{x=1}^{6} (x-4) * (1/6) = -0.5 \\ U_B(X) &= E(G_B(X)) = \sum_{x=1}^{5} (x-4) * (1/6) \ + 14 * (1/6) = 1.5 \end{aligned}$$

ou seja, como resultado final de seu envolvimento na aposta, A espera perder 0.5 gramas, enquanto B espera ganhar 1.5 gramas de ouro. Usando o critério da utilidade esperada B deveria fazer a aposta, enquanto A todavia não.

Critério de Média-e-Variância

Em finanças e economia é usual considerar investidores (ou agentes econômicos, produtores, consumidores, etc.) que desejam maximizar seus ganhos (ou minimizar seus custos). Quando os ganhos do investidor dependem do valor de variáveis aleatórias, é razoável assumir que este deseje maximizar o valor esperado do ganho. Todavia, apenas a esperança do ganho, pode ser uma caracterização incompleta das preferências do investidor. Muitos investidores são "avessos ao risco", preferindo usar um processo cujo ganho tem um dado valor esperado e variância, a um segundo processo cujo ganho tem valor esperado ligeiramente superior, mas mas variância bem maior. Usaremos a seguir a teoria de utilidade para estudar situações deste tipo.

A função utilidade (ou critério) da média-e-variância é definida como:

$$U(G, X, \alpha, \beta) = \alpha E(G(X)) - \beta Var(G(X))$$

Esta função utilidade (ou critério) é muito empregado em Economia e Finanças, onde o parâmetro β/α é interpretado como uma medida da *aversão ao risco* do investidor. Tivemos oportunidade de desenvolver modelos deste tipo no capítulo 3.

Um critério quadrático pode também ser visto como a aproximação truncada da série de Taylor de uma função genérica. Consideremos a variável aleatória auxiliar $y = G(x)$, e a série de taylor:

$$U(y) = U[E(Y)] + \sum_{k=1}^{\infty} U^{(k)}[E(Y)] * (y - E(Y))^k / k!$$

Truncando a série após o termo quadrático, temos

$$U(y) \approx U[E(Y)] + U'[E(Y)] * (y - E(Y)) + U''[E(Y)] * (y - E(Y))^2/2$$

Do valor esperado desta aproximação obtemos

$$E(U(Y)) \approx U[E(Y)] + (U''[E(Y)]/2) Var(Y)$$

A última aproximação nos dá uma nova interpretação para o parâmetro β no critério da média-e-variância. Em particular, obtemos uma legítima aversão ao risco, i.e. $\alpha > 0$ e $\beta > 0$, na situação convencional de uma utilidade côncava, onde $U()$ é crescente, $U'() > 0$, com incrementos decrescentes, $U''() < 0$. Poderíamos aprofundar o estudo de como obter os parâmetros da função quadrática a partir de pressupostos sobrea a "real" função utilidade do investidor; todavia, para aplicações a finanças, temos disponíveis métodos bem mais diretos, conforme vimos no capítulo 3.