FIS02014 - Um pouco de probabilidade e de estatística

Prof. Basílio X. Santiago

Este texto é inspirado numa aula que Don Winget (Univ. do Texas) ministrou para esta disciplina no semestre 2007-2. Baseia-se também nas notas de aula do mesmo pesquisador/professor, cujas cópias nos foram gentilmente disponibilizadas pelo professor Kepler Oliveira.

Definições básicas

A probabilidade de um dado evento ocorrer num dado experimento pode ser compreendida como a fração de ocorrências ou realizações deste experimento que tem como resultado o dado evento.

Para exemplificar, pensemos num dado. O dado é o experimento. Podemos jogá-lo quantas vezes quisermos. Então o número de realizações do experimento é limitado pela nossa paciência! São 6 eventos possíveis que podem resultar de uma realização (jogada) do dado: os números 1, 2, 3, 4, 5 e 6. Se jogarmos um dado N_{tot} vezes, sendo N_{tot} um número bem alto, digamos que $N(6)$ dessas realizações tenham tido como resultado o número 6. Então , uma estimativa da probabilidade de ocorrer o evento 6 é:

$$p(6) = \frac{N(6)}{N_{tot}}$$

A probabilidade assim determinada é resultado de uma medida, de pura experimentação . É portanto chamada de probabilidade *a posteriori*. Note que, por definição , a probabilidade é sempre expressa por um número no intervalo $[0, 1]$. Se a probabilidade de um dado evento ocorrer é $p = 0$, isso significa que certamente esse evento nunca ocorrerá. Se, por outro lado, $p = 1$, o evento sempre ocorrerá.

Claro que no caso de um dado normal, podemos fazer um cálculo simples, antes de jogar o dado, do valor da probabilidade $p(6)$. Ora, se são N_e possíveis eventos resultantes de uma realização de um experimento e se eles são equiprováveis (têm a mesma probabilidade de ocorrer), então esperamos que N_{tot}/N_e realizações tenham como resultado um evento específico dentre os possíveis. Usando o mesmo estimador de probabilidade acima, a probabilidade de ocorrer o i-ésimo evento (i pode variar no intervalo $[1, N_e]$), portanto, será:

$$p(i) = \frac{N_{tot}/N_e}{N_{tot}} = \frac{1}{N_e}$$

Claro que $p(1) = p(2) = ... = p(i) = ... = p(N_e)$, já que todos os eventos são igualmente prováveis. No caso do dado temos $N_e = 6$, de forma que $p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = 1/6$. Esta probabilidade, estimada sem experimentação , com base em concepções sobre o experimento em questão , pode ser chamada de probabilidade *a priori*.

Nem sempre é possível estimar probabilidades *a priori*, simplesmente porque não dispomos de um modelo adequado para isso. Por exemplo, nem todos os experimentos têm eventos equiprováveis. Ou nem todos os experimentos têm um universo de eventos totalmente conhecido.

Mas um conceito que é muito útil para a estimativa de probabilidades é o de *distribuição de probabilidades*, $p(x)$. Ela resume aquilo que sabemos sobre as probabilidades de diferentes eventos (quantificados pela variável x) ocorrerem num dado experimento. No caso do dado, a distribuição de probabilidades é muito simples, sendo uma função constante de valor 1/6, $p(x) = 1/6$, onde $x = 1, 2, 3, 4, 5, 6$. Note que a soma $p(1) + p(2) + ... + p(6) = 6x1/6 = 1$. Isso significa que a probabilidade de uma jogada do dado resultar em algum valor entre 1 e 6 é de $1 = 100\%$, o que é óbvio!

Note então que a soma das probabilidades dos diferentes eventos significa a probabilidade que qualquer um desses diferentes eventos ocorra. No caso de um dado, por exemplo, podemos calcular $p(1, 2, 3)$, a probabilidade de que uma jogada resulte OU em 1, OU em 2 OU em 3. Logicamente, $p(1, 2, 3) = p(1) + p(2) + p(3) = 3x1/6 = 1/2$.

Muito diferente é a multiplicação de duas ou mais probabilidades, que nos dá a probabilidade de um conjunto de realizações resultar, em seqüência, em determinados eventos. Seja por exemplo a probabilidade de se jogar um dado e obter como resultado o evento (número) 1 e, logo em seguida, jogá-lo novamente e obter como resultado o evento (número) 2. Ela pode ser escrita como $p(12) = p(1)p(2) = 1/36$. A probabilidade de 2 ou mais eventos ocorrerem em sucessão é chamada de *probabilidade adjunta*.

É muito importante também sermos capazes de obter estatísticas sobre uma distribuição de probabilidades $p(x)$. A média, por exemplo, é uma estatística muito útil. Qual o valor médio esperado de um grande número de jogadas de um dado? Se o jogarmos N_{tot} vezes, a média será

$$\overline{x} = \frac{\sum_{i=1}^{N_{tot}} x_i}{N_{tot}},$$

onde x_i pode assumir valores inteiros no intervalo $[1, 6]$. Podemos sempre reescrever a expressão acima agrupando os resultados iguais, multiplicando cada valor possível pelo número de realizações em que esse valor ocorre, $n(x)$. Mas esse número de realizações é proporcional à distribuição de probabilidade, pois $p(x) = n(x)/N_{tot}$. Então temos:

$$\overline{x} = \frac{\sum_{i=1}^{6} x_i \; n(x_i)}{N_{tot}} = \sum_{i=1}^{6} x_i \; p(x_i)$$

Note que a soma na última expressão acima se dá sobre o número de eventos e não mais sobre o número de realizações , pois estas foram agrupadas para eventos iguais.

Como no caso do dado $p(x) = 1/6$, o cálculo da média fica:

$$\overline{x} = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{1+2+3+4+5+6}{6} = 3.5$$

Num caso geral, em que os resultados cobrem um domínio contínuo de valores da variável x, digamos que o domínio $[x_{min}, x_{max}]$, então podemos substituir a soma que aparece no cálculo da média por uma integral

$$\overline{x} = \int_{x_{min}}^{x_{max}} x \; p(x) \; dx$$

Assim como no caso do dado, lembramos que

$$\int_{x_{min}}^{x_{max}} p(x) \; dx = 1$$

Ou seja, dizemos que a distribuição de probabilidades é normalizada para a unidade. Analogamente ao que foi feito com o dado, podemos nos perguntar qual o número de realizações que resultam em valores no intervalo $[x, x + dx]$. Para um total de N_{tot} realizações , esse número $n(x) \; dx$, será:

$$n(x) \; dx = N_{tot} \; p(x) \; dx$$

Na verdade, podemos definir toda uma família de integrais, que são os momentos da distribuição p(x). O momento de ordem n de p(x), μ_n é dado pela expressão :

$$\mu_n = \int_{x_{min}}^{x_{max}} x^n \; p(x) \; dx$$

Assim, a normalização da distribuição p(x) nada mais é do que o momento de ordem $n = 0$. Da mesma forma, a média de valores esperada para a distribuição será o momento de ordem $n = 1$ (ou primeira ordem).

Além da média, outra estatística muito importante é a variância, σ^2, definida como:

$$\sigma^2 = \overline{(x - \overline{x})^2}$$

Podemos usar a definição da média e desenvolver a expressão acima

$$\sigma^2 = \int_{x_{min}}^{x_{max}} (x - \overline{x})^2 \; p(x) \; dx$$

$$\sigma^2 = \int_{x_{min}}^{x_{max}} (x^2 + \overline{x}^2 - 2x\overline{x}) \; p(x) \; dx$$

$$\sigma^2 = \int_{x_{min}}^{x_{max}} x^2 \; p(x) \; dx + \int_{x_{min}}^{x_{max}} \overline{x}^2 \; p(x) \; dx - 2 \int_{x_{min}}^{x_{max}} x \; \overline{x} \; p(x) \; dx$$

$$\sigma^2 = \int_{x_{min}}^{x_{max}} x^2 \; p(x) \; dx + \overline{x}^2 \int_{x_{min}}^{x_{max}} p(x) \; dx - 2 \; \overline{x} \int_{x_{min}}^{x_{max}} x \; p(x) \; dx$$

O primeiro termo no lado direito da última igualdade acima é o momento de 2a ordem ($n = 2$), μ_2. O 2o termo é simplesmente igual à média ao quadrado, $\overline{x}^2$, pois $\int p(x)dx = 1$. O último termo do lado direito é $-2 \; \overline{x}^2$, pois a integral nele presente é a definição da própria média. Logo:

$$\sigma^2 = \mu_2 - \overline{x}^2 = \mu_2 - \mu_1^2$$

Antes de terminar essa parte introdutória, é conveniente estabelecer mais uma definição , a de eventos independentes. Ao jogarmos um dado, o resultado obtido numa jogada não afeta o resultado obtido na jogada posterior. A probabilidade de obter o número 2 em dada realização , independe do resultado das realizações anteriores. Então , os eventos associados ao ato de jogar o dado são *eventos independentes.* Nas seções a seguir, discutiremos algumas distribuições de probabilidade muito importantes, as quais se aplicam a eventos independentes.

Distribuição Binomial

Vejamos agora uma distribuição de probabilidades muito importante, a *distribuição binomial.* Ela se aplica a situações em que há apenas dois eventos independentes possíveis ($N_e = 2$). Exemplos: como usar uma moeda para jogar cara/coroa; sortear valores binários (0 ou 1); um carro de brinquedo num trilho plano que move-se para frente ou para trás ao receber impulsos aleatórios por controle remoto. Chamemos os eventos genericamente de A e B e sejam p e q as probabilidades de A e B ocorrerem, respectivamente. Como apenas esses dois eventos são possíveis, sabemos que $p + q = 1$. A distribuição binomial nos dá então a probabilidade de, após N realizações , o evento A ocorrer n_A vezes:

$$p_N(n_A) = \frac{N!}{n_A!(N - n_A)!} \; p^{n_A} \; q^{N-n_A}$$

Como o experimento permite apenas dois eventos possíveis, sabemos que $N = n_A + n_B$ ou $n_B = N - n_A$. Ou seja, ao dar a probabilidade de que ocorram n_A realizações de um dos eventos, a distribuição binomial nos dá também a probabilidade de que o número de realizações do outro evento seja $n_B = N - n_A$.

Aqui cabe uma outra observação muito importante. Temos que fazer uma diferenciação formal entre a situação real que comporta apenas 2 eventos (A e B) e o experimento, tal como definido na seção introdutória, ao qual a distribuição binomial se aplica. A variável dessa distribuição é n_A. Ou seja, a distribuição binomial lida com um conjunto de eventos definido no domínio $[0, N]$ e o experimento pode ser pensado como um grande número de realizações da situação real, do qual resulta então um evento de valor n_A.

Vamos determinar os momentos de ordens $n = 0$, $n = 1$ e $n = 2$ da distribuição binomial.

Na verdade, pode-se provar que

$$(p+q)^N = \sum_{n_A=0}^{N} \frac{N!}{n_A!(N-n_A)!}\ p^{n_A}\ q^{N-n_A} = \mu_0$$

Ou seja, $(p+q)^N$ é a normalização da função binomial. Mas como $p+q=1$, temos então que $(p+q)^N = 1^N = 1$ para qualquer N. Dessa forma, $\mu_0 = 1$, ou seja, a distribuição binomial é normalizada para a unidade.

Sejam agora os demais momentos, como a média e a variância. Para deduzir expressões para eles, expressaremos os momentos de ordem n como uma soma infinita, ao invés de uma integral, até porque os valores para a variável n_A são inteiros.

$$\mu_n = \sum_{n_A=0}^{\infty} n_A^n\ C_{n_A}^N\ p^{n_A}\ (1-p)^{N-n_A},$$

onde $C_{n_A}^N$ é o número de combinações envolvendo N realizações e contendo n_A termos. Acharemos uma fórmula de recorrência, que nos dará os momentos de ordem mais alta em função dos de mais baixa. Para isso, diferenciamos a expressão acima para p:

$$\frac{d\mu_n}{dp} = \sum_{n_A=0}^{\infty} n_A^n\ C_{n_A}^N\ \left[n_A\ p^{n_A-1}\ (1-p)^{N-n_A} - (N-n_A)\ p^{n_A}\ (1-p)^{N-n_A-1}\right]$$

$$\frac{d\mu_n}{dp} = \frac{1}{p}\ \sum_{n_A=0}^{\infty} n_A^{n+1}\ C_{n_A}^N\ p^{n_A}\ (1-p)^{N-n_A}\ -\ \frac{N}{q}\ \sum_{n_A=0}^{\infty} n_A^n\ C_{n_A}^N\ p^{n_A}\ (1-p)^{N-n_A}$$

$$+\ \frac{1}{q}\ \sum_{n_A=0}^{\infty} n_A^{n+1}\ C_{n_A}^N\ p^{n_A}\ (1-p)^{N-n_A}$$

$$\frac{d\mu_n}{dp} = \frac{1}{p}\ \mu_{n+1}\ -\ \frac{N}{q}\ \mu_n\ +\ \frac{1}{q}\ \mu_{n+1}$$

Resolvendo para μ_{n+1}, temos:

$$\left(\frac{1}{p}+\frac{1}{q}\right)\ \mu_{n+1} = \frac{N}{q}\ \mu_n\ +\ \frac{d\mu_n}{dp}$$

$$\frac{\mu_{n+1}}{pq} = \frac{N}{q}\ \mu_n\ +\ \frac{d\mu_n}{dp}$$

$$\mu_{n+1} = Np\ \mu_n\ +\ pq\ \frac{d\mu_n}{dp}$$

Usando $\mu_0 = 1$, temos então :

$$\mu_1 = Np\ \mu_0\ +\ pq\ \frac{d\mu_0}{dp} = Np$$

Para $n = 2$ temos então :

$$\mu_2 = Np\ \mu_1\ +\ pq\ \frac{d\mu_1}{dp} = N^2p^2\ +\ Npq$$

Usando a fórmula da variância, teremos finalmente:

$$\sigma^2 = \mu_2 - \mu_1^2 = N^2p^2\ +\ Npq - N^2p^2 = Npq$$

Distribuição de Poisson

Consideremos agora outra distribuição de probabilidades, a *distribuição de Poisson*. Ela resulta da distribuição binomial quando o número de realizações N tende a infinito ao mesmo tempo em que a probabilidade de um dos 2 eventos do binômio, p, tende a zero, de forma a manter constante o produto $\mu_1 = Np$. Neste limite, pode-se provar que a distribuição binomial se torna

$$p_N(n_A) = \frac{\mu_1^{n_A}}{n_A!}\ e^{-\mu_1}$$

Essa é a expressão para a distribuição de Poisson. Podemos agora calcular os momentos da distribuição de Poisson.

$$\mu_0 = \sum_{n_A=0}^{\infty} \frac{\mu_1^{n_A}}{n_A!}\ e^{-\mu_1}$$

$$\mu_0 = e^{-\mu_1}\ \sum_{n_A=0}^{\infty} \frac{\mu_1^{n_A}}{n_A!}$$

Mas a soma infinita do lado direito da última igualdade é justamente a expansão em série de Taylor de e^{μ_1}. Logo temos:

$$\mu_0 = e^{-\mu_1}\ e^{\mu_1} = 1$$

Assim, de forma semelhante ao que ocorre para a distribuição binomial, a de Poisson já está normalizada para a unidade.

A média de uma distribuição de Poisson é simplesmente igual à média da binomial, já que a primeira resulta da segunda no limite de realizações infinitas, mas mantendo $\mu_1 = Np$ inalterado. Mas podemos deduzir uma fórmula de recorrência para a distribuição de Poisson e confirmar este resultado.

$$\frac{d\mu_n}{d\mu_1} = \sum_{n_A=0}^{\infty} n_A^n \left[n_A \ \frac{\mu_1^{n_A-1}}{n_A!} e^{-\mu_1} - \frac{\mu_1^{n_A}}{n_A!} e^{-\mu_1} \right]$$

$$\frac{d\mu_n}{d\mu_1} = \frac{1}{\mu_1} \sum_{n_A=0}^{\infty} n_A^{n+1} \ \frac{\mu_1^{n_A}}{n_A!} e^{-\mu_1} \ - \ \sum_{n_A=0}^{\infty} n_A^n \ \frac{\mu_1^{n_A}}{n_A!} e^{-\mu_1}$$

$$\frac{d\mu_n}{d\mu_1} = \frac{1}{\mu_1} \mu_{n+1} \ - \ \mu_n$$

$$\mu_{n+1} = \mu_1 \frac{d\mu_n}{d\mu_1} \ + \ \mu_1 \mu_n$$

Usando agora que $\mu_0 = 1$, de forma que sua derivada seja nula, temos

$$\mu_1 = \mu_1 \mu_0 = \mu_1$$

Finalmente, para o momento de 2a ordem, teremos:

$$\mu_2 = \mu_1 + \mu_1^2$$

A variância será então dada por

$$\sigma^2 = \mu_2 - \mu_1^2 = \mu_1 = Np$$

Ou seja, a variância em torno da média na distribuição de Poisson é igual à própria média.

Aplicação astronômica da distribuição de Poisson:

contagem de fótons num detetor

A contagem de fótons num detetor quântico, como um CCD, segue a distribuição de Poisson. Isso porque fontes astronômicas são fracas, de forma que poucos fótons chegam ao detetor. Ou seja, se considerarmos p como sendo a probabilidade de um fóton atingir o detetor num instante, temos justificado o limite em que p tende a zero. Por serem fontes fracas, os tempos de integração são longos, o que significa um grande número N, tendendo a infinito, de instantes (realizações) para a incidência de um fóton. Finalmente, esperamos que uma fonte emita, em média, a mesma quantidade de fótons, o que é válido para qualquer fonte não variável durante um tempo de exposição . Assim, $\mu_1 = Np$ é constante.

Note que o valor esperado para a contagem de fótons, N_f, proveniente de uma fonte num dado tempo de exposição , será também constante,

$$N_f = \frac{dn}{dt}\, t_e,$$

onde dn/dt é a contagem por unidade de tempo (em unidade, por exemplo, de fótons/s) e t_e é o tempo de exposição . Mas se fizermos várias exposições com o mesmo tempo de integração , vamos ver que esta contagem varia, de acordo com a distribuição de Poisson. A flutuação típica no número de fótons será dada então pela raiz quadrada da variância. Esta última, no caso da distribuição de Poisson, é igual ao valor esperado.

$$\sigma^2 = N_f$$

$$\sigma = \sqrt{N_f}$$

Essa flutuação representa uma forma de ruído na contagem de fótons, ruído este associado ao próprio sinal gerado no detetor, ou seja, à própria contagem de fótons. Em resumo, enquanto a contagem de fótons num dado tempo de exposição cresce linearmente com t_e, o ruído Poissônico a ela associado cresce com $\sqrt{t_e}$.