Probabilidad y Estadística I
Semana 6

Variables aleatorias discretas: ejercicios. Variables aleatorias continuas: función de densidad de probabilidad

Profesor: Nicolás López MSc

Universidad del Rosario

Contenido

(Blanco) Para la variable aleatoria X se tiene:

$$F(x) = \begin{cases} 0, & \text{para } x < -\sqrt{2} \\ 3/5, & \text{para } -\sqrt{2} \leq x < \pi \\ 1, & \text{para } x \geq \pi \end{cases}$$

¿Cuál es la *fmp* de X?

Universidad del Rosario

(Blanco) Para la variable aleatoria X se tiene:

$$F(x) = \begin{cases} 0, & \text{para } x < -\sqrt{2} \\ 3/5, & \text{para } -\sqrt{2} \leq x < \pi \\ 1, & \text{para } x \geq \pi \end{cases}$$

La *fmp* de X está dada por

$$f(x) = \begin{cases} 3/5, & \text{para } x = -\sqrt{2} \\ 2/5, & \text{para } x = \pi \\ 0, & \text{eoc} \end{cases}$$

(Blanco) Para la variable aleatoria X se tiene:

$$F(x) = \begin{cases} 0, & \text{para } x^{<}0 \\ x^2/4, & \text{para } 0 \leq x < 2 \\ 1, & \text{para } x \geq 2 \end{cases}$$

¿Es X discreta o continua?

(Blanco) Para la variable aleatoria X se tiene:

$$F(x) = \begin{cases} 0, & \text{para } x < 0 \\ x^2/4, & \text{para } 0 \leq x < 2 \\ 1, & \text{para } x \geq 2 \end{cases}$$

Al no tener saltos en $F(x)$, se tiene que X es una v.a. **continua**.

(Blanco) Para la variable aleatoria X se tiene:

$$P(X=x)=\begin{cases} 5/9, & \text{para } x=-1 \\ 4/9, & \text{para } x=3/2 \\ 0, & \text{eoc} \end{cases}$$

¿Qué valores toma la v.a.? ¿Cómo se define $F(x)$?

(Blanco) Para la variable aleatoria X se tiene:

$$P(X=x)=\begin{cases} 5/9, & \text{para } x=-1 \\ 4/9, & \text{para } x=3/2 \\ 0, & \text{eoc} \end{cases}$$

Se tiene $F(x)$ como

$$F(x)=\begin{cases} 0, & \text{para } x<-1 \\ 5/9, & \text{para } -1\leq x<3/2 \\ 1, & \text{para } x\leq 3/2 \end{cases}$$

Sea $X =$"Número de personas que ingresan diariamente a UCI en Teusaquillo". Dónde $X \sim P(\lambda = 10)$. ¿Cuál es la probabilidad que sean admitidos 3 personas o menos a la UCI en un día normal?

Sea X ="Número de personas que ingresan diariamente a UCI en Teusaquillo". Dónde $X \sim P(\lambda = 10)$. La probabilidad que sean admitidos 3 personas o menos a la UCI en un día normal está dada por:

$$P(X \leq 3) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$

Dónde $X \sim P(\lambda)$, con lo cual

$$P(X = x) = \frac{\lambda^x}{x!} \exp -\lambda$$

Para $x = 0, 1, .., n$. Se tiene como resultado:

$$P(X \leq 3) = 0{,}01$$

Contenido

Universidad del Rosario

Introducción

Recordemos algunas definiciones

1. Variable
2. Experimento aleatorio.
3. Espacio y espacio muestral.

Introducción

Para el escenario continuo, la v.a. real toma valores en un conjunto continuo en los reales. En contraste, las probabilidades puntuales son siempre iguales a cero en este caso, es decir:

$$P(X = x) = 0, \text{ para cualquier } x \in \mathbb{R}$$

De manera intuitiva, podemos ver que esto tiene sentido bajo **regularidad estadística**. ¿Qué tan frecuente resulta un experimento con v.a. continua en el **exacto** mismo resultado?. ¿Cómo contrasta con el escenario discreto?.

Introducción

Entonces ¿cómo establecer una medida de probabilidad en este escenario continuo? Volvamos a pensar en estatura como variable, ahora como variable aleatoria real continua, proveniente de un experimento aleatorio.

Introducción

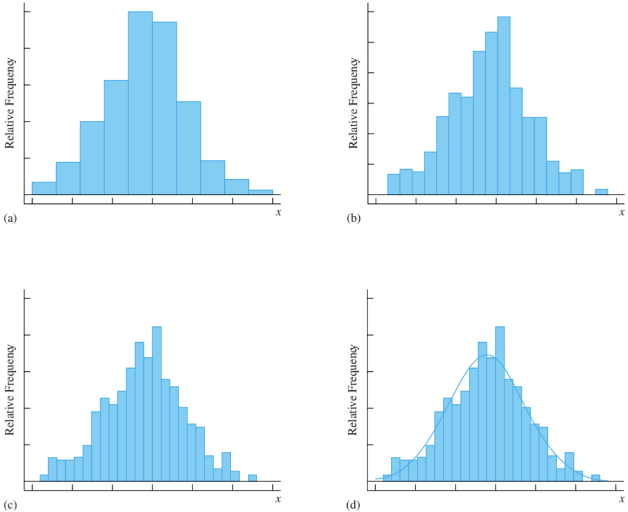
Supongamos una muestra de tamaño n de una población determinada. Medimos su estatura, construimos un histograma de frecuencias relativas con la regla de la raíz cuadrada y añadimos su polígono de frecuencias correspondiente. ¿Cómo se vería el gráfico con?

- $n = 9$, $k = 3$.
- $n = 49$, $k = 7$.
- $n = 81.$, $k = 9$.
- $n = 100$, $k = 10$.
- $n = 1000$, $k = 30$.

Universidad del Rosario

Introducción

Figura 1: Histogramas con múltiples n y k de una población determinada



Introducción

Probablemente se haya notado que:

1. Para cualquier n, la suma de las alturas del histograma es siempre igual a uno (al igual que la fmp del caso discreto).
2. Para n pequeño resultan intervalos vacíos ¿significa esto que en la población no hay sujetos con estas estaturas?
3. A medida que n crece, la frecuencia relativa en cada intervalo se vuelve mas y más precisa (mas cercana a la probabilidad de que el sujeto esté en un intervalo dado).
4. Podemos seguir incrementando los sujetos y el número de intervalos, pero llega un punto el el cual el polígono de frecuencias no cambia mucho.
5. El polígono de frecuencias representa la distribución de los datos provenientes del experimento aleatorio.

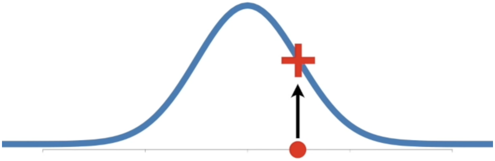
Podríamos pensar en una función $f_X(x)$ (similar a la fmp), ahora sobre la discretización de la variable aleatoria continua.

Variables aleatorias continuas

Función de densidad

Se tiene de manera análoga a la fmp a la **función de densidad de probabilidad** (fdp) como una regla que le asigna a cualquier valor $x \in \mathrm{R}$ su **densidad** o **verosimilitud**. Recuerde que la fmp asigna a cualquier valor $x \in \mathrm{R}$ su probabilidad, la fdp asigna su densidad.

Figura 2: Verosimilitud ($+$) de un punto arbitrario ($\cdot$) bajo una fdp (línea azul)



¿A qué es igual el área bajo la fdp? ¿qué valores puede tomar la fdp?.

Variables aleatorias continuas

Función de distribución

La función de distribución para el caso continuo se define de la misma manera que antes. Para cualquier $x \in \mathbb{R}$ se tiene:

$$F_X(x) = P(X \leq x)$$

Sin embargo, esta no se calcula sobre una suma finita valores, sino respecto a una **infinita**.

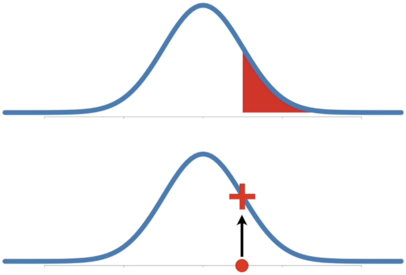


Figura 3: Función de distribución en un punto arbitrario $(\cdot)$ como el área blanca a su izquierda.

Variables aleatorias continuas

Función de distribución

Esta suma infinita acumula todas las alturas de la densidad anteriores a x, y al ser una probabilidad (pues representa $P(X \leq x)$), esta área es siempre menor o igual a cero.

Ejemplo. FDP y probabilidades

Sea X una variable aleatoria con fdp dada por:

$$f_X(x) = \begin{cases} 0{,}875 & \text{si x} \in [0.0{,}0.5) \\ 0{,}375 + x & \text{si x} \in [0.5{,}1.0) \\ 0 & \text{eoc} \end{cases}$$

¿A qué son iguales las siguientes probabilidades?

- $P(X = 0{,}5)$ y $P(X = 1{,}0)$.
- $P(X \leq 0{,}0)$ y $P(X \leq 1{,}0)$.
- $P(X \geq 0{,}5)$ y $P(X \leq 0{,}5)$.
- $P(X \geq 0{,}0)$ y $P(X \leq 1{,}0)$.
- $P(X \geq 10{,}0)$ y $P(X \leq -2{,}0)$.

Ejemplo. FDP y probabilidades

Sea X una variable aleatoria con fdp dada por:

$$f_X(x) = \begin{cases} 0{,}875 & \text{si x} \in [0.0{,}0.5) \\ 0{,}375 + x & \text{si x} \in [0.5{,}1.0) \\ 0 & \text{eoc} \end{cases}$$

¿A qué son iguales las siguientes probabilidades?

- $P(X = 0{,}5) = 0$ y $P(X = 1{,}0) = 0$.
- $P(X \leq 0{,}0) = 0$ y $P(X \leq 1{,}0) = 1$.
- $P(X \geq 0{,}5) = 0{,}5625$ y $P(X \leq 0{,}5) = 0{,}4375$.
- $P(X \geq 0{,}0) = 1$ y $P(X \leq 1{,}0) = 1$.
- $P(X \geq -10{,}0) = 1$ y $P(X \leq -2{,}0) = 0$.

Universidad del Rosario

Variables aleatorias continuas

Definición

Una v.a. se dice continua si existe una función f_X denominada función de densidad (fdp) de X tal que

$$P(X \in B) = \int_B f(x)dx$$

Para cualquier conjunto B en los reales. Recuerde que

$$P(a < X < b) = P(a \leq X < b) = P(a \leq X \leq b)$$

De manera análoga a la fmp, para que una función sea considerada fdp debe cumplir $f(x) \geq 0$ para todo x (pero no necesariamente $f(x) \leq 1$) y además:

$$P(-\infty < X < +\infty) = 1$$

Universidad del Rosario

Variables aleatorias continuas

Valor esperado (media) y varianza

Para el escenario discreto, se tenía que

$$E(X) = \sum_{i=1}^{n} x_i P(X = x_i)$$

Y además

$$V(X) = \sum_{i=1}^{n} (x_i - E(X))^2 P(X = x_i)$$

Para el escenario continuo se tiene el mismo cálculo, pero en lugar de una suma de valores, se realiza una generalización para el caso continuo.

Universidad del Rosario

Variables aleatorias continuas

Valor esperado (media) y varianza

Sea X una v.a. continua. Se tiene que

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx$$

Y además

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

Y finalmente

$$V(X) = E\left[(X - E(X))^2\right] = \int_{-\infty}^{+\infty} (x - E(X))^2 f_X(x) dx$$

Universidad del Rosario

Contenido

De manera análoga al escenario discreto, existen modelos de probabilidad parametrizados para caracterizar múltiples fenómenos. Cada uno de estos modelos de probabilidad tiene su fdp, su función de distribución, valor esperado y varianza:

- Uniforme continua.
- Exponencial.
- T de Student.
- F.
- Gamma.
- Normal.
- ...

(Blanco) Un estudiante que llega siempre a las 6:00 am a esperar el bus a la Universidad ha observado que el bus llega en cualquier momento entre las 6:00am y 6:10am. En este caso la v.a. X que denota el tiempo de espera del estudiante se distribuye uniforme continua en $[0, 10]$, es decir, $X \sim U([a = 0, b = 10])$.

Variables aleatorias continuas

Uniforme continua

Se dice que $X \sim U([a,b])$ con $a < b$ números reales si X tiene fdp dada por

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si x } \in [a,b] \\ 0 & \text{eoc} \end{cases}$$

Con

$$E(X) = \frac{a+b}{2}$$

Y

$$V(X) = E(X^2) - E^2(X) = \frac{(b-a)^2}{12}$$

Cuando la v.a. X tiene como soporte $[0, +\infty)$ (por ejemplo el tiempo hasta un evento determinado) podemos asumir que sigue una distribución exponencial. Se dice que $X \sim e(\lambda)$ con $\lambda > 0$ número real si X tiene fdp dada por

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{eoc} \end{cases}$$

Note que

$$P(X \geq t) = \int_t^{+\infty} \lambda e^{-\lambda x} dx = -e^{-\lambda x}|_t^{+\infty} = e^{-\lambda t}$$

Es decir que la probabilidad que X exceda un valor $t \geq 0$ decrece exponencialmente con t.

Variables aleatorias continuas
Exponencial

Se tiene entonces que si $X \sim e(\lambda)$ con $\lambda > 0$

$$E(X) = 1/\lambda$$

Y

$$V(X) = 1/\lambda^2$$

El tiempo hasta que ingrese la primera llamada a un contact center (c.c) es modelado como una distribución exponencial con media de 10 minutos. Si queremos calcular la probabilidad que ingrese una llamada entre los primeros 20 a 40 segundos de operación, se tiene

$$X = \text{'No. de llamadas que ingresan al c.c en 10 mins'}$$

Se tiene que

$$E(X) = \frac{1}{\lambda} = 10 \quad \longrightarrow \quad \lambda = 0{,}1$$

Note que X es medida en minutos, por lo que la probabilidad preguntada es igual a

$$P(20/60 \leq X \leq 40/60) = P(X \leq 4/6) - P(X \leq 2/6)$$

La probabilidad preguntada también puede ser escrita como

$$P(20/60 \leq X \leq 40/60) = P(X \geq 2/6) - P(X \geq 4/6)$$

Que por el resultado anterior es igual a $e^{-2/60} - e^{-4/60} = 0{,}031$ aproximadamente.

Contenido

Función de distribución

Definición

Para una v.a. discreta contamos con la fmp mientras que para una continua se tiene la fdp. En contraste la función de distribución $F_X(x)$ se tiene para ambos tipos de variables:

$$F_X(x) = P(X \leq x) = \begin{cases} \sum_{k \leq x} p(k) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{x} f(t)dt & \text{si } X \text{ es continua} \end{cases}$$

La cual cumple

- Si $x \leq y$, $F(x) \leq F(y)$.
- $F(x) \longrightarrow 0$ si $x \longrightarrow -\infty$.
- $F(x) \longrightarrow 1$ si $x \longrightarrow +\infty$.

Universidad del Rosario

Función de distribución

Definición

Recuerde que $F_x(x)$ presenta saltos si X es discreta, los cuales son iguales a $P(X = x)$, es decir, podemos obtener $p(x)$ con $F(x)$ y viceversa. Si X es continua, $F(x)$ no presenta discontinuidades, pero, ¿como se relaciona con $f(x)$ en el caso continuo?

Recuerde que

$$F_X(x) = \int_{-\infty}^{x} f(t)dt$$

Derivando se obtiene

$$f_X(x) = \frac{dF_X}{dx}(x)$$

Universidad del Rosario