TEC | Tecnológico de Costa Rica

Suma y promedio de Normales
Algunas desigualdades

Escuela de Matemática

Semana 14

Giovanni Sanabria Brenes

Luis Ernesto Carrera Retana

Erick Chacón Vargas

Mario Marín Sánchez

Rebeca Solís Ortega

Suma y promedio de Normales

Teorema 1 (La suma y promedio de distribuciones normales)**.**

Sean $X_1, X_2, X_3, \cdots, X_n$ variables aleatorias mutuamente independientes. Suponga que cada variable X_i sigue una distribución normal con media μ_i y variancia σ_i^2 , o sea $X_i \sim N(\mu_i, \sigma_i^2)$, $\forall i = 1, \cdots, n$. Se cumple que:

1. La suma de distribuciones normales (S_n) es normal, o sea:

$$S_n = X_1 + X_2 + \cdots + X_n \sim N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$$

2. El promedio de distribuciones normales ($\bar{X}$) es normal, o sea:

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \sim N\left(\frac{\mu_1 + \mu_2 + \cdots + \mu_n}{n}, \frac{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2}{n^2}\right)$$

Ejemplo 1.

Por experiencia se sabe que el tiempo que le toma a una persona llenar un formulario se distribuye normalmente con una media de 10 minutos y una desviación estándar de 2 minutos.

1. Si cinco personas llenan el formulario, ¿cuál es la probabilidad de que en promedio tarden menos de 11 minutos?
2. Si cinco personas deben llenar el formulario, ¿cuál es la probabilidad de que todas ellas tarden menos de 48 minutos?

Solución:

1. Sean $x_1, x_2, \cdot, x_5$ el tiempo que tarda cada una de las personas en llenar el formulario, como se sabe que cada uno de los tiempos sigue una distribución normal se cumple:

$$\begin{aligned}\bar{X} = \frac{x_1 + x_2 + \cdots + x_5}{5} \quad &\sim N\left(\frac{\mu_1 + \mu_2 + \cdots + \mu_5}{5}, \frac{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_5^2}{5^2}\right) \\ &= N\left(\frac{10 + 10 + 10 + 10 + 10}{5}, \frac{2^2 + 2^2 + 2^2 + 2^2 + 2^2}{5^2}\right) \\ &= N\left(10, \frac{4}{5}\right)\end{aligned}$$

Por lo que se tiene:

$$P[\bar{X} < 11] = \Phi\left(\frac{11 - 10}{\frac{2}{\sqrt{5}}}\right) = \Phi\left(\frac{\sqrt{5}}{2}\right) \approx 0.8686$$

2. En este caso se cumple:

$$\begin{aligned} S_5 = x_1 + x_2 + \cdots + x_5 &\sim N\left(\mu_1 + \mu_2 + \cdots + \mu_5, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_5^2\right) \\ &= N\left(10+10+10+10+10, 2^2+2^2+2^2+2^2+2^2\right) \\ &= N(50, 20) \end{aligned}$$

   Por lo que se tiene:
   $P[S_5 < 48] = \Phi\left(\frac{48-50}{2\sqrt{5}}\right) = \Phi\left(-\frac{1}{\sqrt{5}}\right)$

   Por simetría tenemos:
   $\Phi\left(-\frac{1}{\sqrt{5}}\right) = 1 - \Phi\left(\frac{1}{\sqrt{5}}\right) = 0.33$

El siguiente teorema es consecuencia directa del teorema anterior.

Teorema 2 (La suma y promedio de distribuciones normales idénticas)**.**

Sean $X_1, X_2, X_3, \cdots, X_n$ variables aleatorias mutuamente independientes que siguen la misma distribución normal, es decir tiene la misma media (μ) y la misma variancia (σ^2), o sea $X_i \sim N(\mu, \sigma^2)$, $\forall i = 1, \cdots, n$. Se cumple que:

1. La suma de distribuciones normales (S_n) es normal, o sea:

$$S_n = X_1 + X_2 + \cdots + X_n \sim N(n\dot{\mu}, n\dot{\sigma}^2)$$

2. El promedio de distribuciones normales ($\bar{X}$) es normal, o sea:

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

1. Desigualdades de Markov y Chebyshev

Si X es una variable aleatoria discreta con rango R_X y suponga que $P[X < 0] = 0$, se tiene:

$$\begin{aligned} E(X) &= \sum_{R_X} x P[X = x] \\ &= \sum_{x<t} x P[X = x] + \sum_{x \geq t} x P[X = x] \\ &\geq \sum_{x \geq t} x P[X = x], \, pues \, P[X < 0] = 0 \\ &\geq \sum_{x \geq t} t \, P[X = x] \\ &= t \, P[X \geq t] \end{aligned}$$

Este análisis, que se puede hacer en forma equivalente para distribuciones continuas, da lugar al siguiente teorema conocido como la desigualdad de Markov.

Teorema 3 (Desigualdad de Markov)**.**

Si X es una variable aleatoria, cuya esperanza es $E(X)$, para la cual $P[X < 0] = 0$ se cumple la desigualdad

$$P[X \geq t] \leq \frac{E(X)}{t} \tag{1}$$

Esta desigualdad permite hacer aproximaciones acerca del comportamiento de variables aleatorias tomando en cuenta únicamente la esperanza. Veamos el siguiente ejemplo.

Ejemplo 2.

En una caja hay 10 bolillas rojas y 6 negras. Se extraen con remplazo 8 bolillas y se registra el número de bolillas rojas extraídas. Determine la probabilidad de que se obtengan 6 o más bolillas rojas.

Solución:

Sea X la variable:número de bolillas rojas que se extraen. Se cumple que X sigue una distribución binomial:

$$X \sim B(8, 10/16)$$

Así tenemos:

$$P[X \geq 6] = \sum_{k=6}^{8} C(8,k)(10/16)^k(6/16)^{8-k} = 0.3697$$

Utilizando la desigualdad de Markov se obtiene que

$$P[X \geq 6] \leq E(X)/6 = 5/6 = 0,8333.$$

Comparando estos dos valores nos damos cuenta que la cota que se obtiene por la desigualdad de Markov no necesariamente es buena.

Es importante notar que si en la desigualdad (1) utilizamos $t = n\mu_X$ obtenemos

$$P[X \geq n\mu_X] \leq \frac{1}{n}$$

Es decir la probabilidad de que los valores de una variable aleatoria estén a más de n veces la media es menor a $1/n$.

Si además de la media o esperanza se conoce la varianza entonces existe la posibilidad de hacer acotaciones con un poco más de precisión. Supongamos que X es una variable aleatoria con esperanza μ_X y varianza σ_X^2. Si consideramos la variable aleatoria $(X - \mu_X)^2$, aplicando la desigualdad de Markov a esta variable, cuya esperanza es precisamente σ_X^2, se obtiene

$$P[(X-\mu_X)^2 \geq t^2] \leq \frac{\sigma_X^2}{t^2}$$

Dado que $(X-\mu_X)^2 \geq t^2$ es equivalente a $|X-\mu_X| \geq t$ se obtiene el teorema conocido como la desigualdad de *Chebyshev*.

Teorema 4 (Desigualdad de Chebyshev)**.**

Si X es una variable aleatoria con media μ_X y varianza σ_X^2. Entonces para cualquier valor $t>0$ se tiene

$$P[|X-\mu_X)| \geq t] \leq \frac{\sigma_X^2}{t^2} \tag{2}$$

La desigualdad de Chebyshev permite acotar la probabilidad de que los valores de la distribución queden alrededor de la media. Estas aproximaciones no necesariamente son buenas, no obstante mejorar los resultados que se pueden obtener con esta desigualdad implicaría restringir mucho las hipótesis iniciales, como se verá en un ejemplo posterior.

Ejemplo 3.

Una persona puede digitar un texto en un tiempo que sigue una distribución con media 50 minutos y desviación 10 minutos. Determine una cota para la probabilidad de que esta persona tarde entre 30 y 50 minutos digitando un texto.

Solución:

Para estimar una cota para la probabilidad solicitada se puede recurrir a la desigualdad de Chebyshev y se obtiene

$$P[30 \leq T \leq 70] = 1 - P[|X-50| \geq 20] \geq 1 - \frac{100}{400} = \frac{3}{4}$$

Aún cuando este tipo de estimaciones pueden resultar innecesarias cuando se dispone de la distribución de probabilidad el ejemplo siguiente sirve para comparar los resultados que se obtienen usando la de la desigualdad (2).

Ejemplo 4.

El tiempo que tarda un computador en resolver un problema sigue una distribución exponencial con media 2 minutos. Determine la probabilidad que el tiempo de solución de un problema al azar esté entre 0 y 6 minutos.

Para estimar dicha la probabilidad utilizamos la desigualdad de Chebyshev, y obtenemos:

$$P[0 \leq T \leq 6] = 1 - P[|X-2| \geq 4] \geq 1 - \frac{4}{16} = 0.75$$

Es decir con la desigualdad de Chebyshev obtenemos que la probabilidad de que el tiempo esté entre 0 y 6 minutos es superior 0.75.

Por otro lado, si usamos la distribución en forma directa obtenemos que:

$$P[0 \leq T \leq 6] = \int_0^6 \frac{e^{-x/2}}{2}\,dx = 1 - e^{-3} = 0.95$$

Lo que nos indica que las cotas que se obtienen de la desigualdad de Chebyshev pueden no ser muy buenas, no obstante como veremos no es tan fácil mejorar las cotas que se obtienen con esta desigualdad sin imponer restricciones adicionales.

El siguiente ejemplo, es ilustrativo en ese sentido.

Ejemplo 5.

Sea X una variable aleatoria discreta cuya distribución de probabilidad se da en la siguiente tabla:

	$x_1 = -2$ $P[X = x_1] = 1/8$	$x_2 = 0$ $P[X = x_2] = 3/4$	$x_3 = 2$ $P[X = x_3] = 1/8$

Determine $P[|X - \mu_X| \geq 2]$

Solución

Es muy sencillo verificar que $E[X] = 0$ y que $VAR[X] = 1$. Si aplicamos la desigualdad de Chebyshev obtenemos que:

$$P[|X - \mu_X| \geq 2] \leq \frac{1}{4}$$

que en este coincide con el valor real, pues

$$P[|X - \mu_X| \geq 2] = P[X = 2] + P[X = -2] = \frac{1}{8} + \frac{1}{8} = \frac{1}{4}$$

Este ejemplo indica que aun cuando las cotas obtenidas de la desigualdad (2) no siempre son buenas a veces son exactas.

Ejercicios propuestos

1. Los pesos de una población se distribuyen normalmente con una media de 60 kilos y una varianza de 25 kilos. Se desea elegir una muestra de manera tal que haya una probabilidad superior al 90% de que el peso promedio de la muestra esté entre 55 y 65 kilogramos.

   Solución:

   No se conoce n, pero se desea que:

$$P\left[55 < \frac{x_1 + \cdots + x_n}{n} < 65\right] = 0.90$$

Como la normal es simétrica con respecto a $\mu = 60$, entonces esto equivale a

$$P\left[\frac{x_1 + \cdots + x_n}{n} < 65\right] = 0.95.$$

Como $\bar{X} = \dfrac{x_1 + \cdots + x_n}{n} \sim N\left(60, \dfrac{25}{n}\right)$ se debe resolver

$$\Phi\left(\frac{65-60}{\frac{5}{\sqrt{n}}}\right) = 0.95$$

2. Una máquina que produce chocolates elabora barras que pesan en promedio 100 gramos con una varianza de 9 gramos^2. Con base a esta información resuelva los siguientes problemas:

   a) Si se empacan cajas con 100 de estos chocolates, detemine la distribución de probabilidad de la variable: total de barras en la caja cuyo peso sobrepasa los 102 gramos.

   Solución:
   En este caso nos encontramos ante una situación que involucra dos variables, así sea:

   - X la variable: el peso de una barra de chocolate.
   - Y la variable: total de barras que exceden los 102 gramos.

   Note que para la variable X se está ante una distribución normal: $N(100, 9)$, donde se pide:

$$\begin{aligned} P[X > 102] &= 1 - P[X \leq 102] \\ &= 1 - \Phi\left(\frac{102-100}{3}\right) \\ &\approx 1 - \Phi(0.66) \\ &\approx 1 - 0.7454 \\ &= 0.2546 \end{aligned}$$

   Luego se puede analizar que la variable Y sigue una distribución binomial, cuya probabilidad asociada es la calculada con $P[X > 102]$, así la distribución de probabilidad de dicha variable corresponde a:

$$b(x; 100, 0.2546) = \begin{cases} \binom{100}{x} 0.2546^x \, (0.7454)^{100-x} & \text{si x = 0,1,2...,n} \\ 0 & \text{en cualquier otro caso} \end{cases}$$

b) Se establece que las barras deben pesar entre 99 y 102 gramos, caso contrario se consideran inapropiadas. Un control de calidad consiste en revisar las barras que salen hasta obtener la primer barra inapropiada. Determine la probabilidad de que este control lleve a revisar más de 25 barras.

Solución:
En este caso nos encontramos ante una situación que involucra dos variables, así sea:

- X la variable: el peso de una barra de chocolate.
- Y la variable: cantidad de barras a revisar.

Note que para la variable X se está ante una distribución normal: $N(100, 9)$, donde se pide:

$$\begin{aligned} P[99 \leq X \leq 102] &= \Phi\left(\frac{102-100}{3}\right) - \Phi\left(\frac{99-100}{3}\right) \\ &\approx \Phi(0.66) - \Phi(-0.33) \\ &\approx 0.7454 - 0.3707 \\ &= 0.3747 \end{aligned}$$

Así existe una probabilidad de 0.3747 que las barras se clasifiquen como apropiadas, por lo que la probabilidad de que sean inapropiadas es de 0.6253.
Luego se puede analizar que la variable Y sigue una distribución geométrica, por lo que la probabilidad final buscada corresponde a:

$$\begin{aligned} P[Y > 25] &= 1 - P[Y \leq 25] \\ &= 1 - \sum_{i=1}^{25} (0.6253)(0.3747)^{i-1} \\ &\approx 0 \end{aligned}$$

c) Si una caja contiene 100 barras determine la probabilidad de que el peso promedio de las barras en la caja sea inferior a 99 gramos.

Solución:
Sean $x_1, x_2, \cdot, x_{100}$ el peso de cada una de las barras, como se sabe que los pesos siguen una distribución normal se cumple:

$$\bar{X} = \frac{x_1 + x_2 + \cdots + x_{100}}{100} \quad \sim N\left(100, \frac{9}{100}\right)$$

Por lo que se tiene:

$$P[\bar{X} < 99] = \Phi\left(\frac{99-100}{\sqrt{\frac{9}{100}}}\right) = \Phi(-3.33) \approx 0.0004$$

3. Sea X una aleatoria con media μ y varianza σ^2 recuerde que utilizando la desigualdad de Chebishev se tiene que para todo $k > 0$:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

a) Sea $k > 0$, verifique que $P(|X-\mu| < k\sigma) \geq \dfrac{k^2-1}{k^2}$.
Note que

$$P(|X-\mu| < k\sigma) = 1 - P(|X-\mu| \geq k\sigma)$$

y por desigualdad de Chebishev $\left[P(|X-\mu| \geq k\sigma) \leq \dfrac{1}{k^2}\right]$ se tiene que

$$P(|X-\mu| < k\sigma) = 1 - P(|X-\mu| \geq k\sigma) \geq 1 - \frac{1}{k^2} = \frac{k^2-1}{k^2}.$$

b) Si X es una variable aleatoria continua tal que $E(X) = 25$ y $Var(X) = 4$. Acote inferiormente $P(20 < X < 30)$.
Note que

$$P(20 < X < 30) = P(|X-25| < 5) = P(|X-\mu| < \frac{5}{2}\sigma) \geq \frac{\left(\frac{5}{2}\right)^2 - 1}{\left(\frac{5}{2}\right)^2} = 0.84$$