Estadística 1 (Química)

Práctica 3 - Sumas de variables aleatorias

1. Se realizan mediciones independientes del volumen inicial (X) y final (Y) en una bureta. Supongamos que las mediciones inicial y final siguen el modelo de errores independientes, es decir,

$$X = \alpha + \varepsilon_X, \qquad Y = \beta + \varepsilon_Y$$

donde α y β son los volúmenes desconocidos, ε_X y ε_Y, los errores de medición, son variables aleatorias independientes con media 0 y varianza σ^2. Sea $Z = Y - X$.

   a) Hallar $\mathbb{E}(Z)$ y $\mathbb{V}(Z)$.

   b) En una titulación, la lectura inicial en una bureta es de 0.00ml y la lectura final es de 15.60ml. Para ambas mediciones se sabe que el error de medición tiene una desviación estándar de 0.05ml.

      1) Calcular el valor estimado del volumen utilizado.

      2) ¿Cuál es la desviación estándar de su error de medición?

2. Modelo para medición con error aditivo. Se desea determinar una magnitud μ. Para ello se realizarán n medidas repetidas, es decir, se realizarán n mediciones de la misma magnitud en idénticas condiciones, que denotaremos con $X_1, \ldots, X_n$. Asumimos el siguiente modelo para las variables aleatorias X_i

$$X_i = \mu + \varepsilon_i$$

donde μ es la verdadera magnitud desconocida, y ε_i es la variable aleatoria que denota el error de la iésima medición. Asumimos que $\varepsilon_1, \ldots, \varepsilon_n$ son variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) con esperanza cero y varianza $\sigma^2 = 0{,}25$. Notar que los errores son no observables. El supuesto de que $\mathbb{E}(\varepsilon_i) = 0$ refleja la creencia en que el método de medición empleado es exacto. Es decir que no produce errores sistemáticos. La varianza del error, $\sigma^2 = \mathbb{V}(\varepsilon_i)$ representa la precisión del método de medición empleado. Sea

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

el promedio (o media muestral) de las n observaciones.

Asumir ahora que el error de medición tiene una distribución normal con media cero. Este modelo probabilístico se conoce como el Modelo de Gauss sin sesgo. Asumir también que el error de medición tiene desvío estándar $\sigma = 0{,}5$, o sea que $\varepsilon_i \sim \mathcal{N}(0; 0{,}25)$.

   a) Obtener la distribución de $\overline{X}_n$, su esperanza y su varianza.

   b) Calcular la probabilidad de que el promedio de $n = 10$ mediciones y de $n = 100$ mediciones diste de la verdadera magnitud μ en menos de 0.1 unidades. Notar que no fue necesario conocer el valor de μ para realizar este cálculo.

   c) Obtener una expresión para la probabilidad de que el promedio de n mediciones diste de la verdadera magnitud μ en menos de 0,1 unidades en función de n. Estudiar la monotonía y el límite cuando n tiende a infinito de esta probabilidad.

   d) Determinar cuán grande debe ser n para que $P\left(|\overline{X}_n - \mu| < 0{,}1\right) \geq 0{,}99$.

e) Para incluir en el futuro, en alguna parte de esta practica. Calcular la siguiente probabilidad

$$\mathbb{P}\left(\overline{X}_n - \frac{1{,}96 * \sqrt{0{,}25}}{\sqrt{n}} < \mu < \overline{X}_n + \frac{1{,}96 * \sqrt{0{,}25}}{\sqrt{n}}\right)$$

3. Considerar nuevamente el modelo de mediciones propuesto en el ejercicio anterior, suponiendo ahora que ε_i tiene una distribución **desconocida**, pero se sabe que

$$E(\varepsilon_i) = 0 \ , \quad V(\varepsilon_i) = 0{,}25$$

   a) Hallar $\mathbb{E}(\overline{X}_n)$ y $\mathbb{V}(\overline{X}_n)$.

   b) Para $n = 10$ y $n = 100$ mediciones, usando la desigualdad de Chebyshev, encontrar una cota inferior para

$$P\left(|\overline{X}_n - \mu| < 0{,}1\right).$$

   Comparar el resultado obtenido con el hallado en el ítem (b) del ejercicio anterior.

   c) Determinar cuán grande debe ser n para que $P\left(|\overline{X}_n - \mu| < 0{,}1\right) \geq 0{,}99$, usando nuevamente la desigualdad de Chebyshev. Comparar el resultado obtenido con el valor hallado en el ítem (d) del ejercicio anterior.

4. Se desea conocer la proporción de personas que están a favor de la despenalización del aborto en una ciudad. Sea p la proporción poblacional que está a favor de la despenalización. Observar que p es un número fijo y desconocido. Para estimar a p se eligen n personas al azar y se les pregunta a cada una de ellas su opinión. Para i entre 1 y n, sean

$$X_i = \begin{cases} 1 & \text{si la iésima persona encuestada está a favor de la despenalización} \\ \\ 0 & \text{en otro caso.} \end{cases}$$

   Asumimos que las X_i son v.a.i.i.d.

   a) Expresar la proporción muestral de encuestados a favor de la despenalización en términos de las variables X_i.

   b) Proponga una expresión que sea una variable aletoria que permita aproximar p para n suficientemente grande. En adelante denominaremos a esta variable aleatoria **estimador**.

   c) Hallar una cota superior para la probabilidad de que el estimador y el verdadero parámetro p difieran en más de 0.1, que no dependa de valores desconocidos. Es decir, acotar superiormente la siguiente expresión,

$$P\left(\left|\overline{X}_n - p\right| > 0{,}1\right)$$

   de modo que la cota sólo dependa de n. ¿Qué pasa con esta probabilidad cuando n aumenta? ¿Cómo puede el encuestador mejorar su estimación de p?

5. Considerar nuevamente el modelo propuesto en el ejercicio 2, asumiendo ahora que los errores se distribuyen de manera uniforme: $\varepsilon_i \sim \mathcal{U}(-\sqrt{3}/2, \sqrt{3}/2)$. Calcular de forma aproximada la probabilidad $P\left(|\overline{X}_n - \mu| < 0{,}1\right)$, para $n = 100$ y determinar cuán grande debe ser n para que $P\left(|\overline{X}_n - \mu| < 0{,}1\right) \geq 0{,}99$. Comparar con los resultados obtenidos en el ejercicio 2 y en el ejercicio 3. ¿Qué observa?

6. Se tira 100 veces un dado de 6 caras. Usar la aproximación normal para hallar la probabilidad de que:

a) salga “6” entre 15 y 20 veces, inclusive.

b) la suma de los resultados obtenidos sea menor que 300.

c) el número de veces que el resultado sea par esté entre 40 y 60 veces, inclusive.

d) el número de veces que el resultado sea par sea mayor o igual que el número de veces que el resultado sea impar.

7. El cromato y dicromato de postasio forman soluciones intensamente coloreadas. Los iones de cromato y dicromato se encuentran en equilibrio químico dependiente del pH. Se quiere estudiar si el dióxido de carbono burbujeado puede alterar el equilibrio de una solución de cromato y desplazarla al dicromato. Un sensor mide la absorbancia característica del cromato, cuando la reacción se desplaza al dicromato, se activa un mecanismo que automáticamente agrega hidróxido de potasio hasta revertir el efecto del dióxido de carbono. El tiempo (en minutos) que la solución demora en desplazarse al dicromato es una variable aleatoria Exponencial con parámetro $\lambda = 2$. Se puede suponer que los tiempos que tarda la solución en desplzarse completamente son independientes entre sí.

   a) Un investigador se va 30 minutos de su puesto y deja programadas 62 alícuotas de hidróxido de potasio ¿Cuál es, aproximadamente, la probabilidad de que el investigador encuentre la solución con la coloración del cromato?

   b) ¿Cuántas alícuotas de hidróxido de potasio tiene que dejar si quiere encontrar la solución sin desplazarse a dicromato con una probabilidad aproximada mayor o igual que 0,99?

8. Para rellenar una zona del río se utilizan 2 camiones (A y B). La distribución de la carga diaria (en toneladas métricas) transportada por el camión A tiene función de densidad

$$f(x) = \begin{cases} \frac{x}{52} & \text{si } 11 < x < 15, \\ 0 & \text{en otro caso} \end{cases}$$

   El camión B lleva una carga diaria en toneladas con esperanza 18 Tm y desvío estándar 1,3 Tm.

   a) Calcular esperanza y varianza de la carga diaria transportada por A.

   b) Calcular esperanza y desvío estándar de la carga total llevada por los dos camiones en un día, asumiendo que las cargas transportadas por ambos camiones son independientes.

   c) Calcular **aproximadamente** la probabilidad de que la carga total transportada en 256 días esté entre 7950 y 8000 Tm. Asuma que las cargas transportadas en días distintos son independientes.

   d) ¿Puede calcular la probabilidad pedida en c) **exactamente**?

9. Supongamos que el peso (en gramos) de un frasco de mermelada sigue una distribución normal con un peso medio de μ_A para la marca A y μ_B para la marca B, y con desvíos $\sigma_A = 8g$ y $\sigma_B = 6g$, respectivamente. Se eligen al azar $n_A = 40$ frascos de la marca A y $n_B = 35$ frascos de la marca B, y se pesan sus contenidos.

   a) Defina las variables aleatorias involucradas en este problema. Indique la distribución exacta o aproximada de las medias muestrales asociadas a ambas marcas. Indique la distribución exacta o aproximada de la diferencia entre la media muestral para la marca A y la media muestral para la marca B. Enuncie los resultados teóricos en los que se basa. Observe que no es necesario conocer los valores de μ_A y μ_B sino sólo de la diferencia para poder determinar la distribución de $\overline{X}_{n_A} - \overline{Y}_{n_B}$.

b) Asumamos que los pesos medios o esperados de ambas marcas son iguales. Calcular la probabilidad de que la distancia entre las medias muestrales sea a lo sumo 3.

c) Calcular la probabilidad de que la distancia anterior sea por lo menos 5, asumiendo que los pesos medios de ambas marcas son iguales.

d) No se conocen los valores de μ_A y μ_B, sólo se sabe que el valor observado de la diferencia entre la **media muestral** para la marca A y la **media muestral** para la marca B es 5, ¿sería razonable inferir que $\mu_A - \mu_B = 0$? Para contestar calcule, asumiendo que $\mu_A - \mu_B = 0$, la probabilidad de que la distancia entre las medias muestrales sea 5 ó más. Si esta probabilidad fuera "muy pequeña" uno tendería a dudar de que la diferencia entre las medias poblacionales es 0.

e) ¿Se modifican las distribuciones de 9*a*) si $n_A = 400$ y $n_B = 350$? Justificar.