TEMA 5. DISTRIBUCIONES DE PROBABILIDAD

INTRODUCCIÓN

Este tema es fundamental en estadística, ya que combina conceptos de **estadística descriptiva** y **probabilidad.** Las distribuciones de probabilidad son modelos teóricos que describen cómo se espera que se comporten los datos, en lugar de solo describir datos que ya han ocurrido. Son la base para la estadística inferencial, donde se usan datos de muestra para generalizar sobre una población.

CONTENIDO

5.1. Conceptos Básicos: Variable aleatoria: discreta y continua.

5.2. Distribuciones de probabilidad de variable discreta: Función probabilidad. Construcción y representación tabular, gráfica y analítica de una distribución de probabilidad de variable discreta; valor esperado y varianza de probabilidad. Modelos de probabilidad de variable discreta y sus aplicaciones en el área de la Administración de empresas. Ensayo Bernoulli, Distribución Binomial, Distribución de Poisson.

5.3. Distribuciones de probabilidad de variable continua: Función Densidad de probabilidad. Construcción y representación tabular, gráfica y fórmula matemática; valor esperado y varianza de probabilidad. Modelos de probabilidad de variable continua y sus aplicaciones. Distribución Normal. Aproximación de los modelos Binomial y Poisson a la Distribución Normal. Otros modelos de probabilidad de variable continua aplicables a la Administración de empresas.

5.4. Uso de herramientas tecnológicas: Uso de software para determinar probabilidades con el computador, usando los modelos de probabilidad estudiados, así como para conocer otros modelos

DESARROLLO

5.1. Conceptos Básicos

Variable aleatoria

Es una variable cuyo valor numérico es el resultado de un experimento aleatorio. Pueden ser de dos tipos:

1. Discreta: Toma valores que son contables, usualmente números enteros.

Ejemplo: Número de clientes que llegan a una tienda en una hora, número de productos defectuosos en un lote.

2. Continua: Toma cualquier valor dentro de un rango dado.

Ejemplo: La altura de los empleados en una empresa, el tiempo que un cliente espera en la fila.

5.2 Distribuciones de probabilidad de variable discreta: Función de probabilidad

Construcción y representación tabular, gráfica y analítica de una distribución de probabilidad de variable discreta

Una Distribución de Probabilidad es una lista de todos los resultados posibles de algún experimento aleatorio y de la probabilidad relacionada con cada resultado.

Las distribuciones de probabilidad son similares a las distribuciones de frecuencia, pero representan probabilidades teóricas de una población, mientras que las distribuciones de frecuencia representan probabilidades empíricas de muestras. Se usan para describir la probabilidad de cada valor de una variable aleatoria.

Suponga el caso en el que se lanzan tres monedas, entonces se puede definir X como la variable aleatoria que representa el número de veces que sale “Cara”. Así, X puede tomar los valores 0, 1, 2, 3. Las probabilidades de la variable X se indican a continuación:

Espacio muestral: Ω = {CCC, CCS, CSC, CSS, SCC, SCS, SSC, SSS} n = 8 (casos posibles)

Evento X = 0: {SSS} k = 1 (casos favorables)
Evento X = 1: {CSS, SCS, SSC} k = 3
Evento X = 2: {CCS, CSC, SCC} k = 3
Evento X = 3: {CCC} k = 1

Aplicando la definición clásica de probabilidad, P(X=x) = k/n. Por lo tanto:

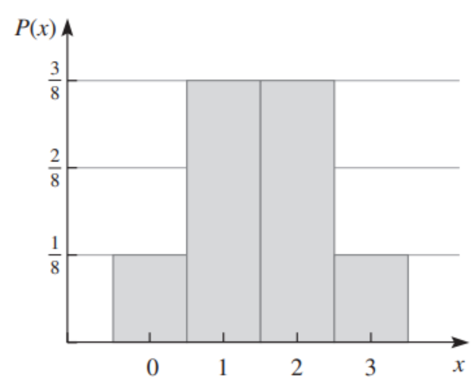
P(X=0) = 1/8
P(X=1) = 3/8
P(X=2) = 3/8
P(X=3) = 1/8

Vale la pena notar que cada una de las probabilidades es menor que 1. También puede verse que la suma de todas ellas es justamente 1.

Finalmente, la representación tabular de la distribución de probabilidad de X, en este caso, es:

X	0	1	2	3
P(X)	1/8	3/8	3/8	1/8

La Gráfica correspondiente a la distribución anterior es la siguiente:



En un histograma de probabilidad, el área de cada barra es igual a la probabilidad de x.

Las distribuciones de probabilidad también pueden definirse analíticamente mediante una función matemática llamada *función de probabilidad* en el caso que la variable aleatoria sea discreta. La función de probabilidad asigna una probabilidad a cada valor posible de la variable aleatoria discreta. Así, para el ejemplo que venimos trabajando, la función de probabilidad es:

$$P(x) = \binom{3}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{3-x} \qquad \text{para } x = 0, 1, 2, 3$$

Valor esperado y varianza de probabilidad

El valor esperado es la media aritmética de una distribución de probabilidad y se calcula mediante la siguiente fórmula:

$$\mu = E(X) = \Sigma[(x_i)P(x_i)]$$

Mientras que la varianza de una distribución de probabilidad es el promedio de las desviaciones al cuadrado respecto a la media. Para obtener su valor se utiliza la fórmula siguiente:

$$\sigma^2 = \Sigma[(x_i - \mu)^2 P(x_i)]$$

Siguiendo con el ejemplo de lanzar la moneda tres veces, donde X representa el número de Caras que se obtiene, la siguiente tabla muestra cómo se calculan el *valor esperado* y la *varianza* de dicha distribución:

X	*P(x)*	*x * P(x)*	*(x - μ)² * P(x)*
0	1/8	0 * 1/8 = 0	$(0 – 3/2)^2$ * 1/8 = 9/32
1	3/8	1 * 3/8 = 3/8	$(1 – 3/2)^2$ * 3/8 = 3/32
2	3/8	2 * 3/8 = 6/8	$(2 – 3/2)^2$ * 3/8 = 3/32
3	1/8	3 * 1/8 = 3/8	$(3 – 3/2)^2$ * 1/8 = 9/32
∑	*1*	*μ = 12/8 = 3/2*	*σ^2 = 24/32 = 3/4*

Desviación Estándar (σ): Es la raíz cuadrada de la varianza y mide la dispersión en las mismas unidades que la variable aleatoria.

<u>Ejemplo</u>: Si continuamos con los datos del ejemplo anterior, la desviación estándar es:

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.75} = 0.866 = \mathbf{0.87}$$

Modelos de probabilidad de variable discreta

El ***Ensayo de Bernoulli***, llamado así en honor al matemático suizo Jacob Bernoulli (1654 – 1705), es un procedimiento aleatorio en el que sólo se tienen dos resultados posibles: éxito o fracaso. Estos resultados son independientes y sus probabilidades complementarias. En la Distribución de Bernoulli, la función de densidad viene dada por:

$$P(x) = p^x q^{1-x}$$

En este modelo, p es la probabilidad de éxito y q = 1-p es la probabilidad de fracaso. La media μ = p, y la desviación estándar viene dada por σ = √pq. Este modelo puede considerarse como un caso particular del modelo binomial que se verá a continuación.

La ***Distribución Binomial***, corresponde a una variable aleatoria de n ensayos de Bernoulli, donde la probabilidad de cada resultado se mantiene constante de un ensayo al siguiente.

La función de distribución asociada a la variable de Bernoulli, viene dada por:

$$P(x) = \binom{n}{x} p^x (1-p)^{n-x},$$

Donde, x es el número de éxitos (x = 1, 2, …, n);
p es la probabilidad de éxito; y
n es el número de ensayos.

Su valor esperado viene dado por μ = np; y su varianza se calcula mediante σ2 = np(1-p).

Este modelo de probabilidad representa situaciones como la del ejemplo que hemos examinado antes en relación al lanzamiento de una moneda tres veces.

Para aplicar el modelo anterior, sólo debemos identificar que las pruebas correspondan a ensayos de Bernoulli, cuantas veces se realiza y determinar la probabilidad de éxito de las pruebas. Para el ejemplo en cuestión, esto se haría de la siguiente forma:

Ensayo de Bernoulli:	Lanzar la moneda.
Cuántas veces se realiza:	n = 3.
Variable x:	número de caras que se obtienen al lanzar tres veces la moneda.
Probabilidad de éxito:	p = ½.

Con estos datos, ya podremos calcular cualquier probabilidad de la variable X, como por ejemplo P(X=2), la probabilidad de obtener 2 caras, sustituyendo directamente estos valores en la fórmula de la Distribución Binomial y realizando las operaciones matemáticas correspondientes:

$$P(x=2) = \binom{3}{2} p^2 (1-p)^{3-2} = 3(½) = 3(½)^2(1-½)^{3-2} = 3\,(1/8) = 3/8$$

(Práctica: obtener las probabilidades faltantes, calcular la media y la varianza de esta distribución).

<u>Otro ejemplo</u>: La probabilidad de que un cliente compre un producto es de 0.3. Si 10 clientes entran en la tienda, ¿cuál es la probabilidad de que exactamente 3 compren el producto?

Se tiene: n = 10, x = 3, p = 0.3 y q= 0.7. Entonces al sustituir en la fórmula de la binomial obtenemos la probabilidad buscada:

$$P(3) = \binom{10}{3} 0.3^3 0.7^7 = 2.668$$

La *<u>Distribución de Poisson</u>*, ideada por el matemático francés Simeon Poisson (1781 – 1840), mide la probabilidad de un evento aleatorio sobre algún intervalo de tiempo o espacio. Tiene por condición que se cumplan los dos supuestos siguientes:

1) La probabilidad de ocurrencia del evento es constante para dos intervalos cualesquiera de tiempo o espacio λ
2) La ocurrencia del evento en un intervalo es independiente de la ocurrencia de otro intervalo cualquiera.

$$P(x) = \frac{\mu^x e^{-\mu}}{x!}$$

La función de probabilidad de Poisson viene dada por la siguiente expresión:

En donde, x es el número de eventos: x = 0, 1, 2, …
μ es el número promedio de ocurrencias por unidad de tiempo o espacio, y
e = 2.71828 es la base del logaritmo natural.

En la distribución de Poisson, la varianza σ^2 es igual que la media μ y se denota por el parámetro λ.

Ejemplo: Si en promedio llegan 2 clientes por hora a un negocio, ¿cuál es la probabilidad de que lleguen 4 clientes en una hora?

En este caso se tiene que $\lambda = 2$, $x = 4$. Por lo tanto, la probabilidad de que lleguen 4 clientes es:

$$P(4) = (e^{-2}\, 2^{4}) / 4! = 0.0902$$

Otro ejemplo: Si el 2% de los libros encuadernados en el taller de restauración de la biblioteca tienen encuadernación defectuosa. Obtener la probabilidad de que 5 de 400 libros encuadernados en este taller tengan encuadernaciones defectuosas.

Para resolver este problema usamos la distribución de Poisson definiendo X como el número de libros que tienen encuadernación defectuosa. Conociendo que el 2% de los libros encuadernados quedan defectuosos, entonces se espera que en 400 libros halla en promedio 8 libros con encuadernación defectuosa ($\mu = 8$). Por lo tanto, la probabilidad de que 5 libros ($X = 5$) tengan encuadernación defectuosa es:

$$\mathrm{P}[X = 5] = \frac{8^5 e^{-8}}{5!} = 0.092$$

5.3. Distribuciones de probabilidad de variable continua: Función Densidad de probabilidad

Función de Densidad de Probabilidad: Es una función que define la forma de la distribución de probabilidad para una variable continua. El área bajo la curva de la función en un intervalo dado representa la probabilidad de que la variable tome un valor en ese intervalo. La probabilidad en un punto específico es cero para una variable continua. A continuación, se presenta la más emblemática de las distribuciones de probabilidad continuas que se conoce como la distribución Normal.

Modelos de probabilidad de variable continua y sus aplicaciones

Distribución Normal

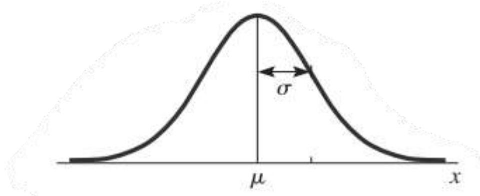
La Distribución Normal: Es la más importante de las distribuciones de probabilidad. Muchas variables aleatorias continuas tienen distribuciones que se aproximan a una distribución normal. La distribución normal se define por su media (μ) y su desviación estándar (σ).

Construcción y representación tabular, gráfica y fórmula matemática

Analíticamente, se define mediante una fórmula matemática.

$$y = f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}} \quad \text{para todo real } x$$

La función de densidad Normal se representa gráficamente con una curva continua en forma de campana:

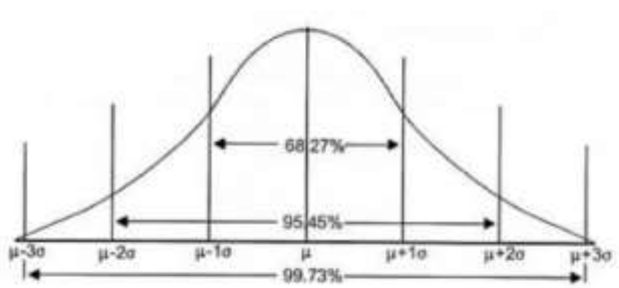


Valor esperado y varianza de probabilidad

Se calculan mediante integrales de la función de densidad, lo cual está fuera del alcance de este curso introductorio, pero se pueden calcular usando software estadístico.

La Distribución Normal Estándar

La distribución normal estándar es una distribución normal que tiene una media de 0 y una desviación estándar de 1 y la probabilidad se calcula encontrando el área bajo la curva de la función de densidad. La regla empírica indica que los intervalos de 1, 2 y 3 desviaciones estándar alrededor de la media contendrán 68, 95 y 99.7% de los datos, respectivamente.



Ejemplo:

El tiempo que un cliente espera en la fila de un supermercado se distribuye normalmente con una media de 5 minutos y una desviación estándar de 1 minuto. ¿Cuál es la probabilidad de que un cliente espere entre 4 y 6 minutos? Para esto, se utilizan tablas o software que calculan las probabilidades bajo la curva normal.

Aproximación de los modelos Binomial y Poisson a la Distribución Normal.

La distribución normal puede utilizarse para aproximar la distribución binomial cuando np y nq son mayores o iguales a 5. También se puede aproximar la distribución de Poisson a la normal para valores grandes de λ.

Este tipo de aproximación es muy útil cuando los cálculos directos son complejos o engorrosos, por ejemplo, en los casos en que se requieran muchos cálculos de factorial con la distribución binomial.

Para utilizar la distribución normal como aproximación se calcula la media y la desviación estándar de la distribución binomial o de Poisson y se trabaja con ellas como si la distribución fuera normal, haciendo las correcciones por continuidad, que transforman un valor discreto en un intervalo continuo.

Otros modelos de probabilidad de variable continua aplicables a la Administración de empresas.

Existen otros modelos como la distribución exponencial, uniforme, etc., que pueden ser de utilidad, pero para los propósitos de esta guía no se entrará en detalles.

5.4 Uso de herramientas tecnológicas

Software Estadístico: Programas como Minitab, Excel, STATDISK y calculadoras TI-83/84 son herramientas muy útiles para calcular probabilidades y visualizar distribuciones.

Estos programas permiten:

- ✓ Generar números aleatorios de diferentes distribuciones.
- ✓ Calcular probabilidades para distribuciones discretas y continuas.
- ✓ Graficar distribuciones de probabilidad.
- ✓ Calcular estadísticos descriptivos.
- ✓ Realizar simulaciones de procesos aleatorios.

Ejercicios propuestos

Variables aleatorias: Clasifique las siguientes variables como discretas o continuas:

a) Número de llamadas recibidas en un call center por hora.
b) Tiempo de espera de un cliente en un banco.
c) Número de autos vendidos por un concesionario en un mes.
d) Peso de un paquete enviado por mensajería.

Cálculo de la varianza: Para el experimento aleatorio de lanzar tres monedas, calcule la varianza a partir de la tabla de distribución de la variable aleatoria correspondiente, agregando dos columnas auxiliares, una para x^2 y otra para $[x^2 P(x)]$ y obteniendo las sumatorias respectivas para utilizarlas en la fórmula: $\sigma^2 = \Sigma [x^2 P(x)] - \mu^2$. Compare el resultado obtenido con el indicado en el ejemplo de la teoría explicada en la sección 5.3.

Distribución binomial: La probabilidad de que un artículo sea defectuoso es de 0.05. Si se inspeccionan 20 artículos, calcule:

a) La probabilidad de que exactamente 2 sean defectuosos.
b) La probabilidad de que al menos 3 sean defectuosos.
c) La media y desviación estándar del número de artículos defectuosos.

Distribución de Poisson: En promedio, llegan 5 solicitudes de información a una oficina cada hora. Calcule:

a) La probabilidad de que lleguen exactamente 3 solicitudes en una hora.
b) La probabilidad de que lleguen más de 6 solicitudes en una hora.

Distribución Normal: Las ventas diarias de una tienda se distribuyen normalmente con una media de $500 y una desviación estándar de $50. Calcule:

a) La probabilidad de que las ventas diarias estén entre $450 y $550.
b) El valor de las ventas diarias que separa al 10% de las ventas más bajas.

Herramientas tecnológicas: Utilice Excel o una calculadora con funciones estadísticas para:

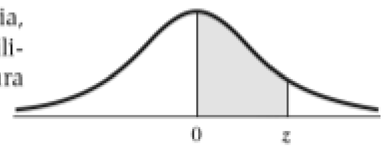
a) Generar 100 números aleatorios de una distribución normal con media 10 y desviación estándar 2.
b) Calcular la probabilidad de obtener 5 éxitos en 10 ensayos binomiales con $p=0.4$.
c) Graficar la distribución de Poisson con $\lambda = 3$.

Conclusión Este tema proporciona las herramientas necesarias para comprender y modelar la incertidumbre en los negocios. Es clave para la toma de decisiones basadas en datos y para el análisis estadístico en administración y otras áreas. El uso de la tecnología es fundamental para trabajar con distribuciones de probabilidad y realizar análisis más profundos.

TABLA DE LA DISTRIBUCIÓN Z

Áreas de la distribución normal estándar

Las entradas de esta tabla son las probabilidades de que una variable aleatoria, con una distribución normal estándar, tome un valor entre 0 y z; la probabilidad está representada por el área sombreada bajo la curva de la figura siguiente. Las áreas para valores negativos de z se obtienen por simetría.



Segundo lugar decimal en z

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998
3.5	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998	0.4998
3.6	0.4998	0.4998	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999	0.4999
3.7	0.4999									
4.0	0.49997									
4.5	0.499997									
5.0	0.4999997									

Para detalles específicos acerca del uso de esta tabla para hallar: probabilidades, vea páginas 317-320; coeficientes de confianza, páginas 338-339, 340-341; valores p, páginas 432-433, 435; valores críticos, páginas 317-320, 338-339.