Intro à Probabilidade

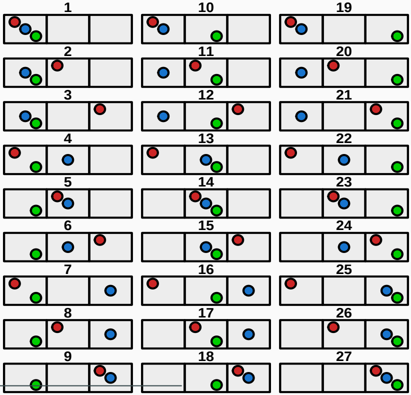
Eduardo Horta

Aula 5

Princípios de Análise Combinatória

Exemplo

- Considere o seguinte "experimento": dispor, ao acaso, três bolas (a = vermelha, b = azul e c = verde, respectivamente) em três urnas[1] (permitindo-se, possivelmente, mais de uma bola em uma mesma urna).



[1] Cada arranjo/configuração na figura representa um evento simples.

Exemplo

- Qual seria um espaço amostral adequado para representar os possíveis resultados do experimento acima?
- Sugestão: vamos representar cada resultado ω por uma terna ordenada $(\omega_a, \omega_b, \omega_c)$, onde cada uma das coordenadas deve ser um dos números 1, 2 ou 3.
- Interpretação: por exemplo, o evento elementar "bola a na primeira urna, bola b na terceira urna e bola c na primeira urna" corresponde ao elemento $\omega = (1, 3, 1)$ do espaço amostral.
- Assim, temos

$$\Omega = \{1, 2, 3\} \times \{1, 2, 3\} \times \{1, 2, 3\}$$

Exemplo

- Considere agora a seguinte generalização do “experimento” descrito acima: dispor, ao acaso, r bolas em n urnas.
- Vamos denotar a primeira bola pelo símbolo $\bar{1}$, a segunda bola pelo símbolo $\bar{2}$, etc.
- Ainda, vamos denotar a primeira urna pelo número 1, a segunda urna pelo número 2, etc.
- Então, podemos descrever qualquer configuração como sendo uma r-upla da forma $(\omega_1, \omega_2, \ldots, \omega_r)$, onde cada ω_j é um elemento do conjunto $\{1, 2, \ldots, n\}$.

Exemplo: generalização

- Assim, por exemplo, se a j-ésima entrada ω_j for k, isso significa que a bola $\bar{j}$ foi colocada na urna k.
- Segue que um espaço amostral adequado para descrever esse experimento é o conjunto

$$\Omega = \{1, 2, \ldots, n\}^r$$

  isto é, o conjunto de todas as r-uplas como acima.
- **Exercício**: no espaço amostral Ω como acima, represente o evento "todas as bolas em uma mesma urna".
- **Exercício**: quantos elementos tem o espaço amostral Ω nesse caso?

O Princípio Fundamental da Contagem

- **Princípio Fundamental da Contagem**: se uma primeira ação pode ocorrer de m_1 maneiras diferentes, se uma segunda ação pode ocorrer de m_2 maneiras diferentes, e assim por diante, isto é, se uma r-ésima ação pode ocorrer de m_r maneiras diferentes, então o número de maneiras diferentes em que as r ações podem ocorrer simultaneamente é igual a $m_1 m_2 \cdots m_r$.

Alguns experimentos equivalentes

- “We use this example to illustrate the important fact that the nature of the sample points is irrelevant for our theory. To us the sample space (together with the probability distribution defined in it) defines the idealized experiment. We use the picturesque language of balls and cells, but the same sample space admits of a great variety of different practical interpretations. To clarify this point, and also for further reference, we list here a number of situations in which the intuitive background varies; all are, however, abstractly equivalent to the scheme of placing r balls into n cells, in the sense that the outcomes differ only in their verbal description. The appropriate assignment of probabilities is not the same in all cases and will be discussed later on.” (Feller, Vol I)

Exemplos:

- **Aniversários**: as possíveis configurações de aniversários de r pessoas correspondem aos diferentes arranjos de r bolas em $n = 365$ urnas (assumindo-se, por simplicidade, que o ano sempre possui 365 dias).
- Nesse caso, enumeramos (de forma arbitrária, mas única) as r pessoas com os símbolos de $\bar{1}$ até $\bar{r}$, e fazemos a identificação 1 = "primeiro de janeiro", 33 = "2 de fevereiro", etc.
- Assim, se Joãozinho está representado pelo símbolo $\bar{7}$, para saber qual o dia de seu aniversário basta olharmos em qual urna (dentre as urnas 1 a n) a bola de número 7 foi colocada.

Exemplos:

- **Dados**: os possíveis resultados do lançamento de r dados correspondem às possíveis configurações de r bolas em $n = 6$ urnas.
- **Moedas**: os possíveis resultados do lançamento de r moedas correspondem às possíveis configurações de r bolas em $n = 2$ urnas.
- **Coleção de figurinhas**: os diferentes tipos de figurinha são representados pelas urnas; cada figurinha adquirida corresponde a uma bola.

Exemplos

- **Distribuição de genes**: os descendentes de um organismo vivo herdam de seu progenitor certos genes. Se um gene particular pode aparecer em n formas, digamos $A_1, A_2, \ldots, A_n$, então os descendentes podem ser classificados por tipo genético. Cada descendente corresponde a uma bola, enquanto os genótipos correspondem às urnas.
- **Exercício**: dê os detalhes (espaço amostral, evento elementar típico, etc) em cada um dos exemplos acima.

Bolas e urnas: quantas configurações?

- Voltando ao caso geral: podemos utilizar o Princípio Fundamental da Contagem para determinar de quantas maneiras é possível colocar r bolas em n urnas.
- De fato, se considerarmos o problema iterativamente, então para a primeira bola podemos escolher dentre n urnas; para a segunda bola, da mesma forma, temos n possíveis escolhas, e assim por diante.
- Na notação anterior, temos $m_1 = m_2 = \cdots = m_r = n$. Logo, o número de possíveis configurações de r bolas em n urnas é dado por

$$\underbrace{n \times n \times \cdots \times n}_{r \text{ vezes}} = n^r$$

Aula 6

Amostras ordenadas

Amostras ordenadas

- Consideremos agora o problema de selecionar, ao acaso, r unidades de uma “população” constituída de um total de n unidades ($r \leq n$).
- Cada possível seleção (levando-se em conta a ordem) de r unidades é dita uma **amostra** de tamanho r dessa população.
- É conveniente considerar que essa seleção será feita sequencialmente.
- Esse procedimento pode ser executado de duas maneiras: *com reposição* ou *sem reposição*.

Amostras ordenadas

- No procedimento *com reposição*, selecionamos um elemento dentre os n possíveis, anotamos qual foi o elemento selecionado e, após, passamos à seleção do segundo elemento, *sendo que a unidade selecionada na primeira etapa pode ser selecionada novamente*. Repetimos o procedimento até que r unidades tenham sido selecionadas.
- Pelo mesmo raciocínio utilizado anteriormente, é fácil concluir que existem n^r possíveis amostras (com reposição) de tamanho r de uma população constituída de n elementos.
- De fato, se cada unidade da população for representada por uma urna, e cada seleção for representada por uma bola, voltamos ao espaço amostral discutido acima.

Amostras ordenadas

- No procedimento *sem reposição*, selecionamos um elemento dentre os n possíveis, anotamos qual foi o elemento selecionado e, após, passamos à seleção do segundo elemento, *dentre as unidades ainda não selecionadas*. Repetimos o procedimento até que r unidades tenham sido selecionadas.
- Nesse caso, temos n possíveis escolhas para a primeira seleção, $n-1$ possíveis escolhas para a segunda seleção, e assim por diante.
- Segue que o número de possíveis amostras (sem reposição) de tamanho r de uma população constituída de n elementos é igual a

$$(n)_r = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1)$$

Importante

$$(n)_r = \frac{n!}{(n-r)!}$$

Amostras ordenadas

- Em suma, temos o seguinte:
  Teorema*: Dados uma população constituída de n elementos e um tamanho de amostra fixado r, existem n^r diferentes amostras ordenadas com reposição, e $(n)_r$ diferentes amostras ordenadas sem reposição.*

Permutações

- Um caso especial do resultado acima ocorre quando consideramos um cenário de amostra *sem* reposição com $r = n$.
- Nesse caso, uma tal amostra nada mais é do que um reordenamento (ou seja, uma permutação) dos elementos da população.
- Segue que uma população constituída de n elementos admite $n!$ amostras de tamanho n, sem reposição.
- Dito de outra forma: o número de diferentes maneiras pelas quais se pode ordenar um conjunto de n elementos é igual a $n!$

Exemplos

- Três pessoas, $\bar{1}$, $\bar{2}$ e $\bar{3}$, formam uma amostra ordenada (sem reposição) retirada da população humana. Seus aniversários formam uma amostra (com reposição) do conjunto de números $\{1, 2, \ldots, 365\}$.
- **Palavras aleatórias**: vamos chamar de *palavra com 10 letras* qualquer concatenação (string) de comprimento 10 formada pelas letras do alfabeto $\{A, B, C, \ldots, X, Y, Z\}$ (permitindo, possivelmente, “palavras” sem sentido). Como repetições são permitidas, existem 26^{10} tais palavras.
- Sr. e Sra. Smith formam uma amostra de tamanho $r = 2$ retirada da população humana. Da mesma forma, os dois considerados juntamente formam uma amostra de tamanho $r = 1$ retirada da população de todos os casais.

Comentários

- Selecionar, sucessivamente (e ao acaso), r elementos de uma população constituída de n elementos é um *experimento* cujos possíveis resultados são amostras de tamanho r.
- O termo “ao acaso” não é definido explicitamente, e vamos interpretá-lo como sendo relacionado ao tipo de amostra que estamos considerando.
- Por ora, será conveniente assumir que cada amostra de tamanho r tem probabilidade

$$\mathbb{P}\{\omega\} = \begin{cases} \frac{1}{n^r}, & \text{se amostra com reposição} \\ \frac{1}{(n)_r}, & \text{se amostra sem reposição} \end{cases}$$

Um corolário

- Suponhamos que uma amostra de tamanho r foi selecionada (ao acaso, com reposição) de uma população constituída de n elementos.
- Pelo Teorema acima, existem n^r tais amostras, das quais $(n)_r$ não apresentam repetições (isto é, cada elemento da população é selecionado apenas uma vez).
- Segue que, numa amostra como essa, a probabilidade de selecionarmos uma amostra em que não ocorrem repetições é igual a

$$\frac{(n)_r}{n^r} = \frac{n(n-1)\cdots(n-r+1)}{n^r}$$

Exemplo

- Considere uma "população" formada pelos dígitos $0, 1, \ldots, 9$.
- Cada *string* de 5 dígitos pode ser interpretada como uma amostra de tamanho $r = 5$ retirada de uma população de tamanho $n = 10$.
- Em vista do resultado acima, a probabilidade de 5 dígitos selecionados ao acaso, com reposição, serem todos distintos é igual a

$$\frac{(10)_5}{10^5} = 0.3024$$

Bolas e urnas: o retorno!

- Se n bolas são colocadas ao acaso em n urnas, então a probabilidade de todas as urnas serem ocupadas é $n!/n^n$.
- Ao aumentarmos o valor de n, essa quantidade converge para zero rapidamente: com $n = 7$, seu valor é de 0.00612. Com $n = 10$, obtém-se 0.00036288.
- **Dados**: em um lançamento de 6 dados equilibrados, vemos que a probabilidade de observarmos todos os possíveis resultados é de apenas $6!/6^6 = 0.01543$.

Exemplo: Aniversários

- Podemos interpretar os dias de aniversário de r pessoas como sendo uma amostra (com reposição) da "população" constituída dos números $1, 2, \ldots, 365$ (por simplicidade, vamos ignorar o 29/fevereiro).
- Utilizando o resultado acima, podemos concluir que a probabilidade de todas as pessoas fazerem aniversário em dias diferentes é igual a
$$\frac{(365)_r}{365^r}$$

Subpopulações

Subpopulações

- *Lembrando*: usamos o termo **população** para designar uma coleção de n elementos *sem qualquer menção a ordem*.
- Em particular, isso significa que duas populações são consideradas distintas se, e somente se, uma delas possui ao menos um elemento que não pertence à outra.
- Consideremos agora uma *subpopulação*, de tamanho r, obtida de uma dada população constituída de n elementos. *De novo: não estamos levando em conta a ordem dos elementos da subpopulação*.
- Isto é, não estamos pensando em uma *amostra* de tamanho r dessa população: todas as amostras que contém os mesmos elementos correspondem, ao desconsiderarmos a ordem de seus elementos, a uma mesma subpopulação.

Subpopulações

- Agora, uma dada subpopulação de tamanho r pode ser ordenada de $r!$ diferentes maneiras (conforme vimos acima).
- Casa uma dessas ordenações corresponde a uma única amostra (sem reposição) de tamanho r da população.
- Segue que o número de amostras (sem reposição) de tamanho r é igual a $r!$ vezes o número de subpopulações de tamanho r, isto é,

$$(n)_r = r! \times (\text{número de subpopulações de tamanho } r).$$

- O número de subpopulações de tamanho r é, portanto,

$$\binom{n}{r} = \frac{(n)_r}{r!}.$$

Coeficientes binomiais

- Um número da forma

$$\binom{n}{r} = \frac{(n)_r}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

  é dito um **coeficiente binomial**.
- Usualmente, lemos um tal número como “n escolhe k”, visto que representa de quantas diferentes maneiras podemos escolher r elementos de um conjunto de n elementos.
  Teorema *Uma população de n elementos possui $\binom{n}{r}$ diferentes subpopulações de tamanho $r \leq n$.*

Um corolário interessante

- Como os elementos de uma subpopulação de tamanho r são inteiramente determinados pelos $n - r$ elementos que *não* pertencem a essa subpopulação, vemos que a cada subpopulação de tamanho $n - r$ corresponde uma (e só uma) subpopulação de tamanho r.
- Logo, temos o seguinte:

$$\binom{n}{r} = \binom{n}{n-r}$$

- Será conveniente *definirmos*

$$(n)_0 = 1, \qquad 0! = 1, \qquad \binom{n}{0} = 1,$$

Ferramentas

http://www.di.fc.ul.pt/~jpn/r/combinatorics/combinatorics.html