Um pouco mais sobre chi-quadrado

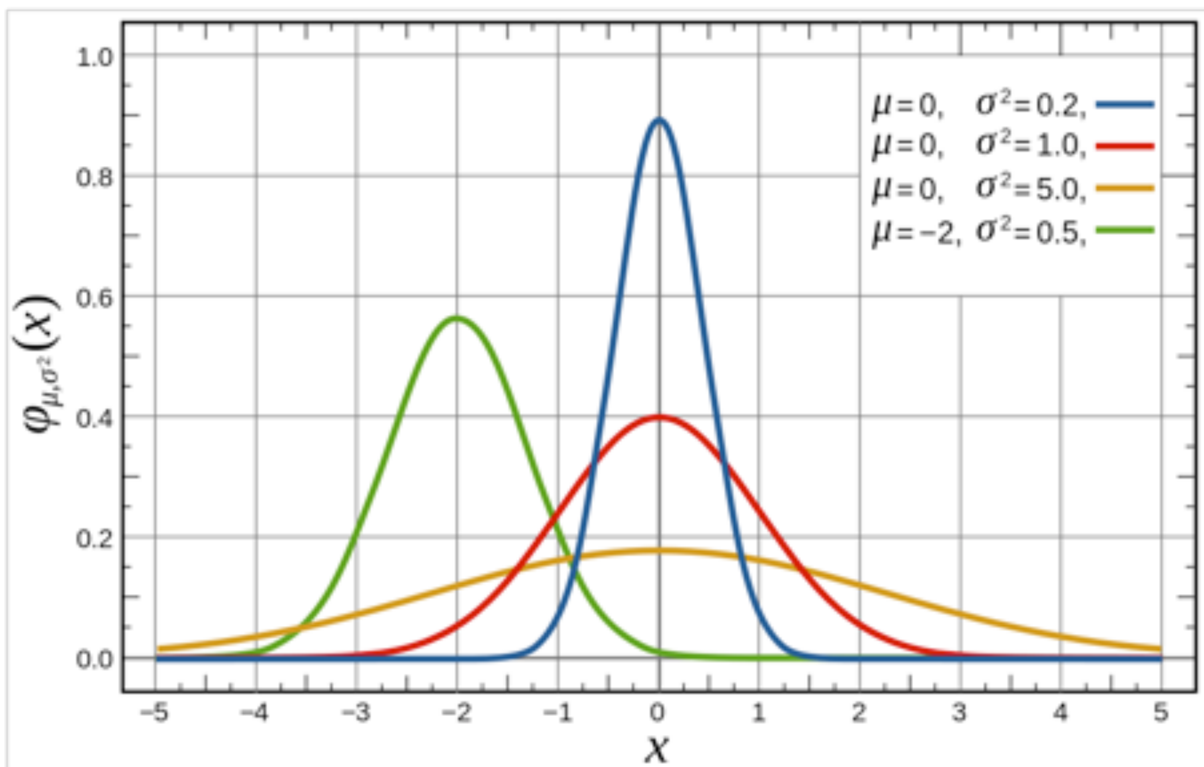
Alexandre Suaide

(texto sobre esse assunto será disponibilizado em breve)

Função densidade de probabilidade gaussiana

- Bastante comum no dia a dia da física experimental

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



- Será que tudo que é derivado de uma grandeza gaussiana também possui FDP gaussiana?

Experimento virtual

- Seja uma grandeza qualquer (x) que possua F.D.P. gaussiana de valor verdadeiro μ e variância σ conhecidos.

- Imagine um experimento no qual se realizam ν medidas independentes desta grandeza e o resultado do experimento é o valor médio dessas medidas, ou seja:

$$\bar{x} = \frac{1}{\nu}\sum_{i=1}^{\nu} x_i$$

- Calcula-se também o chi-quadrado destas medidas em relação ao valor verdadeiro, ou seja:

$$\chi^2 = \sum_{i=1}^{\nu}\left(\frac{x_i - \mu}{\sigma}\right)^2$$

- Quais as FDPs do valor médio e do valor de chi-quadrado? São gaussianas? qual a dependência dessas FDPs com o número de medidas realizadas, ν.

Experimento virtual

- Para isso, imagine que o experimentador repita, diariamente, o experimento descrito anteriormente e calcule, para cada dia, o valor da média e o valor de chi-quadrado e coloque os resultados em um histograma. No final da vida dele ele vai ter repetido isso tantas vezes que é possível conhecer as FDPs dessas duas grandezas.

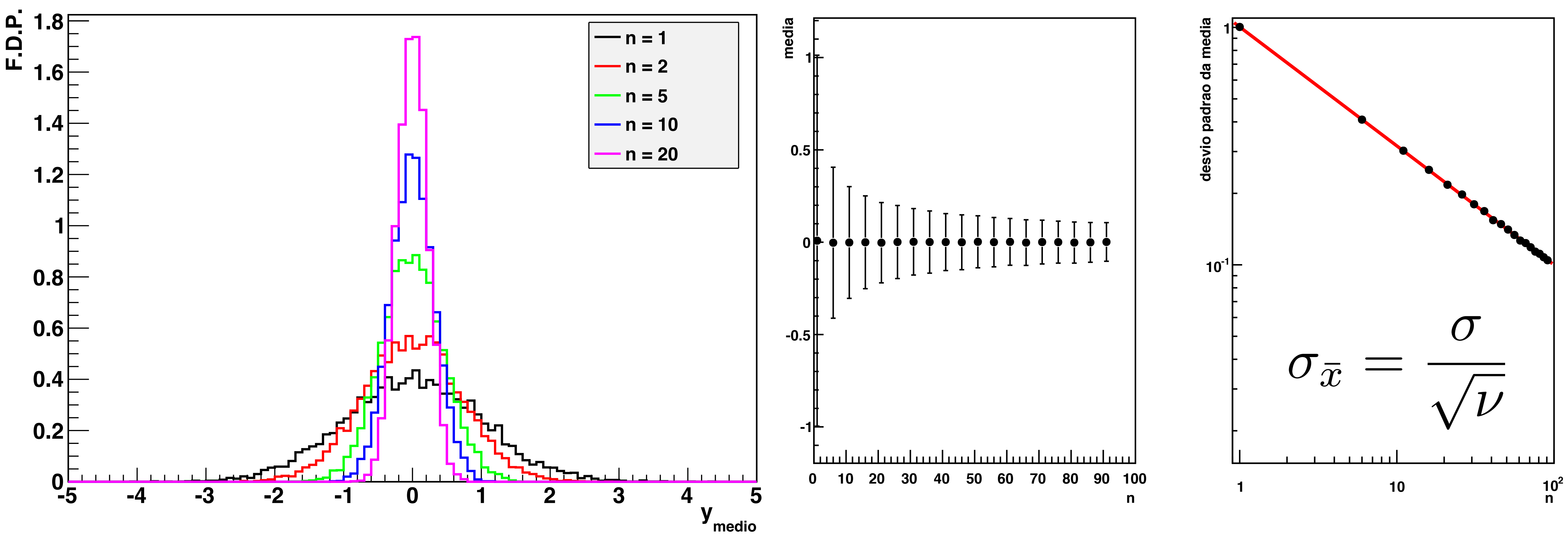
- Para facilitar, vamos considerar $\mu = 0$ e $\sigma = 1$. Não há perda de generalidade pois sempre podemos fazer uma mudança de variáveis do tipo:

$$\frac{x_i - \mu}{\sigma} \rightarrow y_i$$

- Por sorte existem computadores e podemos simular este experimento computacionalmente, bem como a repetição do mesmo indefinidamente.

Distribuição do valor médio

- Continua sendo uma gaussiana
  - A média não se altera com ν (n na figura)
  - Na medida em que ν (n na figura) aumenta, a gaussiana fica mais estreita



E para o chi-quadrado?

- Como os termos que compõe o chi-quadrado são sempre positivos não é esperado que ele tenha média zero.

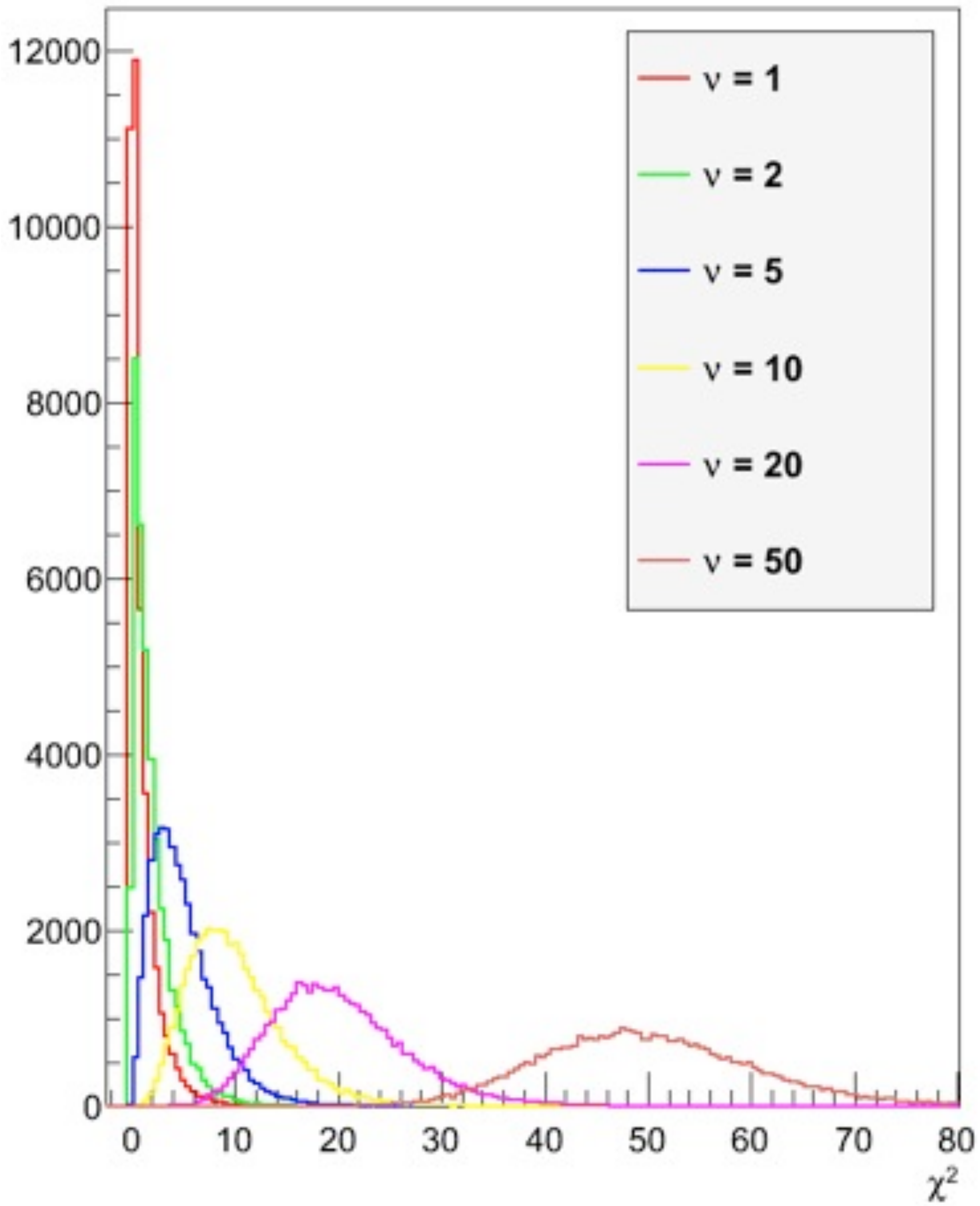
- Também não é esperado que a média seja independente de ν. Isso porque, quanto mais termos se somam, maior fica o chi-quadrado

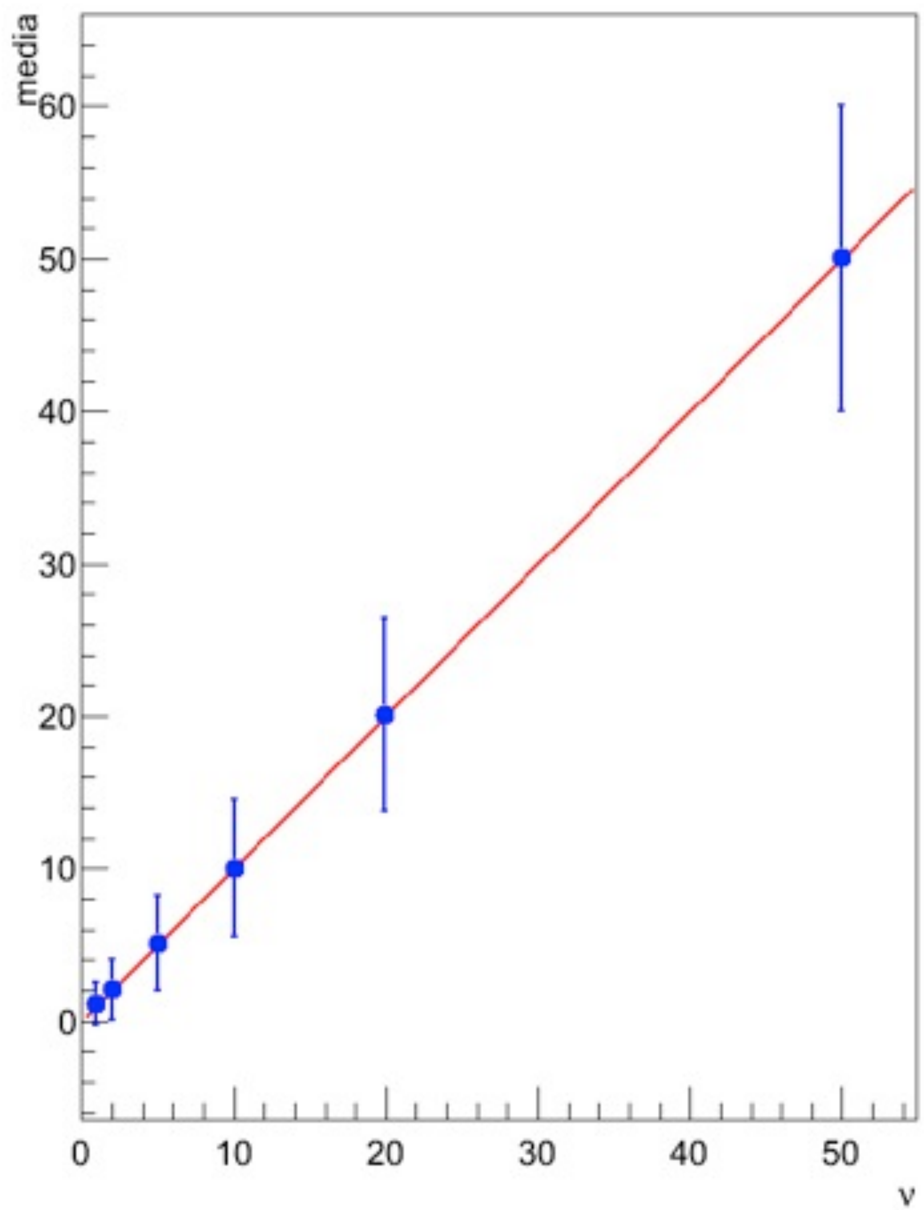
$$\chi^2 = \sum_{i=1}^{\nu} \left(\frac{x_i - \mu}{\sigma} \right)^2$$

E para o chi-quadrado?

- A FDP do chi-quadrado não é gaussiana
  - Se aproxima de uma gaussiana quando ν cresce
  - O valor médio da distribuição de chi-quadrado é ν.

$$\chi^2 = \sum_{i=1}^{\nu} \left(\frac{x_i - \mu}{\sigma} \right)^2$$

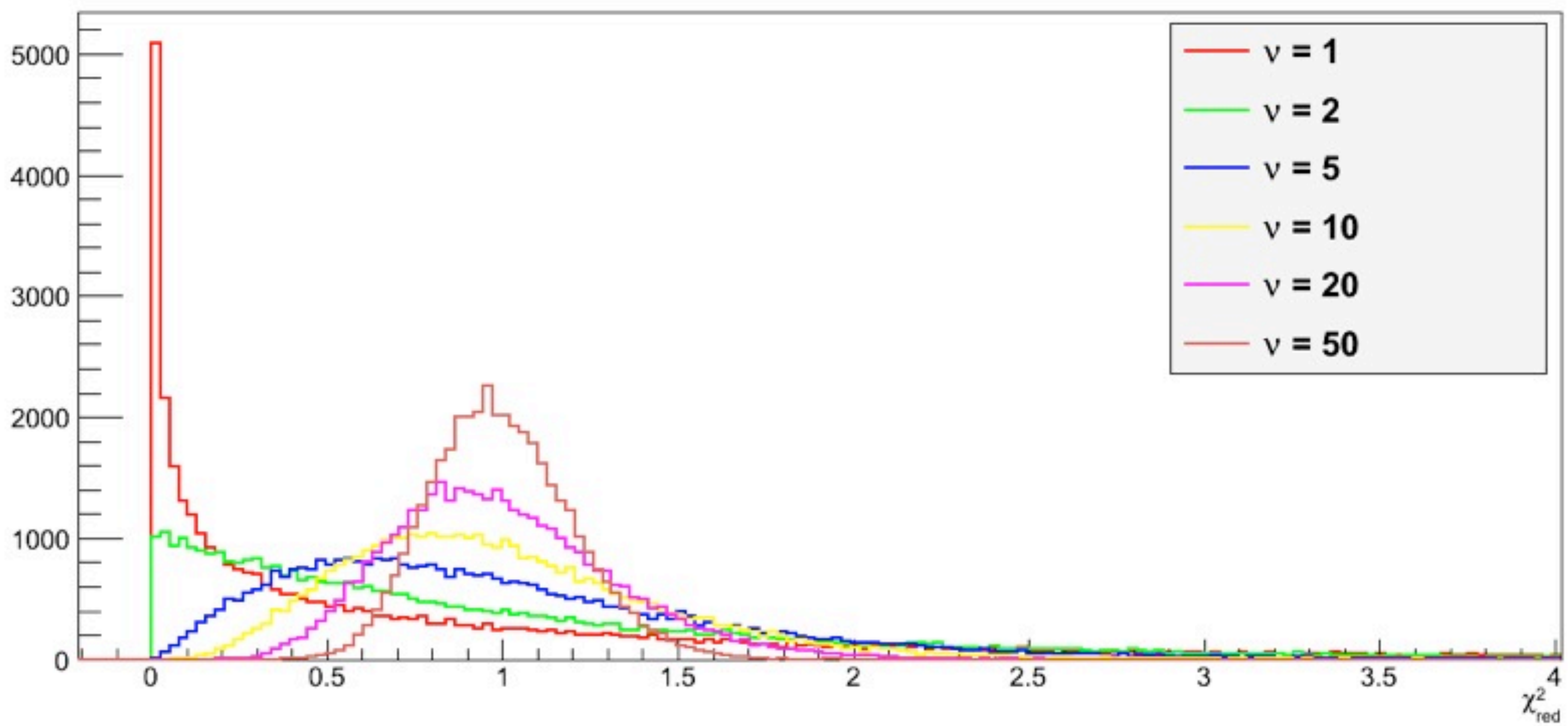




O chi-quadrado reduzido

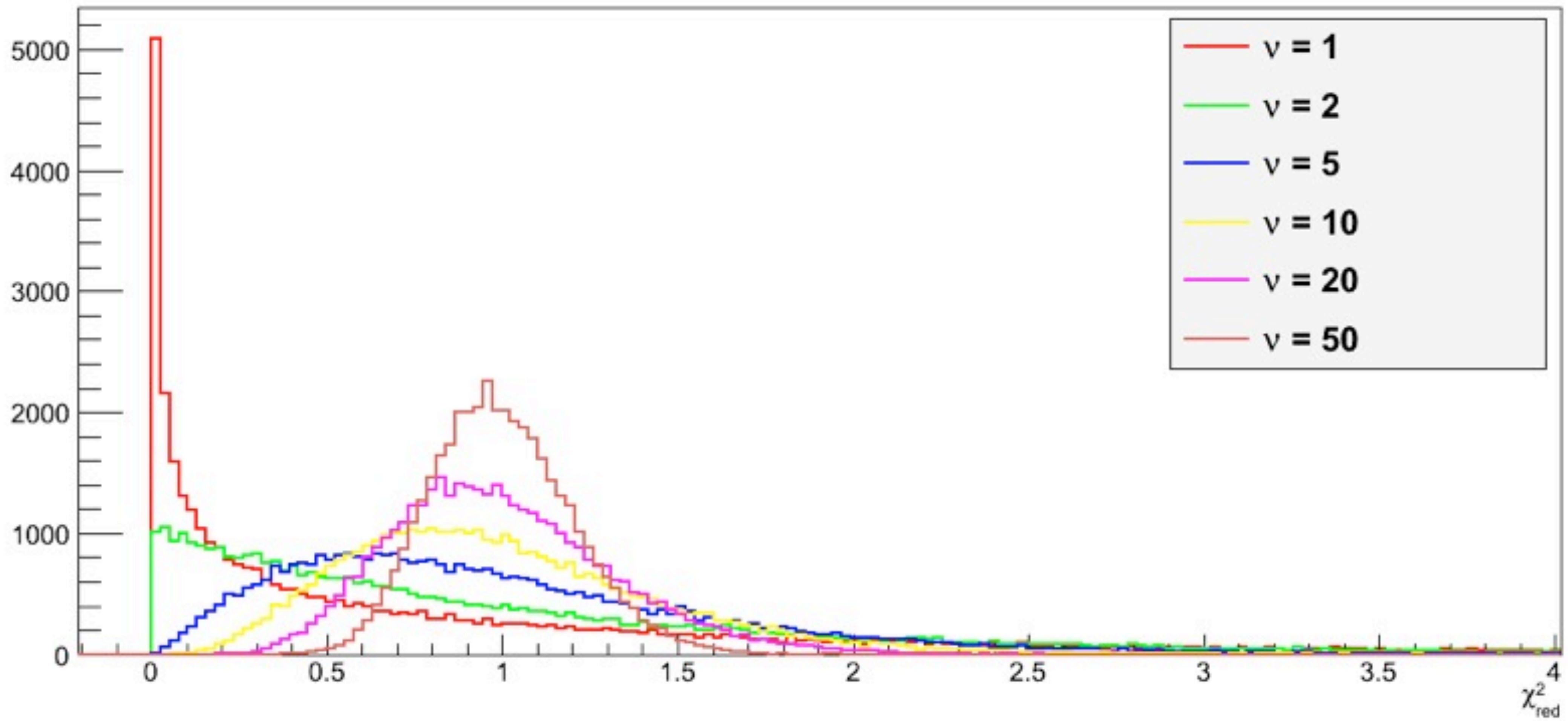
- Se o valor médio do chi-quadrado é ν, podemos definir uma grandeza com a mesma distribuição do chi-quadrado porém de média 1, chamada de chi-quadrado reduzido.

$$\chi^2_{red} = \frac{1}{\nu}\chi^2$$



Porque o chi-quadrado reduzido é importante?

- Como o **valor médio do chi-quadrado reduzido é 1**, pode-se criar um teste (**assunto da próxima aula**) para verificar a qualidade de ajustes de funções a dados experimentais
  - Note que, para valores grandes de ν, é pouco provável obter valores baixos ou altos para o chi-quadrado reduzido.
    - Ou seja, um ajuste no qual a função passa em cima de todos os pontos tem probabilidade estatística nula de descrever os dados experimentais.



Mas quem é ν?

- Olhando lá atrás, “... Imagine um experimento no qual se realizam ν medidas independentes desta grandeza e o resultado do experimento é o valor médio dessas medidas.. ”
  - ν é o número de medidas estatisticamente independentes no experimento. É chamado de **número de graus de liberdade** da amostra.
    - Número de graus de liberdade corresponde à quandtidade de valores independentes que podem variar livremente no cálculo de uma grandeza estatística.
    - O que é independente em uma amostra?

Número de graus de liberdade

- Imagine que tomamos um conjunto de dados e queremos calcular a variância deste conjunto de dados. Usamos a definição:

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

- Imagine a situação na qual não conhecemos o valor verdadeiro da amostra. Neste caso, a melhor estimativa do valor verdadeiro da amostra corresponde ao valor médio da mesma. Neste caso a variância seria calculada como:

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

- Note que *N* --> *N-1* na expressão acima. Porque disso?

Número de graus de liberdade

- No momento em que calculamos a média, somente N-1 pontos são totalmente livres para variar. O último ponto da amostra necessariamente deve ter o valor:

$$x_n = N\bar{x} - \sum_{i=1}^{N-1} x_i$$

- Ou seja, na verdade, um dos dados da amostra não é independente dos outros. Ele está vinculado aos demais por conta do valor médio calculado.

- Neste caso, o número de graus de liberdade disponíveis é ***N-1***. Esta é uma interpretação pela qual aparece o N-1 no cálculo da variância quando não se conhece o valor verdadeiro da amostra.

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

Número de graus de liberdade no ajuste de funções

- O chi-quadrado de um conjunto de dados em relação a uma função é dado por:

$$\chi^2 = \sum_{i=1}^{N} \left(\frac{y_i - f(x_i)}{\sigma_i} \right)^2$$

- Se conhecemos f(x), o número de graus de liberdade é N, pois os dados podem variar livremente um em relação ao outro.

- Porém, se f(x) for estimada dos dados, minimizando o chi-quadrado, então o número de graus de liberdade é dado pela diferença entre o número de dados (N) e o número de parâmetros ajustados em f(x), ou seja:

$$\nu = N - p$$

  - O motivo é o mesmo do N-1 no caso da variância. Para cada valor estimado de uma amostra, um valor dessa amostra deixa de ser independente dos demais

Para estudar

- Fundamentos da teoria de erros, J. H. Vuolo
  - Capítulo 12, em especial o tópico sobre interpretação de chi-quadrado
  - Capítulo 14, Qualidade de ajuste
- Vou disponibilizar um texto em breve sobre o assunto.