Laboratorio 4

Turno A, Primer cuatrimestre de 2024

Repaso de nociones de estadística y errores en mediciones

Pablo Cobelli

Departamento de Física, FCEN UBA

cobelli@df.uba.ar   materias.df.uba.ar/l4a2024c1   pablocobelli/laboratorio4

Breve repaso de nociones de estadística y errores en mediciones

- Estadística descriptiva
- Tipos de errores en mediciones
- Incerteza en lecturas únicas
- Propagación de errores
- Error estándar vs desviación estándar
- Propagación de errores usando Monte Carlo
- Principio de cuadrados mínimos
- Correlación entre variables
- Bondad del ajuste: χ^2 y p-valor

Propagación de errores en mediciones indirectas (parte 1)

- Medimos indirectamente volumen de cilindro: $V = \pi \frac{D^2}{4} h \equiv V(D, h)$.
- Tenemos N mediciones de D y de h: D_i y h_i; calculamos $\overline{h}$, σ_h y $\overline{D}$, σ_D.
- Nos preguntamos por $\overline{V}$ y por σ_V.

- En adelante pensamos en forma más general: $q = q(x, y)$, donde x e y son magnitudes medidas directamente N veces, y buscamos determinar $\overline{q}$ y σ_q indirectamente.

Propagación de errores en mediciones indirectas (parte 2)

- (x_i, y_i), con $i = 1, \ldots, N$, medidas directamente
- $q \equiv q(x, y)$, magnitud a determinar indirectamente

- Si los errores son *suficientemente* pequeños, entonces*:

$$
\begin{aligned}
q_i &= q(x_i, y_i), \\
q_i &\approx q(\bar{x}, \bar{y}) + \left.\frac{\partial q}{\partial x}\right|_{\bar{x},\bar{y}} (x_i - \bar{x}) + \left.\frac{\partial q}{\partial y}\right|_{\bar{x},\bar{y}} (y_i - \bar{y})
\end{aligned}
$$

Podemos calcular promedio y desviación estándar a partir de esta aproximación de Taylor de primer orden usando las definiciones usuales:

$$
\bar{q} = \frac{1}{N} \sum_{i=1}^{N} q_i, \qquad y \qquad \sigma_q = \frac{1}{N} \sum_{i=1}^{N} (q_i - \bar{q})^2
$$

Propagación de errores en mediciones indirectas (parte 3)

Calculemos la media de q:

$$
\begin{aligned}
\overline{q} &= \frac{1}{N}\sum_{i=1}^{N} q_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[q(\overline{x},\overline{y}) + \left.\frac{\partial q}{\partial x}\right|_{\overline{x},\overline{y}}(x_i-\overline{x}) + \left.\frac{\partial q}{\partial y}\right|_{\overline{x},\overline{y}}(y_i-\overline{y}) + O(\cdot)\right] \\
&= q(\overline{x},\overline{y}).
\end{aligned}
$$

Propagación de errores en mediciones indirectas (parte 4)

Calculemos la varianza de q:

$$\sigma_q^2 = \frac{1}{N}\sum_{i=1}^{N}(q_i - \overline{q})^2$$

Propagación de errores en mediciones indirectas (parte 4)

Calculemos la varianza de q:

$$
\begin{aligned}
\sigma_q^2 &= \frac{1}{N}\sum_{i=1}^{N}(q_i - \overline{q})^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\left.\frac{\partial q}{\partial x}\right|_{\overline{x},\overline{y}}(x_i - \overline{x}) + \left.\frac{\partial q}{\partial y}\right|_{\overline{x},\overline{y}}(y_i - \overline{y})\right]^2
\end{aligned}
$$

Propagación de errores en mediciones indirectas (parte 4)

Calculemos la varianza de q:

$$
\begin{aligned}
\sigma_q^2 &= \frac{1}{N}\sum_{i=1}^{N}(q_i-\overline{q})^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\left.\frac{\partial q}{\partial x}\right|_{\overline{x},\overline{y}}(x_i-\overline{x}) + \left.\frac{\partial q}{\partial y}\right|_{\overline{x},\overline{y}}(y_i-\overline{y})\right]^2 \\
&= \left(\frac{\partial q}{\partial x}\right)^2 \frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2 + \left(\frac{\partial q}{\partial y}\right)^2 \frac{1}{N}\sum_{i=1}^{N}(y_i-\overline{y})^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y})
\end{aligned}
$$

Propagación de errores en mediciones indirectas (parte 4)

Calculemos la varianza de q:

$$
\begin{aligned}
\sigma_q^2 &= \frac{1}{N}\sum_{i=1}^{N}(q_i-\overline{q})^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\left.\frac{\partial q}{\partial x}\right|_{\overline{x},\overline{y}}(x_i-\overline{x})+\left.\frac{\partial q}{\partial y}\right|_{\overline{x},\overline{y}}(y_i-\overline{y})\right]^2 \\
&= \left(\frac{\partial q}{\partial x}\right)^2\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2+\left(\frac{\partial q}{\partial y}\right)^2\frac{1}{N}\sum_{i=1}^{N}(y_i-\overline{y})^2+2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y}) \\
&= \left(\frac{\partial q}{\partial x}\right)^2\sigma_x^2+\left(\frac{\partial q}{\partial y}\right)^2\sigma_y^2+2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\sigma_{xy}
\end{aligned}
$$

Propagación de errores en mediciones indirectas (parte 4)

Calculemos la varianza de q:

$$
\begin{aligned}
\sigma_q^2 &= \frac{1}{N}\sum_{i=1}^{N}(q_i-\overline{q})^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\left.\frac{\partial q}{\partial x}\right|_{\overline{x},\overline{y}}(x_i-\overline{x})+\left.\frac{\partial q}{\partial y}\right|_{\overline{x},\overline{y}}(y_i-\overline{y})\right]^2 \\
&= \left(\frac{\partial q}{\partial x}\right)^2\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2+\left(\frac{\partial q}{\partial y}\right)^2\frac{1}{N}\sum_{i=1}^{N}(y_i-\overline{y})^2+2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})\,(y_i-\overline{y}) \\
&= \left(\frac{\partial q}{\partial x}\right)^2\sigma_x^2+\left(\frac{\partial q}{\partial y}\right)^2\sigma_y^2+2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\sigma_{xy}
\end{aligned}
$$

Propagación de errores en mediciones indirectas (parte 4)

Calculemos la varianza de q:

$$
\begin{aligned}
\sigma_q^2 &= \frac{1}{N}\sum_{i=1}^{N}(q_i-\overline{q})^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}\left[\left.\frac{\partial q}{\partial x}\right|_{\overline{x},\overline{y}}(x_i-\overline{x}) + \left.\frac{\partial q}{\partial y}\right|_{\overline{x},\overline{y}}(y_i-\overline{y})\right]^2 \\
&= \left(\frac{\partial q}{\partial x}\right)^2\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})^2 + \left(\frac{\partial q}{\partial y}\right)^2\frac{1}{N}\sum_{i=1}^{N}(y_i-\overline{y})^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\frac{1}{N}\sum_{i=1}^{N}(x_i-\overline{x})(y_i-\overline{y}) \\
&= \left(\frac{\partial q}{\partial x}\right)^2\sigma_x^2 + \left(\frac{\partial q}{\partial y}\right)^2\sigma_y^2 + 2\frac{\partial q}{\partial x}\frac{\partial q}{\partial y}\sigma_{xy}
\end{aligned}
$$

Notar que este resultado es general*, aunque aproximado**

Método de Monte Carlo para propagación de errores

Propagación de errores “clásica” vs el método de Monte Carlo

- Propagación “clásica”
  - informa al diseño de experiencias
  - informa sobre las precisiones relativas a utilizar
  - requiere de un cálculo complejo y/o no siempre es calculable*

- Propagación vía Monte Carlo
  - de acuerdo al caso, puede requerir muchas realizaciones
  - su programación es relativamente fácil
  - puede extenderse con algoritmos avanzados*

Error estándar vs desviación estándar

- El error estándar de una serie de mediciones x_i viene dado por:

$$\epsilon_x = \sigma_{\overline{x}} = \frac{\sigma_x}{\sqrt{N}}$$

- De dónde se obtiene este resultado?
- El error de medición... es el error estándar o la desviación estándar?

Coeficiente de correlación entre variables

Principio de cuadrados mínimos

Residuos y grados de libertad

Mejor estimado, en el sentido de cuadrados mínimos, de un set de datos de una única variable y, que medimos N veces. Nuestro "modelo" para el comportamiento es el más simple:

$$y(x) = c.$$

Calculemos la suma cuadrada de los residuos (entre modelo y experiencia) y minimicémosla respecto del parámetro c:

$$\chi^2 = \sum_{i=1}^{N} r_i^2 = \sum_{i=1}^{N} \left[y(x) - y_i \right]^2 = \sum_{i=1}^{N} (c - y_i)^2 ,$$

$$\Rightarrow \partial_c \chi^2 = 2 \sum_{i=1}^{N} (c - y_i) = 2 \left(Nc - \sum_{i=1}^{N} y_i \right) = 0 \quad \Rightarrow \quad c = \frac{1}{N} \sum_{i=1}^{N} y_i$$

Residuos y grados de libertad

Veamos que este extremo es efectivamente un mínimo:

$$\partial_c^2 \chi^2 = \partial_c^2 \left[\sum_{i=1}^{N} (c - y_i)^2 \right] = \partial_c \left[2 \sum_{i=1}^{N} (c - y_i) \right] = 2 \sum_{i=1}^{N} 1 = 2\,N > 0.$$

Residuos y grados de libertad

Pensemos en ajuste de cuadrados mínimos lineales con 2 parámetros con un modelo lineal en las variables

$$y(x) = ax + b$$

Los residuos vienen dados por:

$$r_i = y(x_i) - y_i = (ax_i + b) - y_i$$

Usamos 2 condiciones (una para cada parámetro) para minimizar los residuos:

$$\partial_a \chi^2 = 0$$
$$\partial_b \chi^2 = 0$$

Bondad del ajuste: χ^2 y p-valor