

Tratamento Estatístico de Dados em Física Experimental

Atividade 07 – Incertezas em contagens e histogramas

Envie um documento de texto ou arquivo pdf com as rotinas utilizadas e uma foto ou arquivo pdf da tabela com as respostas para o formulário em <https://forms.gle/VwhfBqeq8vynCKNy7>. Esta atividade deve ser entregue até às **23h59** do dia 06/10 (quarta-feira). Até o final do prazo é possível enviar novas versões dos arquivos (até 5 versões poderão ser enviadas em cada questão).

Iniciar o gerador de números aleatórios com semente igual ao seu número USP (no Python: `np.random.seed(nU)` com `nU` o seu número USP).

As rotinas utilizadas para fazer os itens numéricos da atividade (partes “a” e “b”) devem ser enviadas (na forma de arquivos texto ou arquivos pdf) como a Questão 0 desta atividade. O comando para iniciar o gerador de números de aleatórios deve fazer parte dessas rotinas.

Considere a função densidade de probabilidade $f(x) = \frac{3}{125}x^2$, para $x \in [0, 5]$. Escreva uma rotina para gerar dados que sigam essa função densidade de probabilidade (pode usar qualquer dos métodos ensinados na disciplina: método da exclusão ou método da inversão).

a) Parte numérica 1: uma primeira simulação

a.1) Gere um primeiro conjunto de $N = 450$ dados e preencha a coluna (1) do Quadro 1 com o número de ocorrências, n , obtido em cada intervalo. Lembre-se que o número de ocorrências é um número inteiro, então os valores de n devem ser escritos apenas até a casa da unidade, pois as casas decimais não são algarismos significativos.

a.2) Usando a aproximação $\tilde{p} = \frac{n}{N}$, estime a incerteza no número de ocorrências, n , em cada intervalo na simulação considerada no item anterior usando a aproximação $\tilde{\sigma}_n \cong \sqrt{n \left(1 - \frac{n}{N}\right)}$ e preencha a coluna (2) do Quadro 1. Nos casos em que $n = 0$ considere que a incerteza estimada seja $\tilde{\sigma}_n \cong 1$. Escreva essas incertezas estimadas arredondadas até a casa da unidade (para manter a coerência com a forma como os correspondentes valores de n foram escritos no item anterior).

b) Parte numérica 2: 10.000 conjuntos de simulações

b) Gere $n_{REP} = 10.000$ conjuntos de $N = 450$ dados cada e preencha as colunas (3) e (4) do Quadro 1 com o valor médio do número de ocorrências em cada intervalo, \bar{n} , acompanhados dos correspondentes desvios-padrões da média, $\sigma_{\bar{n}}$, e dos desvios-padrões amostrais, s_n , escritos de forma adequada, isto é, segundo as regras de que incertezas devem ser escritas com 1 ou 2 algarismos significativos e que os valores médios e os desvios-padrões amostrais devem ser escritos até a mesma casa decimal do desvio-padrão da média correspondente (regra baseada no fato que a incerteza do desvio-padrão amostral é da mesma ordem de grandeza do desvio-padrão da média).

c) Parte analítica:

c.1) Preencha a coluna (5) do Quadro 1 com as probabilidades de se obter, em cada medição, valores de x em cada um dos 5 intervalos indicados (essas probabilidades devem ser calculadas com base na integral de $f(x)$ nos intervalos correspondentes). Escreva essas probabilidades com 3 casas decimais.

c.2) Com base nas probabilidades calculadas no item anterior, e considerando o caso de experimentos com $N = 450$ dados, preencha as colunas (6) e (7) do Quadro 1 com os valores esperados (valores verdadeiros) do número de ocorrências, $n_0 = Np$, e correspondentes desvios-padrão (verdadeiros), $\sigma_{n_0} = \sqrt{Np(1-p)}$, de cada um dos 5 intervalos. Escreva esses resultados com 3 casas decimais.

Quadro 1 – Resultados numéricos e analíticos para o número de eventos nos 5 intervalos indicados em experimentos com $N = 450$ dados regidos pela função densidade de probabilidade $f(x) = \frac{3}{125}x^2$, no domínio $0 \leq x \leq 5$. Respeitar as restrições de algarismos significativos e/ou casas decimais indicadas nos enunciados.

	(a) – Numérico: primeiro conjunto de 450 dados gerados por Monte Carlo		(b) – Numérico: resultados de 10.000 conjuntos de 450 dados gerados por Monte Carlo		(c) – Analítico: usando $f(x)$ para calcular probabilidades e considerando binomiais com 450 tentativas		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$[x_a, x_b[$	n	$\tilde{\sigma}_n$	$\bar{n} \pm \sigma_{\bar{n}}$	s_n	$P(x_a \leq x \leq x_b)$	n_0	σ_{n_0}
$[0, 1[$							
$[1, 2[$							
$[2, 3[$							
$[3, 4[$							
$[4, 5]$							

Questões para reflexão (não fazem parte das respostas a serem entregues no formulário):

- Compare a estimativa da incerteza de n estimada a partir de um único conjunto [coluna (2)] usando a aproximação da probabilidade pela frequência relativa, com a incerteza real (o desvio-padrão verdadeiro) de cada n [coluna (7)], calculado a partir das probabilidades corretas [coluna (5)]. As incertezas estimadas a partir de um único conjunto parecem ser razoáveis?
- Compare os desvios-padrão amostrais dos 10.000 conjuntos [coluna (4)] com os valores verdadeiros dos desvios-padrão [coluna (7)]. Por que eles não são exatamente iguais? Parece haver tendência sistemática de um deles ser maior que o outro?
- O que acontece com a incerteza do número de contagens em cada canal do histograma, σ_{n_0} , quando aumenta o número de eventos esperados no canal? E com a incerteza relativa, σ_{n_0}/n_0 ?
- Compare as incertezas estimadas de n [coluna (2)] usando binomial com aproximação da probabilidade pela frequência relativa, com estimativas ainda mais simplificadas $\hat{\sigma}_n \cong \sqrt{n}$. São estimativas de incerteza muito diferentes? Para quais canais do histograma essas estimativas são equivalentes?