Conceitos da Teoria de Probabilidade

Física Estatística - F 604

Silvio A. Vitiello

Conjunto

- Se $S = \{A_1, A_2, \ldots, A_n\}$ é o conjunto de *todas* possibilidades que um experimento pode dar, S é chamado de *espaço amostral* deste experimento e cada um de seus elementos de ponto amostral ou evento elementar.
- Empregamos a palavra *amostral* para aludir a origem estatística dos termos.
- Uma coleção parcial de elementos de S é chamada de *subconjunto*. Casos extremos destes *conjuntos*:
    - O conjunto é o próprio S
    - O conjunto não possui nenhum ponto __ *conjunto vazio* __ $\emptyset$
- Um *evento* é um subconjunto do espaço amostral de um experimento.
- Um evento não é elementar se a ele estiver associado mais do que um elemento ou evento de S.

Operações

- $A \cup B$ denota a união de dois conjuntos.

  É o conjunto de todos os pontos pertencentes a A ou B ou a ambos.
- $A \cap B$ denota a intersecção de dois conjuntos.

  É o conjunto de todos os pontos pertencentes a ambos os conjuntos.

Probabilidade

- Probabilidade é um número que resume a possibilidade de obtermos um certo evento em um experimento.
- Alternativamente: A probabilidade $P(A)$ de um evento A é a razão do número de casos favoráveis ao evento m pelo número total n de casos, com tanto que todos eles sejam equiprováveis: $P(A) = \frac{m}{n}$.
- Uma medida de probabilidade é uma função P que satisfaz:
  1. Se A é um possível evento de um experimento então o valor da função, sua probabilidade, é $P(A) \geq 0$
  2. Se dois eventos são mutuamente exclusivos $A \cap B = \emptyset$ o valor de P é $P(A \cup B) = P(A) + P(B)$
  3. $P(S) = 1$
- A *probabilidade* de um evento $A \subset S$ é obtida adicionando as probabilidades atribuídas aos elementos do subconjunto de S que correspondem a A.

Propriedades

- Se os eventos $A_1, A_2, \ldots, A_m$ são mutuamente *exclusivos* e *exaustivos*, então
$$A_1 \cup A_2 \cup \ldots \cup A_m = S$$
Os m eventos formam uma *partição* do espaço amostral S em m subconjuntos.
Se $A_1, A_2, \ldots, A_m$ formam uma partição, então
$$P(A_1) + P(A_2) + \cdots + P(A_m) = 1$$

- $P(A \cup B)$ é a probabilidade do evento A ou do evento B ou ambos ocorrerem.
$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$
$P(A \cap B)$ — a probabilidade de ambos eventos ocorrerem — deve ser subtraída, já que $P(A) + P(B)$ leva em conta a região $A \cap B$ duas vezes.

Eventos independentes

- Experimentos *independentes* correspondem a noção intuitiva de experimentos repetidos sob condições identicas
  - um sistema utilizado ao longo do tempo
  - diversos sistemas idênticos empregados em um mesmo instante $\rightarrow$ *ensemble*
- Se N experimentos idênticos e independentes (sem memória do aconteceu anteriormente) forem repetidos e $N \rightarrow \infty$, esperamos que a fração dos eventos A (N_A/N) aproxime-se de $P(A)$.
- Os eventos A e B são independentes se e somente se $P(A \cap B) = P(A)P(B)$.

  Eventos independentes não são mutuamente exclusivos onde temos $P(A \cap B = \emptyset) = 0$

Ensemble

- Na discussão da entropia, postulamos que são iguais as probabilidades de encontrar um sistema fechado em qualquer estado microscópico consistente com seus vínculos. Esta hipótese atribui estas probabilidades apelando a simetria do sistema.
- Um ensemble de sistemas é uma coleção de sistemas idênticos.
  Cada réplica do sistema no ensemble está em um dos estados acessíveis ao sistema.
- Numa visão originalmente proposta por Gibbs, podemos substituir médias temporais por médias em um ensemble.

Probabilidade condicional

- A *probabilidade condicional* do evento A ocorrer assumindo que B tenha ocorrido é denotada por $P(A|B)$ e dada por

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

- Uma vez que $P(A \cap B) = P(B \cap A)$:

$$P(A|B)P(B) = P(B|A)P(A).$$

- Se A e B são independentes: $P(A|B) = P(A)$
- A probabilidade condicional $P(A|B)$ é essencialmente a probabilidade do evento A se usamos o conjunto B como espaço amostral ao invés de S.

Contagem

- Sejam duas operações tais que uma pode ser feita de m maneiras e a outra em n.
  - Principio de Adição: Se as operações são mutuamente exclusivas então uma operação qualquer pode ser feita de $m+n$ modos.
  - Principio da Multiplicação: Se uma das operações é executada em uma de suas possíveis maneiras e a seguir a outra é executada em uma das suas próprias maneiras então as duas operações podem ser executadas de $m \times n$ maneiras.

Permutações e Combinações

- Uma *permutação* é um dos arranjos de um conjunto de objetos distintos em uma dada ordem. O número de diferentes permutações de um conjunto de N objetos é $N!$
- O número de permutações P_R^N de R objetos tomados de um conjunto de N elementos distintos é $P_R^N = \dfrac{N!}{(N-R)!}$
- Uma *combinação* é a seleção de alguns dos objetos de um conjunto formado de elementos distintos sem preocupação com a ordem da escolha. O número de combinações de R objetos tomados de um conjunto de N elementos é

$$C_R^N = \frac{N!}{R!(N-R)!}$$

O *Santo Império Romano* era um complexo de terras na Europa central e ocidental inicialmente regido pelos francos e depois por reis alemães. Da coroação de Carlos Magno em 800, perdurou por dez séculos.

Voltaire foi uma das mais influentes figuras do iluminismo francês e da Europa no Sec. XVIII, filósofo, poeta, dramaturgo, ensaísta, historiador e escritor.

Variáveis aleatórias

Segundo Voltaire: O Santo Império Romano não era santo, nem império e nem romano. Similarmente uma variável aleatória não é aleatória nem uma variável:

Uma *variável aleatória* é uma função definida em um espaço amostral.

- Alternativamente: Uma quantidade cujo valor é um número determinada por um evento de um experimento é chamada *variável aleatória*.
- Uma variável aleatória X em um espaço amostral S (contável) é uma função numérica que mapeia os elementos de S.
- As possíveis realizações da v.a. X são denotadas por $\{x_i\}$.

Distribuição e valor esperado

- A probabilidade de conjunto de pontos amostrais pode ser determinada pelos valores de variáveis aleatórias

$$\{a \leq X \leq b\} = \{A | a \leq X(A) \leq b\}$$

- Notação $p_n = P(X = x_n)$
- $P(a \leq X \leq b) = \sum_{a \leq x_n \leq b} p_n$
- Função de distribuição de X: $F_X(x) = P(X \leq x) = \sum_{x_n \leq x} p_n$
- $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$
- Valor médio ou esperado: $\langle X \rangle = E(X) = \sum_n X(A_n) P(A_n) = \sum_n x_n P(A_n) = \mu$
- Segundo momento: $E(X^2) = \sum_n X^2(A_n) P(A_n)$

Variáveis aleatórias com densidades

- Função densidade
  1. $\forall x: \qquad f(x) \geq 0$
  2. $\int_{-\infty}^{\infty} f(x)dx = 1$
- $P(a \leq X \leq b) = \int_a^b f(x)dx$
- Função de distribuição de X: $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)du$
- $F_X'(x) = f(x)$
- Valor médio ou esperado: $\langle X \rangle = E(X) = \int_{-\infty}^{\infty} x f(x)dx = \mu$
- Segundo momento: $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$

Variância

$$\mathrm{Var}\ (X) = E(X^2) - E(X)^2 = E((X-\mu)^2)$$

Medida de quanto os valores estão espalhados em torno da média

Distribuição Binomial

- Aplicação: Um número N de experimentos ou de sistemas e dois eventos possíveis
- Experimentos/sistemas independentes
- p é a probabilidade de um dos eventos ocorrer (coroas, spin para cima, etc)
- $q = 1 - p$ é probabilidade do outro evento
- Dois estados para cada sistema $\Rightarrow 2^N$ é o número total de estados possíveis
- $P(\text{um particular estado com } n \text{ spins para cima}) = p^n q^{N-n}$
- $P(\text{de qualquer estado com } n \text{ spins para cima}) = \binom{N}{N-n} p^n q^{N-n}$
- Normalização: $1 = (p+q)^N = \sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n}$

Valor médio da distribuição binomial

Se uma medida de y produz n

- $\langle y \rangle = \sum_{n=0}^{N} n \binom{N}{n} p^n q^{N-n}$
  - $\frac{\partial (p+q)^N}{\partial p} = \frac{\partial \sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n}}{\partial p} \quad \Rightarrow \quad N(p+q)^{N-1} = \sum_{n=0}^{N} n \binom{N}{n} p^{n-1} q^{N-n}$
  - $Np(p+q)^{N-1} = \sum_{n=0}^{N} n \binom{N}{n} p^n q^{N-n}$
- $p+q=1 \quad \Rightarrow \quad Np = \sum_{n=0}^{N} n \binom{N}{n} p^n q^{N-n} = \langle y \rangle$
- $\mathrm{Var}\ (\mathrm{y}) = \mathrm{Npq} \quad \Rightarrow \quad \sigma_{\mathrm{N}} = \sqrt{\mathrm{Npq}}$ e portanto $\frac{\sigma_N}{\langle y \rangle} = \sqrt{\frac{q}{p}} \frac{1}{\sqrt{N}}$
- Se $N \approx N_0$ então $\sigma_N / \langle y \rangle \approx 10^{-12}$ (desvio relativo da média)
  - Portanto apesar dos parâmetros macroscópicos serem variáveis aleatórias, suas flutuações são tão pequenas que podem ser ignoradas.
  - Exemplo: falamos da energia de um sistema em contato com um reservatório térmico, apesar desta grandeza ser uma variável aleatória que flutua constantemente.

Distribuição Gaussiana (Normal)

A distribuição Gaussiana é a mais comum e importante nas aplicações.

O teorema do limite central garante que a soma (ou a média) de um grande número de v.a. distribuídas de forma idêntica e independente aproxima-se da distribuição Gaussiana.

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

onde $\langle x \rangle = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \mu$, $\quad$ $\mathrm{Var}\,(\mathrm{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (\mathrm{x}-\mu)^2 e^{-\frac{(\mathrm{x}-\mu)^2}{2\sigma^2}}\, \mathrm{dx} = \sigma^2$

Distribuição normal padrão: $\mu = 0$ e $\sigma = 1$

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Equivalentemente $x \to x - \mu$ e a seguir $x \to x/\sigma$.

Função de Distribuição

$$
\begin{aligned}
F_X(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}} \, du \\
&= \frac{1}{2} + \frac{1}{2}\operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)
\end{aligned}
$$

Função erro

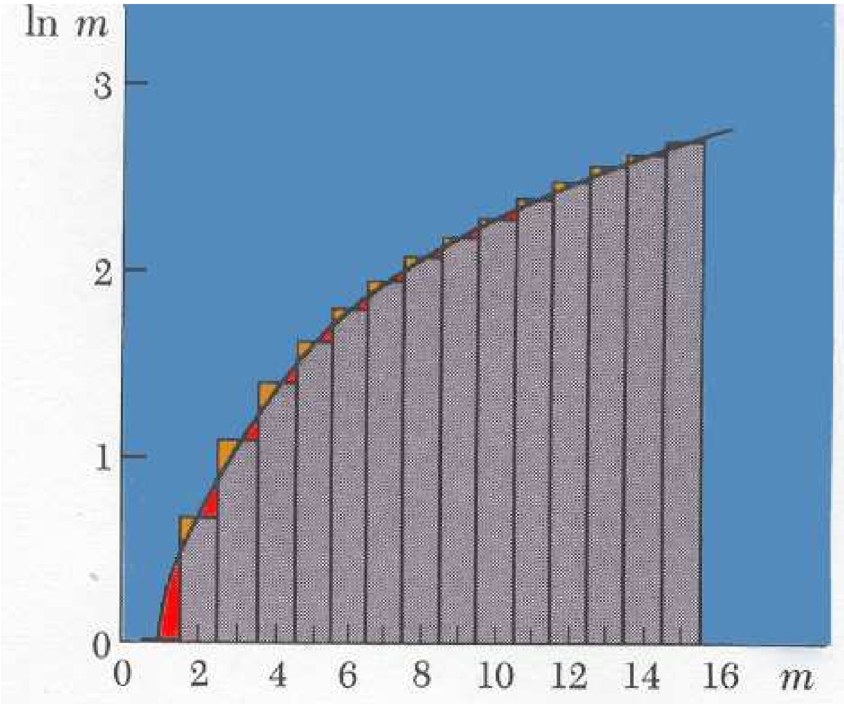
$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-u^2} \, du$$

Função erro complementar

$$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} e^{-u^2} \, du$$

erf(x) + erfc(x) = 1

Valor de ln *n! para n grande*



$$n! = 1 \cdot 2 \cdot \dots \cdot n$$

$$\ln n! = \sum_{m=1}^{n} \ln m$$

$$\ln n! \approx \int_1^n \ln x dx = x \ln x - x|_1^n$$

Se $n >> 1$ então:

$$\boxed{\ln n! \approx n \ln n - n}$$

Aproximação de Stirling: $\ln n! \approx \frac{1}{2}\ln(2\pi n) + n\ln n - n$

Função gama

$$\Gamma(z) = \int_0^\infty t^{z-1} \mathrm{e}^{-t} dt$$

Integrando por partes:
$$\begin{aligned}\Gamma(z) &= -t^{z-1}\mathrm{e}^{-t}\Big|_0^\infty + (z-1)\int_0^\infty t^{z-2}\mathrm{e}^{-t} dt \\ &= (z-1)\int_0^\infty t^{z-2}\mathrm{e}^{-t} dt \\ &= (z-1)\Gamma(z-1) \Rightarrow \boxed{n! = \Gamma(z+1)}\end{aligned}$$

$$I_m = 2\int_0^\infty x^m \mathrm{e}^{-x^2} dx\,; \quad (m > -1) \xrightarrow[2dx=dy/\sqrt{y}]{x^2=y} \quad I_m = \int_0^\infty y^n \mathrm{e}^{-y} dy = \Gamma(n+1)\,; \quad \left(n = \frac{m-1}{2}\right)$$

$$I_{2\ell} = \Gamma(\ell + 1/2) = (\ell - \frac{1}{2})(\ell - \frac{3}{2})\cdots\frac{1}{2}\sqrt{\pi}; \quad (\Gamma(1/2) = \sqrt{\pi})$$

$$I_{2\ell+1} = \Gamma(\ell+1) = \ell!$$

Distribuição Gaussiana como limite da Binomial

Sistemas com $N >> 1$ spins, $N_{\parallel}, N_{\nparallel} >> 1 \Rightarrow 2s \not\approx \pm N$. Portanto, s não precisa ser grande

$$\left.\begin{array}{rcl} N_{\parallel} - N_{\nparallel} & = & 2s \\ N_{\parallel} + N_{\nparallel} & = & N \end{array}\right\} \begin{array}{rcl} N_{\parallel} & = & N/2 + s \\ N_{\nparallel} & = & N/2 - s \end{array}$$

$$p_s = \frac{N!}{(N/2+s)!(N/2-s)!} p^{\frac{N}{2}+s} q^{\frac{N}{2}-s}$$

Aproximação de Stirling: $n! \approx \sqrt{2\pi N} N^N \mathrm{e}^{-N}$

$$\begin{aligned}
p_s &\approx \sqrt{\frac{2\pi N}{2\pi(N/2+s)2\pi(N/2-s)}} \frac{N^N}{(N/2+s)^{(N/2+s)}(N/2-s)^{(N/2-s)}} \mathrm{e}^0 \, p^{\frac{N}{2}+s} q^{\frac{N}{2}-s} \\
&= \sqrt{\frac{1}{2\pi N(1/2+s/N)(1/2-s/N)}} \frac{1}{(1/2+s/N)^{(N/2+s)}(1/2-s/N)^{(N/2-s)}} \, p^{\frac{N}{2}+s} q^{\frac{N}{2}-s} \\
&= \sqrt{\frac{1}{2\pi N}} \left(\frac{p^{\frac{N}{2}+s}}{(1/2+s/N)^{(N/2+s+1/2)}} \right) \left(\frac{q^{\frac{N}{2}-s}}{(1/2-s/N)^{(N/2-s+1/2)}} \right) \\
&= \sqrt{\frac{1}{2\pi N pq}} \left(\frac{p}{1/2+s/N} \right)^{(N/2+s+1/2)} \left(\frac{q}{1/2-s/N} \right)^{(N/2-s+1/2)} \\
&\approx \sqrt{\frac{1}{2\pi N pq}} \left(\frac{p}{1/2+s/N} \right)^{(N/2+s)} \left(\frac{q}{1/2-s/N} \right)^{(N/2-s)} \text{ pois } \frac{N}{2} \pm s + \frac{1}{2} \approx \frac{N}{2} \pm s
\end{aligned}$$

Distribuição Gaussiana como limite da Binomial

Esta função possui um pico pronunciado em torno do valor médio de s

$\langle s \rangle = \mu = \langle N_{\|} \rangle - N/2 = N(p-1/2)$, já que $\dfrac{\sigma_N}{\langle y \rangle} \propto N^{-1/2}$

Podemos ver isto explicitamente. É conveniente tomar o ln de

$$p_s = \sqrt{\frac{1}{2\pi N p q}} \left(\frac{p}{1/2 + s/N} \right)^{(N/2+s)} \left(\frac{q}{1/2 - s/N} \right)^{(N/2-s)}$$

$$\ln p_s = -\frac{1}{2} \ln(2\pi N p q) + (\frac{N}{2} + s)\,(\ln p - \ln(1/2 + s/N)) + (\frac{N}{2} - s)\,(\ln q - \ln(1/2 - s/N))$$

$$\frac{\partial \ln p_s}{\partial s} = \ln p - \ln(1/2 + s/N) - 1 - \ln q + \ln(1/2 - s/N) + 1$$

Conforme afirmamos: $\left.\dfrac{\partial \ln p_s}{\partial s}\right|_{s=\mu} = 0$ Vamos expandir $\ln p_s$ em torno de $s = \mu$

$$\frac{\partial^2 \ln p_s}{\partial s^2} = -\frac{1}{N/2 + s} - \frac{1}{N/2 - s}$$

$$\left.\frac{\partial^2 \ln p_s}{\partial s^2}\right|_{s=\mu} = -\frac{1}{Np} - \frac{1}{N(1-p)} = -\frac{p+q}{Npq} = -\frac{1}{Npq}$$

Distribuição Gaussiana como limite da Binomial

Se $\sigma^2 = Npq$

$$\ln p_s|_{s=\mu} = \frac{1}{2}\ln\left(\frac{1}{2\pi\sigma^2}\right)$$

$$\ln p_s = \frac{1}{2}\ln\left(\frac{1}{2\pi\sigma^2}\right) - \frac{1}{2}\frac{1}{\sigma^2}(s-\mu)^2 + \ldots$$

$$p_s \approx \exp\left(\ln\left(\frac{1}{\sqrt{\frac{1}{2\pi\sigma^2}}}\right) - \frac{(s-\mu)^2}{2\sigma^2}\right)$$

$p(s)ds = P($de p estar entre s e $s+ds) \Rightarrow p(s)[(s+1)-s] = p(s) = p_s$

$$p(s) = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{(s-\mu)^2}{2\sigma^2}}$$

Desigualdade de Chebyshev

Seja x uma v.a. com valor médio $\mu = E(X)$ e $\varepsilon > 0$ um número positivo qualquer. Então

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\text{Var (X)}}{\varepsilon^2}$$

Prova: $P(|X-\mu| \geq \varepsilon) = \sum_{|x_i-\mu|\geq\varepsilon} P(x_i)$ (A prova é semelhante no caso continuo)

$$\begin{aligned}
\sum_{x_i} (x_i-\mu)^2 P(x_i) &= \text{Var (X)} \\
\sum_{|x_i-\mu|\geq\varepsilon} (x_i-\mu)^2 P(x_i) &\leq \\
\sum_{|x_i-\mu|\geq\varepsilon} \varepsilon^2 P(x_i) &\leq \\
\varepsilon^2 \sum_{|x_i-\mu|\geq\varepsilon} P(x_i) &\leq \\
\varepsilon^2 P(|X-\mu| \geq \varepsilon) &\leq
\end{aligned}$$

Portanto $P(|X-\mu| \geq \varepsilon) \leq \dfrac{\text{Var (X)}}{\varepsilon^2}$

Lei dos grandes números ou das médias

Sejam $X_1, X_2, \ldots, X_n$ observações independentes com valor esperado $E(X_j) = \mu$ e variância $\mathrm{Var}\ (\mathrm{X}) = \sigma^2$. Seja $S_n = X_1 + X_2 + \cdots + X_n$. Então para qualquer $\varepsilon > 0$,

$$P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \to 0 \text{ com } n \to \infty.$$

Equivalentemente: $P\left(\left|\frac{S_n}{n} - \mu\right| < \varepsilon\right) \to 1 \text{ com } n \to \infty$

Prova: Uma vez que $X_1, X_2, \ldots, X_n$ são independentes e possuem a mesma distribuição, temos que:

$$\mathrm{Var}\ (\mathrm{S_n}) = \mathrm{n}\sigma^2 \qquad \text{e} \qquad \mathrm{Var}\left(\frac{\mathrm{S_n}}{\mathrm{n}}\right) = \frac{\sigma^2}{\mathrm{n}}.$$

Pela desigualdade de Chebyshev: $P(|X - \mu| \geq \varepsilon) \leq \dfrac{\sigma^2}{n\varepsilon}$

Portanto para um dado $\varepsilon > 0$: $P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \to 0 \text{ com } n \to \infty$

$\square$ QED (*quod erat demonstrandum*).

Observações dos lançamentos de uma moeda

Observações de Bernoulli

Seja p a probabilidade de obtermos cara.
Seja $X_j = 1$ no caso de sucesso na obtenção de cara e $X_j = 0$ no caso oposto.
Então $S_n = X_1 + X_2 + \cdots + X_n$ é o número de sucessos em n observações.
$\mu = E(X_j) = 1p + 0(1-p) = p$

De acordo com a lei dos grandes números: $P\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \to 0$ com $n \to \infty$

A firmação acima diz que, em um número n grande de experimentos podemos esperar que a proporção de vezes que o evento cara irá ocorrer é próximo de p.

Isto mostra que nosso modelo matemático de probabilidade concorda com nossa interpretação de frequência da probabilidade.

Teorema do limite central

Se S_N é a soma de N v.a. mutuamente independentes então a função de distribuição de S_N, com $N \to \infty$, é bem aproximada pela função de densidade Gaussiana (normal)

$$f_{\mu,\sigma}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} .$$

Vamos discutir este teorema apenas como aplicado a observações de Bernoulli.
Caso geral é assunto para um curso de teoria da probabilidade.
Considere observações de Bernoulli com probabilidade p de sucesso em cada experimento.
Seja $X_i = 1$ ou -1 respectivamente se o *i-ésimo* evento é um sucesso ou um fracasso.

$$S_N = X_1 + X_2 + \cdots + X_N,$$

é a diferença entre o número de sucessos e fracassos __ o excesso de spin ($2s$).
Sabemos que s ou $\frac{S_N}{2}$ possui como sua distribuição a distribuição binomial de probabilidades.

Portanto como já vimos se $N \to \infty$ esta distribuição é bem aproximada pela distribuição Gaussiana.