Laboratorio 4

Turno A. Primer Cuatrimestre 2024

Promedios y Barras de Error

Nicolás Piovanelli
✉ piovanellinicolas@hotmail.com

Mariana Antillano
✉ Mariana.antillano@gmail.com

Análisis Básico

Algunas definiciones de estadística

- Densidad de probabilidades normalizada $\longrightarrow$ $\int_{-\infty}^{\infty} P_{(x)}\, dx = 1$

- Promedio exacto $\longrightarrow$ $\int_{-\infty}^{\infty} x P_{(x)}\, dx$

- Varianza $\longrightarrow$ $\int_{-\infty}^{\infty} (x-\mu)^2 P_{(x)}\, dx$

- Desviación estándar $\longrightarrow$ $\sqrt{\int_{-\infty}^{\infty} (x-\mu)^2 P_{(x)}\, dx}$

Algunas definiciones de estadística

		Notación
Densidad de probabilidades normalizada	$\int_{-\infty}^{\infty} P_{(x)}\, dx = 1$	$<>$
Promedio exacto	$\int_{-\infty}^{\infty} x P_{(x)}\, dx$	$< x > \equiv \mu$
Varianza	$\int_{-\infty}^{\infty} (x-\mu)^2 P_{(x)}\, dx$	$< x^2 > - \mu^2 \equiv \sigma^2$
Desviación estándar	$\sqrt{\int_{-\infty}^{\infty} (x-\mu)^2 P_{(x)}\, dx}$	$\sqrt{< x^2 > - \mu^2} \equiv \sigma$

¿Cómo obtengo el valor más esperado de mis mediciones?

$$\mu = \int_{-\infty}^{\infty} x P_{(x)}\, dx$$

¿Cómo obtengo el valor más esperado de mis mediciones?

Lo que quiero

$$\mu = \int_{-\infty}^{\infty} x P_{(x)}\, dx$$

Lo que no conozco

¿Cómo obtengo el valor más esperado de mis mediciones?

Lo que quiero $\mu = \int_{-\infty}^{\infty} x P_{(x)}\, dx$ Lo que no conozco

¿Qué es lo que sí conozco?

- Promedio $\equiv \bar{x} = \dfrac{\sum_{n=1}^{N} x_i}{N}$
- Varianza $\equiv s^2 = \dfrac{\sum_{n=1}^{N} (x_i - \bar{x})^2}{N}$

¿Cómo obtengo el valor más esperado de mis mediciones?

Lo que quiero

$$\mu = \int_{-\infty}^{\infty} x P_{(x)}\, dx$$

Lo que no conozco

¿Qué es lo que sí conozco?

- Promedio $\equiv \bar{x} = \frac{\sum_{n=1}^{N} x_i}{N}$
- Varianza $\equiv s^2 = \frac{\sum_{n=1}^{N} (x_i - \bar{x})^2}{N}$

Lo que medí

Dos postulados útiles

Dos postulados útiles

- El promedio de la suma de N variables *independientes* con *la misma distribución* es igual al producto de N con el promedio de una sola de las variables.

$$\mu = < \sum_{n=1}^{N} x_i > = \sum_{n=1}^{N} < x_i > = \mathrm{N} < x_i >$$

Dos postulados útiles

- El promedio de la suma de N variables *independientes* con *la misma distribución* es igual al producto de N con el promedio de una sola de las variables.

$$\mu = < \sum_{n=1}^{N} x_i > = \sum_{n=1}^{N} < x_i > = \mathrm{N} < x_i >$$

- La varianza de la suma de N variables *independientes* con *la misma distribución* es igual al producto de N con la varianza de una sola de las variables.

$$\sigma^2 = < \left(\sum_{n=1}^{N} x_i \right)^2 > - < \left(\sum_{n=1}^{N} x_i \right) >^2 = \sum_{n=1}^{N} (< x_i^2 > - < x_i >^2) = \mathrm{N}\, \sigma^2$$

Una suposición

- Supongamos ahora que sobre la magnitud de interés, no solo realizamos un set de mediciones, si no N sets de mediciones. Calculando el promedio de cada conjunto de datos y realizando este proceso suficientes veces, podemos construir la distribución y calcular el promedio exacto.

$$< \bar{x} > = < \frac{1}{N}\sum_{n=1}^{N} x_i > = \frac{1}{N}\sum_{n=1}^{N} < x_i > = \frac{\cancel{N}}{\cancel{N}} < x > = \mu$$

Usando los postulados enunciados

Es decir que el mejor estimador de μ es $\bar{x}$.

- Cuando un estimador es promediado por la repetición del experimento N veces y este promedio coincide con el promedio exacto, se dice que el estimador es imparcial.

Incerteza del promedio exacto

- Ahora que podemos calcular el promedio exacto con datos medidos, cabe preguntarse la incerteza de esta estimación. Tal como con el promedio, si tomamos N sets de datos y calculamos la varianza de cada uno de los promedios
podemos construir la distribución y así calcular

$$\sigma_{\bar{x}}^2 \equiv < \bar{x}^2 > - < \bar{x} >^2 = \frac{\sigma^2}{N}$$

Usando los postulados enunciados

Incerteza del promedio exacto

- Ahora que podemos calcular el promedio exacto con datos medidos, cabe preguntarse la incerteza de esta estimación. Tal como con el promedio, si tomamos N sets de datos y calculamos la varianza de cada uno de los promedios
podemos construir la distribución y así calcular

$$\sigma_{\bar{x}}^2 \equiv < \bar{x}^2 > - < \bar{x} >^2 = \frac{\sigma^2}{N}$$

No lo conozco, ya que depende de $P_{(x)}$

Usando los postulados enunciados

Incerteza del promedio exacto

- Ahora que podemos calcular el promedio exacto con datos medidos, cabe preguntarse la incerteza de esta estimación. Tal como con el promedio, si tomamos N sets de datos y calculamos la varianza de cada uno de los promedios
podemos construir la distribución y así calcular

$$\sigma_{\bar{x}}^2 \equiv <\bar{x}^2> - <\bar{x}>^2 = \frac{\sigma^2}{N}$$

No lo conozco ya que depende de $P_{(x)}$

Usando los postulados enunciados

- Aún puedo extraer información de los datos que medí, usando la varianza de un set de datos.

$$<s^2> = \frac{1}{N}\sum_{n=1}^{N} <x_i^2> - \frac{1}{N^2}\sum_{n=1}^{N}\sum_{j=1}^{N} <x_i x_j> = \frac{N-1}{N}\sigma^2$$

- Usando el resultado anterior y despejando obtenemos que el mejor estimador para $\sigma_{\bar{x}}^2$ es

$$\sigma_{\bar{x}}^2 = \frac{s^2}{N-1}$$

Nuevamente obtuvimos un estimador imparcial.

Conclusiones

- Dado un experimento en el que se realizan N mediciones, pese a no poder conocer la función de distribución de probabilidades puedo calcularlo, pues el mejor estimador de este está dado por

$$< \bar{x} > = = \mu$$

- Además, considerando la desviación estándar como la medida de incerteza tengo que el mejor estimador de esta será

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{N-1}}$$

- Finalmente tengo que qué el promedio exacto estará dado por

$$\mu = \bar{x} \pm \sigma_{\bar{x}} = \frac{\sum_{n=1}^{N} x_i}{N} \pm \frac{s}{\sqrt{N-1}}$$

Es importante destacar que esto es válido bajo las hipótesis de

Igual función de distribución para todas las variables

Variables ***independientes***

Más de una variable

En caso de tener más de una variable aleatoria, por ejemplo x e y la varianza va a tener además de lo ya mostrado, un término de correlación cruzada. A este término se lo denomina covarianza

$$Cov_{(x,y)} = \sigma_{xy}^2 = < (x - < x >)(y - < y >) >$$

Que para un set de N datos será

$$s_{xy}^2 = \frac{1}{N} \sum_{n=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$$

Análisis Avanzado

- **Funciones lineales:**

Promedio aritmético $\equiv \overline{x} = \frac{\sum_{n=1}^{N} x_i}{N}$

- **Funciones no lineales:**

Varianza $\equiv s^2 = \frac{\sum_{n=1}^{N}(x_i - \bar{x})^2}{N}$

Curtosis $\equiv \langle x^4 \rangle / \langle x^2 \rangle^2$.

- Propagación de errores tradicional
- Método "Jackknife"
- Método "Bootstrap"

Propagación de errores tradicional

Expandimos nuestra función alrededor del promedio exacto hasta orden 2

$$f(\overline{y},\overline{z}) = f(\mu_y,\mu_z)+(\partial_{\mu_y} f)\,\delta_{\overline{y}}+(\partial_{\mu_z} f)\,\delta_{\overline{z}}+\frac{1}{2}\,(\partial^2_{\mu_y\mu_y} f)\,\delta^2_{\overline{y}}+(\partial^2_{\mu_y\mu_z} f)\,\delta_{\overline{y}}\delta_{\overline{z}}+\frac{1}{2}\,(\partial^2_{\mu_z\mu_z} f)\,\delta^2_{\overline{z}}+\cdots$$

Haciendo muchas repeticiones de la toma de datos y promediando esto obtenemos:

$$\langle f(\overline{y},\overline{z})\rangle - f(\mu_y,\mu_z) = \frac{1}{2}\,(\partial^2_{\mu_y\mu_y} f)\,\sigma^2_{\overline{y}} + (\partial^2_{\mu_y\mu_z} f)\,\sigma^2_{\overline{y}\,\overline{z}} + \frac{1}{2}\,(\partial^2_{\mu_z\mu_z} f)\,\sigma^2_{\overline{z}} + \cdots .$$

Notemos que: $\langle \delta^2_{\overline{y}}\rangle = \langle \overline{y}^2\rangle - \langle \overline{y}\rangle^2 \equiv \sigma^2_{\overline{y}}, \quad \langle \delta^2_{\overline{z}}\rangle = \langle \overline{z}^2\rangle - \langle \overline{z}\rangle^2 \equiv \sigma^2_{\overline{z}}, \quad \langle \delta_{\overline{y}}\delta_{\overline{z}}\rangle = \langle \overline{y}\,\overline{z}\rangle - \langle \overline{y}\rangle\langle \overline{z}\rangle \equiv \sigma^2_{\overline{y}\,\overline{z}},$

Conociendo la relación anteriormente calculada entre la varianza y la varianza de la muestra, llegamos a:

$$f(\mu_y,\mu_z) = \langle f(\overline{y},\overline{z})\rangle - \frac{1}{(N-1)}\left[\frac{1}{2}\,(\partial^2_{\mu_y\mu_y} f)\,s^2_y + (\partial^2_{\mu_y\mu_z} f)\,s^2_{yz} + \frac{1}{2}\,(\partial^2_{\mu_z\mu_z} f)\,s^2_z\right] + \cdots .$$

Propagación de errores tradicional

El error al considerar al mejor estimado como la función evaluado en los promedios de nuestra muestra viene dado por:

$$\sigma_f^2 \equiv \langle\, f^2(\overline{y},\overline{z})\,\rangle - \langle f(\overline{y},\overline{z})\rangle^2$$
$$= (\partial_{\mu_y} f)^2\, \langle \delta_{\overline{y}}^2 \rangle + 2(\partial_{\mu_y} f)\,(\partial_{\mu_z} f)\,\langle \delta_{\overline{y}}\delta_{\overline{z}}\rangle + (\partial_{\mu_z} f)^2\,\langle \delta_{\overline{z}}^2\rangle\,.$$

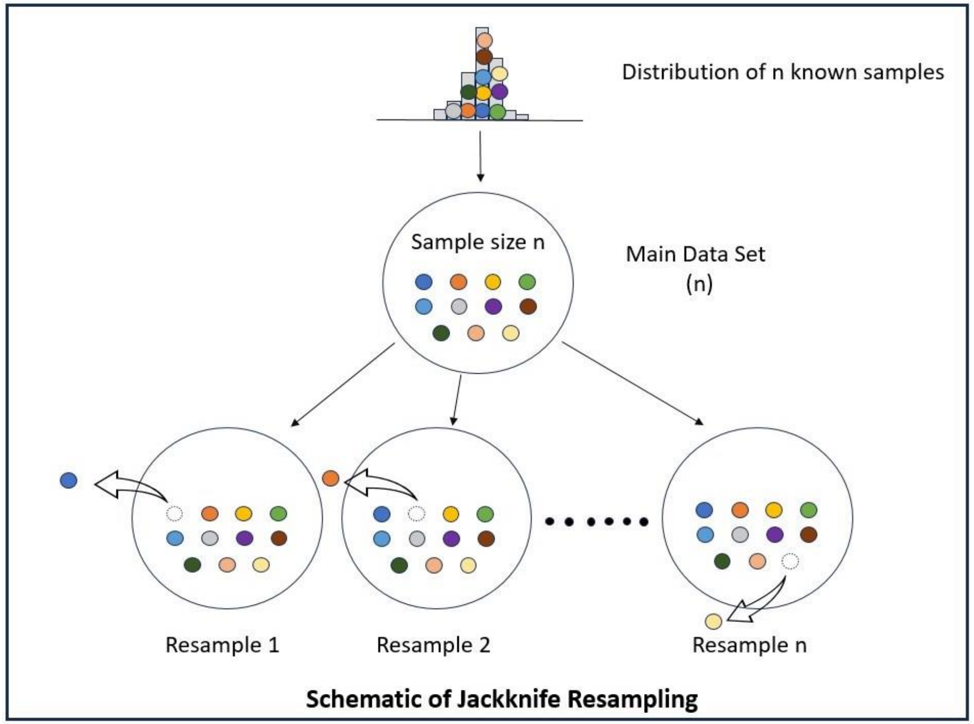
De nuevo tomando en cuenta que la varianza es (N-1)^-1 veces la varianza de la muestra, tenemos que

$$\sigma_f^2 = \frac{1}{(N-1)}\,(\partial_{\mu_y} f)^2\, s_y^2 + 2(\partial_{\mu_y} f)\,(\partial_{\mu_z} f)\, s_{yz}^2 + (\partial_{\mu_z} f)^2\, s_z^2\,.$$

Así,

$$\boxed{f(\mu_y,\mu_z) = f(\overline{y},\overline{z}) \pm \sigma_f\,,}$$

Jackknife



Tomado de https://en.wikipedia.org/wiki/Jackknife_resampling

Implica un remuestreo que sale de la muestra original

Estimado i-ésimo:

$$y_i^J \equiv \frac{1}{N-1}\sum_{j\neq i} y_j \quad \left(= \overline{y} + \frac{1}{N-1}(\overline{y} - y_i)\right).$$

Estimado i-ésimo de la función f:

$$f_i^J \equiv f(y_i^J, z_i^J).$$

Estimado general:

$$\boxed{\overline{f^J} \equiv \frac{1}{N}\sum_{i=1}^{N} f_i^J .}$$

Volviendo a la expresión anterior de propagación de errores,

$$f(\mu_y, \mu_z) = \langle f(\overline{y}, \overline{z})\rangle - \frac{A}{N} - \frac{B}{N^2} + \cdots ,$$

Considerando que los conjuntos de datos jackknife tienen N-1 datos con la misma distribución que la muestra original de N datos,

$$f(\mu_y, \mu_z) = \langle \overline{f^J}\rangle - \frac{A}{N-1} - \frac{B}{(N-1)^2} \cdots .$$

Tomando una combinación lineal apropiada de estas ecuaciones,

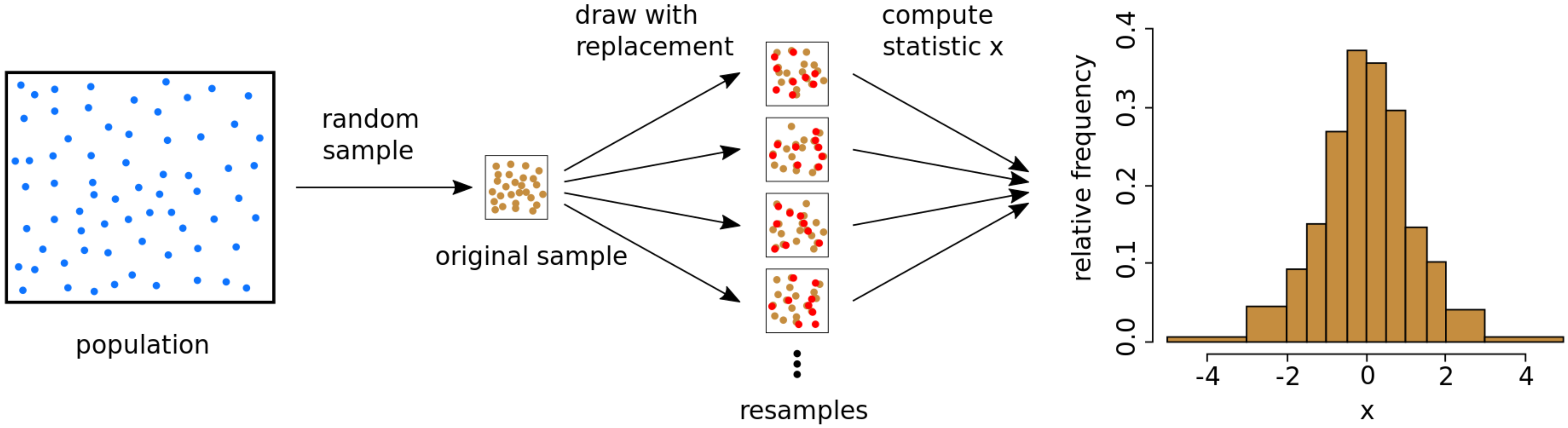
$$f(\mu_y, \mu_z) = N\langle f(\overline{y}, \overline{z})\rangle - (N-1)\langle \overline{f^J}\rangle + O\left(\frac{1}{N^2}\right) .$$

Por otro lado, la varianza en los estimados jacknnifes viene dada por

$$s_{f^J}^2 = \frac{1}{(N-1)^2} \left[(\partial_{\mu_y} f)^2 s_y^2 + (\partial_{\mu_z} f)^2 s_z^2 + 2(\partial_{\mu_y} f)(\partial_{\mu_z} f) s_{yz}^2 \right] ,$$

$$\boxed{\text{the jackknife estimate for } \sigma_f \text{ is } \sqrt{N-1}\, s_{f^J} .}$$

Bootstrap



Tomado de https://en.wikipedia.org/wiki/Bootstrapping_(statistics))