Probabilidad y Estadística I
Semana 7

Variables aleatorias continuas: Distribución normal, ejercicios de distribuciones desde R

Profesor: Nicolás López MSc

Universidad del Rosario

Contenido

Variables aleatorias continuas

Distribución normal

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

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

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

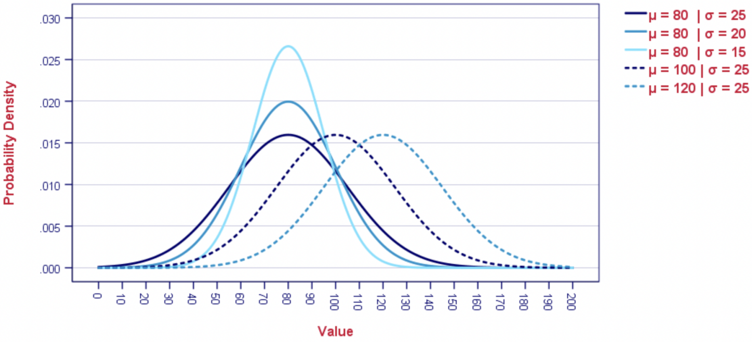
1. μ es un parámetro de **tendencia** o **localización**, mientras que σ es un parámetro de **escala**.
2. e y π son constantes numéricas aproximadamente iguales a 2,71 y 3,14, respectivamente.

Variables aleatorias continuas

Distribución normal

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

Figura 1: Múltiples fdp normales con diferentes parámetros



Variables aleatorias continuas

Distribución normal

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:

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

Si X se distribuye normal, esta sigue la **regla empírica**.

Variables aleatorias continuas

Distribución normal

Figura 2: 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

Universidad del Rosario

Contenido

Si la v.a. continua Z sigue una distribución normal estándar, ¿cuál es la probabilidad de que la variable aleatoria tome un valor menor a 1.63? ¿Cómo se expresa en notación estadística?

Puede verse que el valor es $P(Z < 1{,}63) = 0{,}9484$.

Figura 3: 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

Si la v.a. continua Z sigue una distribución normal estándar, ¿cuál es la probabilidad de que la variable aleatoria tome un valor mayor a -0.5? ¿Cómo se expresa en notación estadística?

Si la v.a. continua Z sigue una distribución normal estándar, ¿cuál es la probabilidad de que la variable aleatoria tome un valor mayor a -0.5? ¿Cómo se expresa en notación estadística?

1. Puede verse que el valor es $P(Z > -0{,}5) = 1 - P(Z < -0{,}5)$, pues $P(A) = 1 - P(A^c)$.
2. Como $P(Z < -0{,}5)$ no se encuentra tabulado, pero f_X es simétrica, $P(Z < -0{,}5) = P(Z > 0{,}5) = 1 - P(Z < 0{,}5) = 1 - 0{,}6915 = 0{,}3085$. Y finalmente $P(Z > -0{,}5) = 1 - P(Z < -0{,}5) = 1 - 0{,}3085 = 0{,}6915$.

Si la v.a. continua Z sigue una distribución normal estándar, ¿cuál es la probabilidad de que la variable aleatoria tome un valor entre -1 y 1? ¿Cómo se expresa en notación estadística?

La probabilidad es la siguiente: $P(-1 < Z < 1) = P(Z < 1) - P(Z < -1)$, con lo cual:

1. Así, al ver directamente en la tabla $P(Z < 1) = 0{,}8413$.
2. Ahora $P(Z < -1) = P(Z > 1) = 1 - P(Z < 1) = 0{,}1587$.

Así $P(-1 < Z < 1) = 0{,}8413 - 0{,}1587 = 0{,}6826$.

Si la v.a. continua Z sigue una distribución normal estándar, ¿cuál es la probabilidad de que la variable aleatoria tome un valor entre -2 y 2? ¿cuál entre -3 y 3? ¿Cómo se expresa en notación estadística?

Variables aleatorias continuas

Distribución normal

Las probabilidades son las siguientes: $P(-2 < Z < 2)$ y $P(-3 < Z < 3)$

1. Al ver directamente en la tabla, $P(Z < 2) = 0{,}9772$.
2. Ahora $P(Z < -2) = P(Z > 2) = 1 - P(Z < 2) = 0{,}0228$.

Así $P(-2 < Z < 2) = 0{,}9772 - 0{,}0228 = 0{,}9544$.

1. Al ver directamente en la tabla, $P(Z < 3) = 0{,}9987$.
2. Ahora $P(Z < -2) = P(Z > 2) = 1 - P(Z < 2) = 0{,}0013$.

Así $P(-3 < Z < 3) = 0{,}9987 - 0{,}0013 = 0{,}9974$. Este resultado junto al anterior justifican la regla empírica: 68 %, 95 %, 99 % en la detección de outliers.

Contenido

Variables aleatorias continuas

Distribución normal. Ejercicios en R

La temperatura durante septiembre está distribuida normalmente con media 18,7ºC y desviación standard 5ºC. Calcule la probabilidad de que la temperatura durante setiembre esté por debajo de 21ºC.

Sea X la v.a. real que denota la temperatura durante setiembre, el enunciado establece que $X \sim N(\mu = 18{,}7; \sigma = 5)$. Se pregunta $P(X < 21)$, con lo cual:

$$P(X < 21) = P\left(\frac{X - 18{,}7}{5} < \frac{21 - 18{,}7}{5}\right) = P(Z < 0{,}46) = 0{,}6772$$

Con lo cual, la probabilidad de que la temperatura durante setiembre esté por debajo de 21ºC es del 67 % aproximadamente.

Universidad del Rosario

Note que la probabilidad preguntada puede ser calculada directamente desde R mediante la función `pnorm()` de la siguiente forma:

$$P(X < 21) = \texttt{pnorm(21,mean=18.7,sd=5)} = 0{,}6772$$

Obteniendo el resultado sin usar las tablas de probabilidad normal.

El IQ infantil a los 5 años está distribuido normalmente con media 100 y desviación standard 16. Calcule la probabilidad de que un niño tomado al azar tenga un IQ mayor o igual a 90.

Variables aleatorias continuas

Distribución normal. Ejercicios en R

Sea X la v.a. real que denota el IQ infantil a los 5 años, el enunciado establece que $X \sim N(\mu = 100; \sigma = 16)$. Se pregunta $P(X > 90)$, con lo cual:

$$P(X > 90) = 1 - \texttt{pnorm(90,mean=100,sd=16)} = 0{,}7340145$$

Verifique este resultado mediante el uso de la tabla de la distribución normal.

Universidad del Rosario

El consumo de gasolina en mpg de un vehículo X está distribuido normalmente con media 25.5 galones y desviación standard de 4.5 galones. Calcule el percentíl 95 de la distribución de X

Sea X El consumo de gasolina en mpg de un vehículo, el enunciado establece que $X \sim N(\mu = 25{,}5; \sigma = 4{,}5)$. Se pregunta el x_0 tal que $P(X < x_0) = 0{,}95$, con lo cual:

$$P(X < x_0) = 0{,}95 \longrightarrow P\left(\frac{X - 25{,}5}{4{,}5} < \frac{x_0 - 25{,}5}{4{,}5}\right) = 0{,}95$$

Con lo cual, se busca

$$P\left(Z < \frac{x_0 - 25{,}5}{4{,}5}\right) = 0{,}95$$

Con lo cual, buscando en la tabla se tiene de manera aproximada que

$$\frac{x_0 - 25{,}5}{4{,}5} = 1{,}65$$

Es decir que $x_0 = 32{,}925$ usando la tabla.

Universidad del Rosario

Note que el percentíl preguntado puede ser calculada directamente desde R mediante la función `qnorm()` de la siguiente forma:

$$P(X < x_0) = 0{,}95 \rightarrow \texttt{qnorm(0.95,mean=25.5,sd=4.5)} = 32{,}9018$$

Obteniendo el resultado sin usar las tablas de probabilidad normal.

El tiempo de espera en una caja de pago está distribuido de manera exponencial, con un tiempo promedio de espera de 5 minutos. Calcule la probabilidad de que un cliente deba esperar menos de 10 minutos y también el percentíl 50 de la distribución.

Sea T la v.a. real que denota el tiempo de espera en una caja de pago, el enunciado establece que $T \sim e(\lambda = 1/5)$, pues $E(T) = 5 = 1/\lambda$. Se pregunta $P(T < 10)$, con lo cual:

$$P(T < 10) = 1 - P(T > 10) = 1 - e^{-10/5} = 0{,}865$$

Este resultado también puede ser encontrado desde R mediante

$$P(T < 10) = \texttt{pexp(10,rate = 1/5)} = 0{,}86466$$

Ahora, el percentíl 50 de la distribución corresponde al tiempo mediano de espera. Este es igual a t_0 que está dado por $P(T < t_0) = 0{,}5$, es decir:

$$P(T < t_0) = 0{,}5 \rightarrow 1 - e^{-t_0/5} = 0{,}5 \rightarrow e^{-t_0/5} = 0{,}5$$

Con lo cual $t_0 = -5 \times \log(0{,}5) = 3.46$. Este resultado también puede ser encontrado desde R mediante

$$P(T < t_0) = 0{,}5 \rightarrow \texttt{qexp(0.5,rate=1/5)} = 3{,}46$$

El cual es el tiempo mediano de espera (en minutos) en una caja de pago.

Una compañía de seguros de vida asegura a 5000 hombres de 45 años. La probabilidad de muerte de cualquier hombre en esa edad es de 0.001 en un año determinado. Calcule la probabilidad de 2 a 4 muertes entre los 5000 hombres.

Sea X la v.a. real que denota el número de muertes en un año determinado, el enunciado sugiere que $X \sim Bin(n = 5000; p = 0{,}001)$, con lo cual la probabilidad preguntada es igual a la suma de las 3 probabilidades siguientes:

$$P(X = 2) = \binom{5000}{2} 0{,}001^2 \times 0{,}999^{5000-2} = 0{,}084$$

$$P(X = 3) = \binom{5000}{3} 0{,}001^3 \times 0{,}999^{5000-3} = 0{,}140$$

$$P(X = 4) = \binom{5000}{4} 0{,}001^4 \times 0{,}999^{5000-4} = 0{,}175$$

Igual a 0.4, que también puede verse como:

$$P(2 \leq X \leq 4) = P(X \leq 4) - P(X < 2) = F_X(4) - F_X(1)$$

Note la diferencia con el caso continuo.

Podemos calcular la respuesta usando R de dos formas diferentes. A través de la fmp f_X:

```
dbinom(2 , size = 5000, p = 0.001) +
dbinom(3 , size = 5000, p = 0.001) +
dbinom(4 , size = 5000, p = 0.001)
```

O a través de la función de distribución F_X

```
pbinom(4,size = 5000 , p = 0.001) -
pbinom(1,size = 5000 , p = 0.001)
```

Obteniendo en ambos casos el mismo resultado.

Continuando con el ejercicio anterior, note que se tiene un $n \geq 100$, $p \leq 0{,}01$ y $np \leq 20$, con lo cual la muerte en este rango de edad representa un evento raro o inusual. Esto implica que podríamos aproximar la v.a. X a una distribución de Poisson de parámetro $\lambda = np = 5$ muertes en promedio. Calcule la probabilidad del escenario anterior mediante la aproximación de Poisson.

En este caso se obtiene desde R mediante f_Y con Y aproximadamente Poisson de parámetro 5:

```
dpois(2 , lambda = 5) +
dpois(3 , lambda = 5) +
dpois(4 , lambda = 5)
```

Y mediante F_Y

```
ppois(4,lambda = 5) -
ppois(1,lambda = 5)
```

Obteniendo nuevamente un valor muy cercano al 0.4 de la distribución exacta. Verifique que sus resultados manuales sean iguales a los computacionales.

Contenido

Distribuciónes conjuntas

Introducción

Hasta ahora nos hemos encargado de caracterizar el comportamiento **univariado** de los experimentos aleatorios. Sin embargo podríamos pensar en dos v.a. X y Y con fmp conjunta (para el caso discreto)

$$p(x, y) = P(X = x, Y = y) = P(X = x \cap Y = y)$$

Teniendo así

$$P((X, Y) \in A) = \sum_{(x,y)\in A} p(x, y)$$

La **distribución conjunta** de (X, Y), mientras que las **marginales** están dadas por

$$p(x) = \sum_{y} p(x, y) \qquad p(y) = \sum_{x} p(x, y)$$

Universidad del Rosario

Distribuciónes conjuntas

Introducción

Recuerde el experimento de lanzar dos dados corrientes. En este caso se mapean a los reales los resultados mediante las v.a. X (en filas) y Y (en columnas) indicando el número obtenido, teniendo así el **soporte conjunto** de las v.a.

$$
\begin{array}{cccccc}
\{(1,1), & (1,2), & (1,3), & (1,4), & (1,5), & (1,6), \\
(2,1), & (2,2), & (2,3), & (2,4), & (2,5), & (2,6), \\
(3,1), & (3,2), & (3,3), & (3,4), & (3,5), & (3,6), \\
(4,1), & (4,2), & (4,3), & (4,4), & (4,5), & (4,6), \\
(5,1), & (5,2), & (5,3), & (5,4), & (5,5), & (5,6), \\
(6,1), & (5,6), & (6,3), & (6,4), & (6,5), & (6,6)\}
\end{array}
$$

Distribuciónes conjuntas

Introducción

Con la cual se obtiene la **fmp conjunta** del anterior arreglo como:

$$
\begin{array}{cccccc}
\{1/36, & 1/36, & 1/36, & 1/36, & 1/36, & 1/36, \\
1/36, & 1/36, & 1/36, & 1/36, & 1/36, & 1/36, \\
1/36, & 1/36, & 1/36, & 1/36, & 1/36, & 1/36, \\
1/36, & 1/36, & 1/36, & 1/36, & 1/36, & 1/36, \\
1/36, & 1/36, & 1/36, & 1/36, & 1/36, & 1/36, \\
1/36, & 1/36, & 1/36, & 1/36, & 1/36, & 1/36,\}
\end{array}
$$

Es decir

$$p(x, y) = 1/36$$

Para $(x, y) \in \{1, ..., 6\} \times \{1, ..., 6\}$. Y la marginal de X (de Y) se obtiene como las sumas de filas (de columnas) del arreglo de probabilidades conjuntas, igual a $1/6$ para $x \in \{1, ..., 6\}$ y $1/6$ para $y \in \{1, ..., 6\}$.

Note que

$$p(x, y) = 1/36$$

Y

$$p(x) = 1/6 \qquad p(y) = 1/6$$

Para $(x, y) \in \{1, \ldots, 6\} \times \{1, \ldots, 6\}$. y $x \in \{1, \ldots, 6\}$ y $1/6$ para $y \in \{1, \ldots, 6\}$. Es decir

$$p(x, y) = p(x)p(y)$$

Lo que resembla a la definición de eventos independientes. En efecto, si $p(x, y) = p_X(x)p_Y(y)$ para todo x y y se dice que las v.a. discretas X y Y son **independientes**.

Distribuciónes conjuntas

Variables independientes. Valor esperado y varianza

Si dos v.a. X y Y son independientes, note lo siguiente

$$E(XY) = \sum_x \sum_y xyp(x,y) = \sum_x \sum_y xp_X(x) \times yp_Y(y) = E(X)E(Y)$$

Como ejercicio queda revisar que también se cumple

$$V(X+Y) = V(X) + V(Y)$$

También bajo independencia de las v.a X y Y.

Distribuciónes conjuntas

Múltiples variables independientes

Si ahora se tienen n v.a. X_1 ... X_n independientes, se cumple que

$$p(x_1, ..., x_n) = \Pi_n p_{X_i}(x)$$

Y también se tiene que

$$V(\sum_i X_i) = \sum_i V(X_i)$$

Nuevamente bajo independencia de las v.a.

Distribuciónes conjuntas

Múltiples variables independientes. Ejemplo 1

Como ejemplo, piense en $X_i \sim Ber(p)$ con $i = 1, ..., n$. Cada realización es independiente entre sí, por lo cual se tiene que $X = \sum_i X_i \sim Bin(n, p)$. Por la linealidad del valor esperado

$$E(X) = \sum_i E(X_i) = np$$

Y por la independecia de las X_i

$$V(X) = \sum_i V(X_i) = np(1 - p)$$

Universidad del Rosario

Distribuciónes conjuntas

Múltiples variables independientes. Ejemplo 2

Ahora piense en X_i con $i = 1, ..., n$ v.a. independientes indicando la aprobación del presidente. En este caso $X_i \sim Ber(p)$ con p desconocido, pero fijo. Puede ser de interés la v.a.

$$\bar{X} = \frac{\sum X_i}{n}$$

Por la linealidad del valor esperado, nuevamente

$$E(\bar{X}) = \frac{\sum_i E(X_i)}{n} = \frac{np}{n} = p$$

Universidad del Rosario

Asumiendo que cada realización es independiente entre sí:

$$V(\bar{X}) = \frac{1}{n \times n} \sum_i V(X_i) = \frac{p(1-p)}{n}$$

Y así, se espera que la media $\bar{X}$ en promedio sea igual a p. A medida que n es mayor, se tiene mayor precisión en la estimación. Este resultado se mantiene para v.a. no Bernoulli, siempre y cuando sean independientes con misma media y varianza.

Contenido

De manera semejante al caso discreto se tiene en el continuo:

$$P((X,Y) \in B) = \int\int_{(x,y)\in B} f_{X,Y}(x,y)dxdy$$

Cuando B es un rectángulo $B = \{(x,y) : a < x < b, b < y < d\}$

$$P((X,Y) \in B) = \int_c^d \int_a^b f_{X,Y}(x,y)dxdy$$

Note que si B es $\mathcal{R}^2$

$$P((X,Y) \in B) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = 1$$

De manera semejante al caso discreto se tienen las funciónes de **densidad marginales** a partir de la **densidad conjunta** como:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$

Y

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx$$

Tanto en el caso discreto como en el contínuo se tiene la función de distribución conjunta. Para el discreto se tiene

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$$

Mientras que para el continuo

$$F_{X,Y}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(s,t)dsdt$$

Y al derivar parcialmente dos veces a $F_{X,Y}(x,y)$, respecto a cada una de las dos variables, se obtiene $f_{X,Y}(s,t)$.

Contenido

Note que las v.a. $X' = X - E(X)$ y $Y' = Y - E(Y)$ representan la versión centrada de X y Y respectivamente. En una realización determinada, la variable aleatoria $X'Y'$ será positiva si ambos valores tienen el mismo signo, mientras que será negativa en caso contrario. Por lo cual

$$E(X'Y') = E([X - E(X)][Y - E(Y)])$$

Determina el valor esperado en la asociación entre las variables X y Y, a esta cantidad se le llama la covarianza de X y Y. A medida que se acerca a cero, implica que las variables son no crrelacionadas.

Universidad del Rosario

Esta cantidad puede ser estandarizada para remover las unidades de medida como

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

En el intervalo $[-1, 1]$, entre más se acerque a 1 en valor absoluto, se tiene una mayor asociación entre las v.a. X y Y.