O teorema de Bayes

- O teorema de Bayes é uma igualdade simples que vem da afirmação de que $prob(A \text{ e } B) = prob(B \text{ e } A)$:

$$prob(B|A) = \frac{prob(A|B)\ prob(B)}{prob(A)}, \tag{4}$$

  no qual o denominador é a probabilidade total.

- Esse teorema é útil quando interpretado como uma regra para indução: os dados e o evento A são considerados como sucessores de B, o grau de crença anterior a realização do experimento.

- Assim sendo $prob(B)$ é chamado de *probabilidade a priori* a qual será modificada pela experiência.

- A experiência é determinada pela *verossimilhança* $prob(A|B)$

- Finalmente, $prob(B|A)$ é a *probabilidade posterior*, ou o nível de crença após a realização do experimento.

- A primeira vista o teorema parece trivial mas seu poder reside na sua interpretação.

O teorema de Bayes

- Considere agora o seguinte exemplo: Suponha uma caixa com cinco moedas, uma das quais é "trucada" e tem 'cara' dos dois lados. Toma-se uma moeda dessa caixa ao acaso e esta é jogada três vezes. Observa-se que em todos casos o resultados foi sempre 'cara'. **Qual é a probabilidade da moeda escolhida ser a 'trucada' dado esses resultados?**
- Sabendo-se de antemão qual é o tipo de moeda é simples calcular a probabilidade de obtermos três 'caras' (*Heads*). No caso da moeda trucada isso é: $P(H|M) = 1$, onde H denota as três 'caras' e M é o evento da moeda especial ter sido a escolhida.
- No caso da moeda comum temos: $P(H|M^c) = (1/2)^3 = 1/8$, onde M^c é o evento *complementar* a M (moeda normal ; $M + M^c \equiv 1$).
- O denominador do teorema de Bayes pode ser encontrado a partir da soma ponderada de todas as probabilidades possíveis:

$$P(A) = P(A|B_1)P(B_1) + ... + P(A|B_n)P(B_n),$$

que no presente caso é: $P(A) = P(H|M)P(M) + P(H|M^c)P(M^c)$

- Podemos agora escrever o teorema da Bayes e obter a resposta à pergunta:

$$P(M|H) = \frac{P(H|M)P(M)}{P(H|M)P(M) + P(H|M^c)P(M^c)} = \frac{1/5}{1 \times (1/5) + (1/8) \times (4/5)} = \frac{2}{3}$$

O teorema de Bayes

- Vejamos agora outro exemplo
- Suponhamos o exemplo de bolas coloridas, M vermelhas e N brancas) numa urna. A probabilidade de tirarmos 3 vermelhas e 2 brancas, segundo o teorema da Bayes, é:

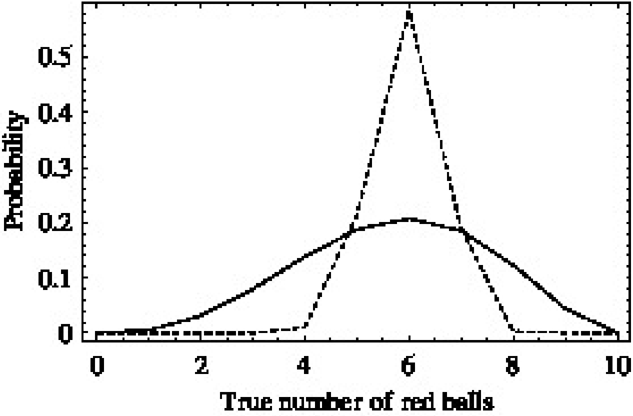
$$prob(\textbf{conteúdo da urna}|\textbf{dados}) \propto prob(\textbf{dados}|\textbf{conteúdo da urna}),$$

onde o lado direito pode ser calculado, segundo algumas suposições.

- **Nota:** O que pode ser difícil se dar conta é que, em geral, temos o problema inverso. Temos os dados (três bolas vermelhas e duas brancas) e queremos inferir o conteúdo da urna. Esse pode ser chamado de "probabilidade inversa".

O teorema de Bayes

- **Exemplo W&J P. 27**

- 

O teorema de Bayes

- Nota-se que em ambos casos o pico das curvas se dá em 6, o que parece trivial. Notem porém que estamos falando da probabilidade da urna contar uma bola vermelha, duas, três etc.
- Estamos descrevendo nossa crença sobre os conteúdos da urna baseado naquilo que vemos (os dados a as informações dada *a priori*).
- O ponto chave dessa discussão é que nós fomos bem sucedidos em responder à questão científica colocada: fizemos uma inferência sobre o conteúdo da urna e podemos fazer afirmações probabilísticas sobre essa inferência.
- Por exemplo, no primeiro caso, a probabilidade da urna conter 3 ou menos bolas vermelhas é de 11.5%.

O teorema de Bayes

- *O Teorema da Bayes nos permite fazer inferências a partir dos dados, ao invés de apenas computar os dados que teríamos caso nós tivéssemos toda a informação relevante sobre o problema.*
- Essa diferenciação soa bastante acadêmica mas, suponhamos que temos duas amostras e queremos determinar se suas médias são diferentes. Vários livros de estatística respondem à questão oposta: dadas duas populações com diferentes médias, que dados você obteria?
- A combinação da interpretar probabilidades como uma medida consistente de crença, mais o teorema de Bayes, permite que se responda à questão que de fato estamos interessados: dados os dados, quais são as probabilidades dos parâmetros contidos no nosso modelo estatístico?

Distribuições de Probabilidade: conceito

- Considere o experimento de jogar quatro moedas (justas). a probabilidade de obter-se zero 'caras' é de $(1/2)^4$; uma 'cara' $4 \times (1/2)^4$; duas 'caras' $6 \times (1/2)^4$ etc.

- A soma das probabilidades de obtermos desde zero 'caras' até quatro 'caras' deve somar 1.0.

- Se x é o número de caras (0,1,2,3,4), temos um conjunto de probabilidades prob(x) $=$ (1/16, 1/4, 3/8, 1/4, 1/16); Isso é uma *distribuição probabilística*.

- Em particular, essa distribuição é discreta *i.e.* existe um conjunto discreto de probabilidades para os resultados do experimento. Nesse tipo de caso temos um mapeamento entre os resultados do experimento e um conjunto de inteiros.

Distribuições de Probabilidade: conceito

- Em outros casos o conjunto de resultados possíveis é mapeado em números reais. Então as distribuições probabilísticas são contínuas.
- É possível lidar com essas distribuições 'discretizando' o intervalo de números reais em pequenos intervalos onde se assume que não há variação da probabilidade.
- Assim, se x é o número real que indica os resultados, podemos associa-lo com uma densidade de probabilidade $f(x)$.
- Então, a probabilidade de obtermos um número próximo a x, dentro do intervalo δx será de $prob(x) \approx f(x)\delta x$.

Distribuições de Probabilidade: conceito

- Formalmente: se x é uma variável aleatória contínua, então $f(x)$ é uma *função densidade de probabilidade (FDP)*, ou uma *distribuição probabilística*, quando:

  1. $prob[a < x < b] = \int_a^b f(x)dx$;
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$ e
  3. $f(x)$ é unívoca e não-negativa para qualquer x real.

- A FDP *cumulativa* correspondente é: $F(x) = \int_{-\infty}^{x} f(y)dy$

- As FDPs podem serem também *multivariadas*, ou seja, funções de mais de uma variável: $f(x, y, z)$

Algumas distribuições mais comuns

- Na tabela 2.1 do W&J podem ser encontradas algumas das mais conhecidas FDPs, bem como os respectivos valores de posição (centro) e escala (dispersão).

- Esses valores correspondem ao dois primeiros momentos das distribuições:

$$\mu_1(media) = \mu = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$\mu_2(variancia) = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu_1)^2\, f(x)\, dx,$$

  onde a raiz quadrada da variância, σ é conhecida como *desvio padrão*.

- A seguir veremos quatro dessas distribuições, binomial, de Poisson e gaussiana ou Normal e por lei de potência.

Distribuição binomial

- Aplica-se nos casos onde existam apenas dois resultados possível – 'sucesso' ou 'falha'
- Essa distribuição dá a probabilidade de n sucessos em N tentativas, onde a probabilidade de êxito em cada tentativa é a mesma e designada por p, sendo que as sucessivas tentativas são independentes entre si.
- Essa probabilidade então é:

$$prob(n) = \binom{N}{n} p^n (1-p)^{N-n}$$

- O primeiro termo, o coeficiente combinatório ou binomial, dá o número de possibilidades de obter-se n sucessos em N tentativas:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

- Esse coeficiente pode ser derivado da seguinte forma. Existem $N!$ modos equivalentes de se arranjar N tentativas. Por outro lado, existem $n!$ permutações dos sucessos e $(N-n)!$ permutações das falhas, que correspondem ao mesmo resultado – ou seja n sucessos arranjados de quaisquer modos.
- Como n sucessos (probabilidade p^n) não são suficientes, mas sim *exatamente* n sucessos, precisamos também de $N-n$ falhas (probabilidade $p^{(N-n)}$).

Distribuição binomial

- $$prob(n) = \frac{N!}{n!(N-n)!} p^n q^{N-n} \; ; \; q = 1 - p$$

- A distribuição de probabilidade binomial

		n				
		0	1	2	3	4
N	1	q	p			
	2	q^2	$2pq$	p^2		
	3	q^3	$3pq^2$	$3p^2q$	p^3	
	4	q^4	$4pq^3$	$6p^2q^2$	$4p^3q$	p^4

- $$(p+q)^N \equiv 1$$

Distribuição binomial

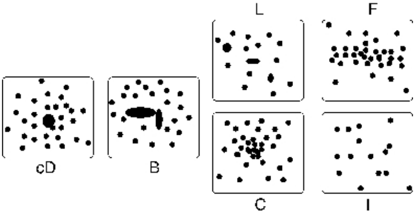
- A distribuição binomial tem um valor médio dado por:

$$\sum_{n=0}^{N} n\, prob(n) = Np$$

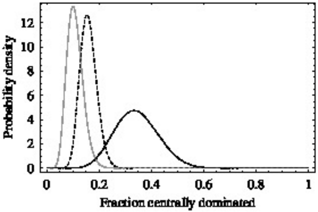
- e uma variância (ou média quadrática) de:

$$\sum_{n=0}^{N} (n - Np)^2\, prob(n) = Np(1 - p) = Npq$$

Distribuição binomial



- Exemplo W&J P. 35

- 

- A distribuição binomial dá origem a outras duas distribuições famosas, a de Poisson e a gaussiana.

Distribuição de Poisson

- A distribuição de Poisson deriva da distribuição binomial no caso limite de eventos (independentes) muito raros em um grande número de tentativas. de modo que, embora $p \rightarrow 0$, $Np \rightarrow$ tem um valor finito.

- Chamando esse valor finito da média de $\mu_1 = \mu$, então a distribuição de Poisson é:

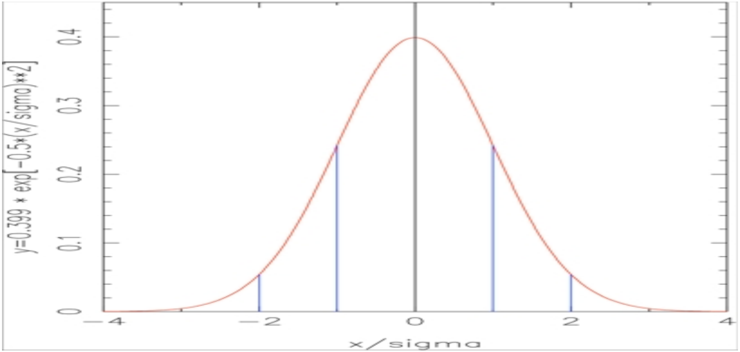
$$prob(n) = \frac{\mu^n}{n!} e^{-\mu}$$

- A variância dessa distribuição, μ_2 é também μ.

Distribuição Normal ou gaussiana

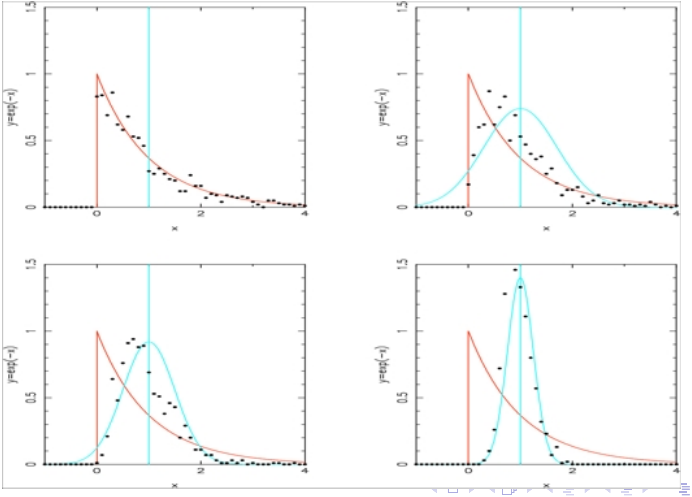
- Ambas distribuições, binomial e de Poisson tendem à gaussiana quando *n* (no caso da distribuição binomial) ou μ (no caso de Poisson) são grandes.
- A distribuição gaussiana ou Normal (univariada) tem a seguinte forma:

$$prob(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$



Distribuição Normal. A área total sob a cruva é 1,00; dentro de $\pm 1\sigma$ é (aproximadamente) 0.68; 0.95 dentro de $\pm 2\sigma$ e 0.997 dentro de $\pm 3\sigma$.

Teorema do valor central ou Porque a Gaussiana é tão importante?

- Definição não-rigorosa do Teorema do Valor Central: dado um conjunto de médias M_n formado a partir da geração de n amostras de uma população x_i com média finita μ e variância σ^2, então sua distribuição tenderá a uma gaussiana com média$=\mu$ e desvio padrão $= \sigma/\sqrt{n}$.
- 

Teorema do valor central ou Porque a Gaussiana é tão importante?

- O teorema do valor central é fundamental nessa área. O que ele diz, em outras palavras, é que quando algumas condições são obedecidas (e elas o são a maioria das vezes), fazer médias produz como resultado uma distribuição gaussiana, *não importando qual é a forma da distribuição a partir da qual a distribuição é gerada.*
- Isso significa que os erros de médias de dados vão sempre ser 'gaussianos'.
- Isso também é o responsável por descrevermos resultado em termos de 'sigmas'.
- Levando isso em conta, um resultado de 2σ teria apenas 5% de ocorrer apenas por acaso. Isso deveria ser suficiente, mas a comunidade científica sempre olha esse tipo de resultado com grande desconfiança. Isso ocorre porque a determinação dos erros (sigmas) também é incerta.
- Um outro ponto que (raramente levado em consideração) é que outra propriedade importante do TVC é que a convergência ocorre rapidamente no centro mas muito mais lentamente nas asas.

Distribuição em lei de potência

- Um outro tipo de distribuição que aparece frequentemente na vida do astrônomo é a distribuição em lei de potência.

- Tome N como o número de objetos ou evento com uma dada propriedade medida (e.g. a luminosidade) que ou é maior do que maior do que um certo valor L (forma integral) ou que está dentro do intervalo dL centrado em L. Sendo γ o expoente de lei de potência temos:

$$N(> L) = K\, L^{\gamma+1} \text{ (forma integral)}$$

ou

$$dN = (\gamma + 1)\, K\, L^{\gamma}\, dL \text{ (forma diferencial)}$$

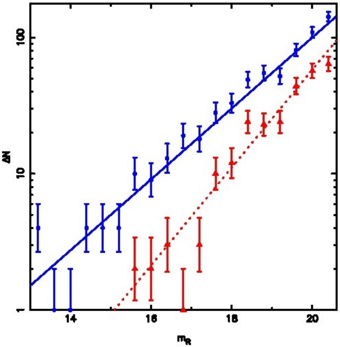
- Essa é uma distribuição *independente de escala*, porque se $f(x) = x^{\gamma}$, então $f(ax) = a^{\gamma}\, x^{\gamma} = Cte.\, x^{\gamma} = Cte.\, f(x)$, que é a definição de independência de escala.

Distribuição em lei de potência

- Essa distribuição difere grandemente das vistas anteriormente.
- Por exemplo, suas média e variância são infinitas, a menos que limites sejam determinados e isso é o que ocorre na maioria dos casos concretos, pois sempre existem limitações físicas em ambos 'lados' da distribuição.
- Lei de potência aparecem em diversos fenômenos cotidianos tais como flutuações no mercado econômico, taca de crescimento de empresas, distribuição de salários e em distribuições de tamanho, incluindo avalanches, terremotos e incêndios florestais.
- De modo geral o expoente é negativo. Tomando os salários como exemplo, há muito mais pessoas com baixos salários do que com com altos.
- Em astronomia encontramos esse tipo de distribuição muito frequentemente, por exemplo: a função de massa inicial de Salpeter, funções de massa, magnitude ou contagens de objetos, funções de luminosidade,a espectro de flutuações primordial e muito mais.
- Aqui também o expoente é sempre negativo. Há muito mais objetos de baixa luminosidade/massa/etc. do que de alta.

Distribuição em lei de potência

Exemplo de distribuição em lei de potência (W&J P. 44)



Contagens de estrelas (triângulos vermelhos) e todos os objetos (círculos azuis) numa imagem de uma dada área do céu. Os dados estão em intervalos de 0,4 magnitudes e são diferenciais (i.e. não cumulativos). Trata-se de uma lei de potência porque a magnitude é uma escala logarítmica inversa: $m_1 - m_2 = -2.5 \log_{10}(f_1/f_2)$, onde f é o fluxo.

Distribuição em lei de potência

- Estudos de objetos distribuídos em lei de potência podem estar sujeitas a grandes viézes, especialmente se houver erros de medida relativamente grandes.

- Outras questões as quais se deve estar atento:

1. Essa lei de potência é diferencial ou integral? Essa é uma forma comum de se errar no expoente por uma unidade.
2. A divisão em intervalos ("binagem") é linear ou logarítmica? Se uma distribuição diferencial é dividida numa escala uniforme em $\Delta log\ L$, ao invés de uniforme em L, o valor da inclinação é reduzido de uma unidade.
3. Tenha muito cuidado ao usar média e desvios padrão baseado em limites para a função exponencial em ambos lados. Por exemplo, supondo que esses limites sejam a e b temos que o primeiro momento (média) é dados por:

$$\mu = \left(\frac{\gamma+1}{\gamma+2}\right)\left[\frac{b^{\gamma+2}-a^{\gamma+2}}{b^{\gamma+1}-a^{\gamma+1}}\right]$$

   (Essa expressão não funciona para $\gamma =$ -1 ou -2, mas expressões podem ser derivadas para esses casos especiais.)
   De toda forma mostra como o resultado depende fortemente dos limites impostos. Isso é ainda mais impróprio para a variância, dada a natureza altamente assimétrica da lei de potência.