Probabilidad y Estadística Fundamental

Principios de variables aleatorias. Variables aleatorias continuas y sus distribuciones

Profesor: Nicolás López

Universidad Nacional de Colombia

Contenido

Contenido

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:

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

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.

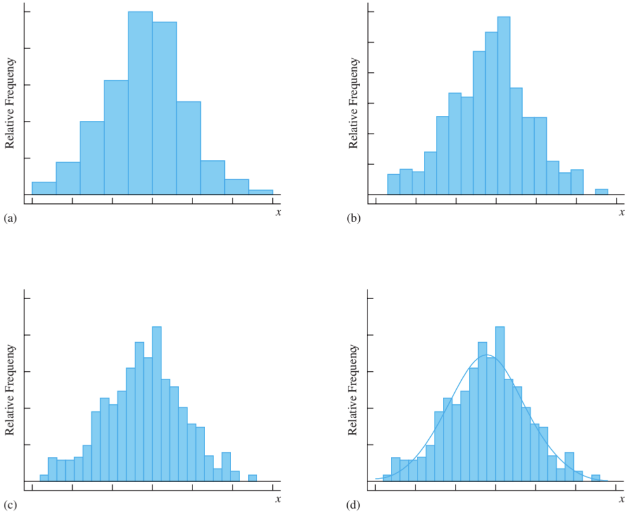
Introducción

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?

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

Introducción

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



Introducción

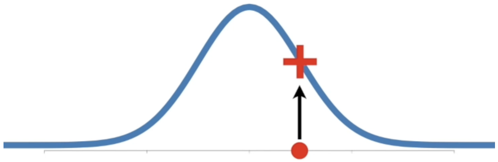
Probablemente se haya notado que:

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.

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 $x \in \mathbb{R}$ su **densidad** o **verosimilitud**. Recuerde que la fmp asigna a cualquier valor $x \in \mathbb{R}$ 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**.

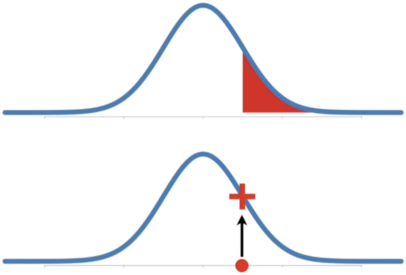


Figura 3: Función de distribución en un punto arbitrario $(\cdot)$ como el área blanca a su izquierda.

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 $P(X \leq -2{,}0) = 0$.

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. Esta requiere elementos de cálculo integral y se dejan para el estudiante interesado los detalles.

Variables aleatorias continuas

Algunas distribuciones continuas

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**.
- ...

Uno de los modelos de probabilidad más utilizados para variables aleatorias reales continuas sigue la siguiente fdp:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \text{ para } x \in \mathbb{R}$$

Si la v.a sigue la fdp descrita, se dice que X sigue una distribución normal de parámetros $\mu \in \mathbb{R}$ y $\sigma \in \mathbb{R}^+$, y se nota $X \sim N(\mu, \sigma)$.

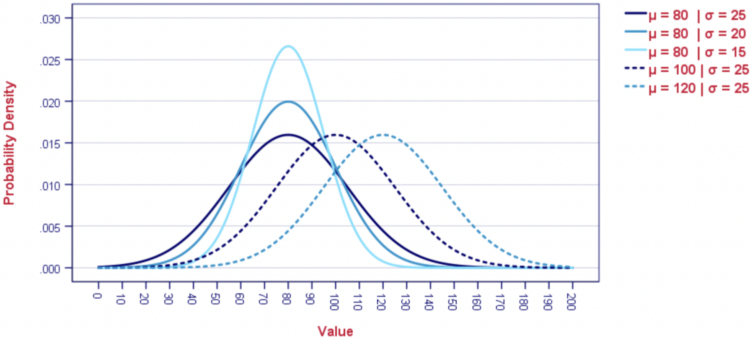
1. μ es un parámetro de **tendencia** o **localización**, mientras que σ es un parámetro de **escala**.
2. e y π son constantes numéricas aproximadamente iguales a 2,71 y 3,14, respectivamente.

Variables aleatorias continuas

Distribución normal

Así como en el caso discreto, los parámetros sirven para caracterizar el fenómeno aleatorio de interés:

Figura 4: Múltiples fdp normales con diferentes parámetros



Todas estas posibles distribuciones siguen la **regla empírica** vista en la detección de atipicidades.

Entre todas las posibles distribuciones normales indexadas por μ y σ, la distribución normal con $\mu = 0$ y $\sigma = 1$ es la más utilizada. Su fdp está dada por:

$$f_Z(z) = \frac{1}{\sqrt{2\pi(1)^2}} e^{-\frac{1}{2}\left(\frac{z-0}{1}\right)^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \text{ para } z \in \mathrm{R}$$

Si la v.a sigue la fdp descrita, se dice que Z sigue una **distribución normal estándar**. Cualquier variable $X \sim N(\mu, \sigma)$ puede ser **estandarizada** a una variable $Z \sim N(0, 1)$ como sigue:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

Figura 5: Probabilidades acumuladas. Normal estándar

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998