Estadística (Química) - 2014

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) Halle $E(Z)$ y $Var(Z)$.

   (b) En una titulación, la lectura inicial en una bureta es de 3.51ml y la lectura final es de 15.67ml. Para ambas mediciones se sabe que el error de medición tiene una desviación estándar de 0.02ml.

      i. Calcule el valor estimado del volumen utilizado.

      ii. ¿Cuál es la desviación estándar de su error de medición?

2. 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. Observe 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) Exprese la proporción muestral de encuestados a favor de la despenalización en términos de las variables X_i.

   (b) Proponga un estimador para p a partir de las variables X_i. Observe que el estimador es una variable aleatoria.

   (c) Halle 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, acote 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?

3. Se desea determinar una magnitud μ. Para ello se realizan n medidas repetidas, es decir, se realizarán n mediciones de la misma magnitud en idénticas condiciones, que denotaremos con $X_1, ..., X_n$. 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 $\varepsilon_1, ..., \varepsilon_n$ son variables aleatorias independientes e idénticamente distribuidas (v.a.i.i.d.) con esperanza cero y varianza $\sigma^2 = 0.25$. Note que los errores son no observables. El supuesto de que $E(\varepsilon_i) = 0$ refleja la creencia en que el método de medición empleado es exacto. Es decir no produce errores sistemáticos. La varianza del error, $\sigma^2 = Var(\varepsilon_i)$ 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.

(a) Halle $E(\overline{X}_n)$ y $Var(\overline{X}_n)$.

(b) Para $n = 10$ y $n = 100$ mediciones, usando la desigualdad de Chebyshev, encuentre una cota inferior para
$$P\left(|\overline{X}_n - \mu| < 0.1\right).$$

(c) Determine cuan grande debe ser n para que $P\left(|\overline{X}_n - \mu| < 0.1\right) \geq 0.99$, usando nuevamente la desigualdad de Chebyshev.

Asuma 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. Asuma también que la desviación estándar es 0.5, es decir $\varepsilon_i \sim \mathcal{N}(0, 0.25)$.

(d) Obtenga la distribución de $\overline{X}_n$, su esperanza y su varianza.

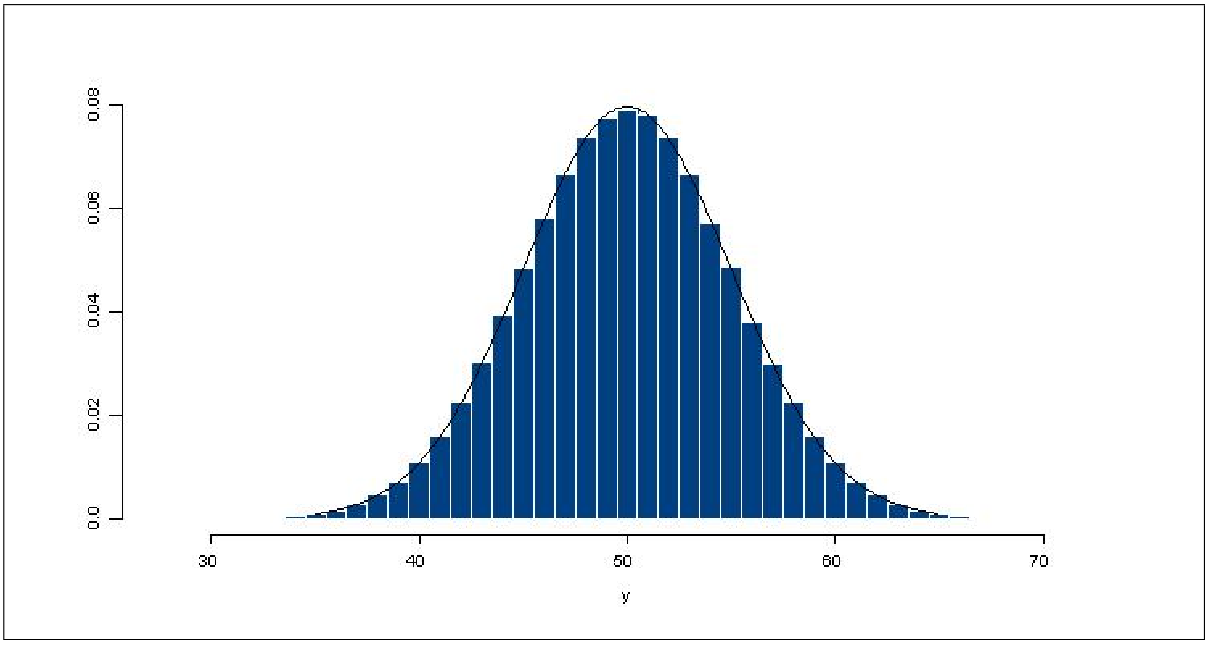
(e) Calcule la probabilidad de que el promedio de $n = 10$ mediciones y de $n = 100$ mediciones diste de la verdadera magnitud μ en menos de 0.1 unidades. Note que no fue necesario conocer el valor de μ para realizar este cálculo. Compare los resultados obtenido con los valores hallados en el ítem b).

(f) Obtenga una expresión para la probabilidad de que la medición diste de la verdadera magnitud μ en menos de 0.1 unidades en función de n. Estudie su monotonía y el límite cuando n tiende a infinito de esta probabilidad.

(g) Determine cuan grande debe ser n para que $P\left(|\overline{X}_n - \mu| < 0.1\right) \geq 0.99$. Compare el resultado obtenido con el valor hallado en el ítem c).

4. Considere nuevemente el modelo propuesta en el ejercicio 3, asumiendo ahora que $\varepsilon_i \sim \mathcal{U}(-\sqrt{3}/2, \sqrt{3}/2)$. Calcule en forma aproximada la probabilidad $P\left(|\overline{X}_n - \mu| < 0.1\right)$, para $n = 100$ y determine cuan grande debe ser n para que $P\left(|\overline{X}_n - \mu| < 0.1\right) \geq 0.99$.

5. La figura siguiente muestra



la función de probabilidad puntual en k graficada sobre el intervalo de longitud uno centrado en k, para todos los enteros k entre 0 y 100, junto con una curva normal $\mathcal{N}(\mu, \sigma^2)$. La función recién descripta está definida por

$$g(x) = \begin{cases} p_X(k) & \text{si } k - \frac{1}{2} \leq x < k + \frac{1}{2}, \text{ para } k \text{ entero entre 0 y 100} \\ 0 & \text{en otro caso} \end{cases}$$

donde $X \sim Bi\,(100, 0.5)$.

(a) ¿Cuáles son los valores de μ y σ^2 que corresponden a la curva normal dibujada?

(b) La probabilidad de obtener 52 caras en 100 tiradas de una moneda equilibrada, ¿es exactamente igual al área entre 51.5 y 52.5 bajo la curva normal o al área bajo el gráfico de g? ¿Son muy distintas entre sí? ¿Qué resultado teórico está usando?

(c) La probabilidad de obtener 52 caras o más en 100 tiradas de una moneda equilibrada puede ser aproximada por ¿cuál zona del gráfico? Calcúlela.

6. Se tira 100 veces un dado de 6 caras. Use 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. Un negocio de mascotas ofrece el siguiente servicio para sus clientes que toman vacaciones. El servicio consiste en alquilarle al dueño de la mascota un dispenser automático de raciones diarias de alimento balanceado. Cuando el animal termina de comer una ración, automáticamente, el dispenser pone a su disposición la ración siguiente. El tiempo (en días) que el gato de Felipe demora en comer una ración de alimento es una variable aleatoria Exponencial con parámetro $\lambda = 2$. Se puede suponer que los tiempos que tarda en comer cada ración son independientes entre sí.

(a) Felipe se va 30 días de vacaciones y contrata este servicio con 62 raciones. Si el gato no tiene comida disponible se escapa de la casa. ¿Cuál es, aproximadamente, la probabilidad de que Felipe encuentre al gato en su hogar al volver de las vacaciones?

(b) ¿Cuántas raciones tiene que comprar si quiere encontrar al gato en su casa al volver de las vacaciones con una probabilidad 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 $18Tm$ y desvío estándar $1.3Tm$.

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

(b) Calcule 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) Calcule **aproximadamente** la probabilidad de que la carga total transportada en 256 días esté entre 7950 y 8000Tm.

(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 $n_A = 35$ 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. Calcule la probabilidad de que la distancia entre las medias muestrales sea a lo sumo 3.

(c) Calcule 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 a) si $n_A = 400$ y $n_B = 350$? Justifique.