Modelos Probabilísticos Discretos

Distribución Bernoulli

Sea X variable aleatoria que puede tomar solo dos valores, digamos 0 y 1, con $P(X=1)=\alpha$. Así,

$$p(x)=\begin{cases} 1-\alpha & x=0 \\ \alpha & x=1 \\ 0 & \text{caso contrario} \end{cases}$$

Entonces $X \sim$ Bernoulli de parámetro α,
$E(X)=\alpha$, $V(X)=\alpha(1-\alpha)$.

Ejemplo 1

Consideremos el ejemplo que consiste en lanzar una moneda deshonesta (la probabilidad de obtener cara es $3/4$) y anotar el resultado.
Sea X la v.a que asigna el número 1 si el resultado observado es cara, y 0 si el resultado observado es cruz.

Ejemplo 2

Supongamos que repetimos el experimento anterior 5 veces (repeticiones independientes) y que ahora
X = número de caras que se observan en las 5 repeticiones del experimento.

Observemos que:

- en cada repetición del experimento hay dos posibles resultados: Exito (sale cara), Fracaso (sale cruz)
- P(E) = 3/4 = p , P(F) = 1/4 = (1-p) = q
- Si X = número de caras que se observan en las 5 repeticiones del experimento $\Rightarrow$ los valores que puede asumir X son $i = 0, 1, 2, 3, 4, 5$

Supongamos que queremos calcular P(X=2)

Observaciones

- Un resultado favorable al evento $\{X = 2\}$ es CCXXX, pero también lo es CXXXC, ... y todas las posibles combinaciones de "ubicar" dos éxitos (C) y tres fracasos (X).
- Los distintos resultados favorables al evento $\{X = 2\}$ se diferenciarán entre sí en el orden en que se presenten los 2 resultados éxito (C) y los 3 fracasos (X), y no pueden ocurrir simultaneamente.
- Todas las 5-uplas favorables al evento $\{X = 2\}$ tienen la misma probabilidad, y la probabilidad de cada una de ellas es p^2q^3
- $P(X = 2) = Np^2q^3$, donde $N=$ número de formas diferentes de ordenar 2 éxitos (C) y (5-2) fracasos (X) en 5 lugares . Esto es, $N = P^5_{2,(5-2)} = \frac{5!}{2!(5-2)!} = C^5_2$

Luego,

$$P(X = 2) = C^5_2 p^2(1 - p)^{(5-2)}$$

Distribución Binomial

Experimento Binomial

Un experimento aleatorio que satisface las siguientes condiciones se denomina experimento binomial

1. Consiste de n repeticiones de un mismo experimento aleatorio, (n fijado previamente)
2. Cada repetición tiene solamente dos resultados posibles, denominados Exito(E) y Fracaso(F)
3. Las repeticiones son idependientes entre sí.
4. En cada repetición, $P(E) = p$, $P(F) = q = 1 - p$.

Distribución Binomial

Sea X = número de éxitos observados en las n repeticiones de un experimento binomial

Entonces la variable aleatoria X tiene distribución Binomial de parÃ¡metros n y p (n repeticiones y probabilidad de éxito p) y se denota

$$X \sim \mathcal{B}(n, p)$$

$$p(x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{cc} \end{cases}$$

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}$$

$E(X) = np$ $\qquad$ $V(X) = npq$ con $q = 1 - p$

Ejemplo 3

Al llegar a un corte de ruta, los autos pueden doblar hacia la izquierda o hacia la derecha. Suponga que los sucesivos conductores eligen una dirección independientemente de lo que hayan hecho los demás, y que la probabilidad de doblar a la izquierda sea 0.6.
a.Entre los próximos 12 autos, cual es la probabilidad que al menos 10 doblen a la izquierda?
b. Cuántos autos se espera que doblen a la izquierda?

Ejemplo 4

En cierta localidad se sabe que la distribución de los grupos sanguíneos de los habitantes es la siguiente: “0”:50%, “A”: 25%, “B”:20% y “AB” :5%. Se hace una selección aleatoria de 6 personas de esa región y se quiere calcular la probabilidad que 2 sean del grupo 0, 2 del A ,1 del B y 1 del C.

- ¿Puede usarse un modelo binomial?
- ¿Por qué?

Distribución Multinomial

Supuestos:

1. Un experimento aleatorio simple que tiene k posibles resultados $A_1, A_2, \ldots, A_k$, se repite n veces en forma independiente.
2. p_i, la probabilidad que el resultado sea A_i ($\sum_{i=1}^{k} p_i = 1$), se mantiene constante en cada repetición.
3. Sea $X_i =$ cantidad de veces que ocurre el resultado A_i a lo largo de las n repeticiones.

Si un experimento aleatorio simple se repitió n veces de acuerdo con los supuestos enunciados se dice que es un experimento Multinomial y

$P(X_1 = r_1, X_2 = r_2, \ldots, X_k = r_k) =$ probabilidad de que A_1 se presente r_1 veces, A_2 se presente r_2 veces, y A_k se presente r_k veces

Distribución Multinomial

Se dice que el vector aleatorio $\mathbf{X} = (X_1, X_2, \ldots, X_k)$ tiene distribución Multinomial con k resultados distintos, con probabilidades $p_1, \ldots, p_k$ y n repeticiones

$$\mathbf{X} \sim \mathcal{M}_k(p_1, p_2, \ldots, p_k, n)$$

$$P(X_1 = r_1, X_2 = r_2, \ldots, X_k = r_k) = \frac{n!}{r_1! r_2! \ldots r_k!} p_1^{r_1} \ldots p_k^{r_k}$$

$r_1 + r_2 + \ldots + r_k = n$

$E(X_i) = np_i,\ Var(X_i) = np_i(1 - p_i),\ Cov(X_i, X_j) = -np_i p_j$

Ejemplo 5

Según una encuesta, en cierta localidad la intención de voto para cuatro candidatos a intendente es la siguiente: Candidato A: 45 % Candidato B: 35 %, Candidato C: 15 % y Candidato D: 5%. Se hace una selección aleatoria de 10 personas de esa localidad. Suponiendo que esas personas no votan en blanco y sólo votan por un candidato, ¿Cuál es la probabilidad de que 1 vote al candidato A, 2 al B, 3 al C y 4 al D?

Ejemplo 6

Un científico inocula varios ratones, uno por vez, con el virus de una enfermedad hasta que uno de ellos la contraiga. Si la probabilidad que un ratón contraiga la enfermedad es $1/6$, cuál es la probabilidad que requiera 7 ratones?

- ¿Es el modelo binomial adecuado para este experimento?
- ¿Por qué?

Distribución Geométrica

Consideremos un experimento aleatorio tal que:

- El experimento consiste de una serie de repeticiones independientes entre sí con sólo dos posibles resultados en cada prueba: Exito (E) y Fracaso(F).
- $P(E) = p$ (constante) en cada prueba.

La v.a. X definida como $X=$ número de pruebas necesarias hasta observar el primer éxito tiene distribución Geométrica de parámetro p.

- Su función de probabilidad de masa está dada por:

$$p(x) = P(X = x) = (1-p)^{x-1}p, \quad x = 1, 2, ...$$

- $\mu = E(X) = 1/p$
- $Var(X) = (1-p)/p^2$

Continuación ejemplo 6

Supongamos ahora que el científico inocula el virus de la enfermedad en ratones, uno a la vez, hasta que **tres** ratones hayan contraído esa enfermedad.

Si la probabilidad de contraer la enfermedad es $1/6$,

Pregunta

Sigue siendo válido el modelo Geométrico?

Distribución Binomial Negativa (Distribución de Pascal)

- Un experimento aleatorio tiene dos posibles resultados: éxito (E) con probabilidad p y fracaso (F) con probabilidad $1-p$.
- Se hacen repeticiones independientes del experimento hasta que ocurran r éxitos.

Definición

Sea X la variable aleatoria definida como el número de experimentos necesarios hasta obtener r éxitos.
Se dice que X tiene distribución **Binomial Negativa** de parámetros p y r, y se denota por:

$$X \sim BN(p, r)$$

$$p(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r, \quad x = r, r+1, r+2, \ldots$$

$$\mu = E(X) = \frac{r}{p}, \qquad \sigma_X^2 = r\frac{1-p}{p^2}$$

Continuación ejemplo 6

Un científico inocula varios ratones, uno a la vez, con el virus de una enfermedad hasta obtener tres ratones que hayan contraído esa enfermedad.

Pregunta

Â¿Cuál es la probabilidad de que se requieran 8 ratones, si la probabilidad de contraer la enfermedad es $1/6$?

- $p = \frac{1}{6}$
- $r = 3$

$$P(X = 8) = \binom{8-1}{3-1} \left(1 - \tfrac{1}{6}\right)^{8-3} \left(\tfrac{1}{6}\right)^{3}$$

Valores posibles de X

$$X = 3, 4, 5, \ldots$$

Ejemplo 7

En un vivero hay 100 plantines, de los cuales 20 están atacados por cierta plaga. Una persona elige al azar 10 plantines y los compra. ¿Cuál es la probabilidad que haya comprado 10 plantines que no estén infectados? ¿ Y que entre esos 10 haya 3 plantines infectados?

¿Puede usarse alguno de los modelos vistos hasta ahora?

Distribución Hipergeométrica

Supuestos:

1. Los N elementos de una población pueden clasificarse en dos categorías excluyentes "Exito" (E) y "Fracaso" (F).
2. n_1 el número de elementos que pueden clasificarse como E.
3. n_2 el número de elementos que pueden clasificarse como F. $n_1 + n_2 = N$
4. Se extrae una muestra aleatoria de tamaño n, sin reposición, de dicha población $(n \leq N)$
5. Sea $X =$ número de elementos de la categoría E presentes en dicha muestra.

X así definida es una v.a. con distribución Hipergeométrica de parámetros N, n_1, n y se denota $X \sim h(N, n_1, n)$

Observación: Como las extracciones se hacen sin reposición, la probabilidad de obtener un elemento de la categoría E cambia de una prueba a otra.

Distribución Hipergeométrica

Si $X \sim h(N, n_1, n)$ entonces

$$P(X = r) = \frac{\begin{pmatrix} n_1 \\ r \end{pmatrix} \begin{pmatrix} N - n_1 \\ n - r \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}} \tag{1}$$

Con $r \in \mathbb{N}_o$ sujeto a: a) $r \leq n_1$ y b) $(n - r) \leq (N - n_1)$

Además

$E(X) = n \cdot \frac{n_1}{N}$, $Var(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{n_1}{N} \cdot (1 - n_1/N)$

Ejemplo 7

En un vivero hay 100 plantines, de los cuales 20 están atacados por cierta plaga. Una persona elige al azar 10 plantines y los compra. ¿Cuál es la probabilidad que haya comprado 10 plantines que no estén infectados? ¿ Y que entre esos 10 haya 3 plantines infectados?

Relación entre modelos hipergeométrico y binomial

Sea $X \sim H(N, n, N_1)$.

Si el número de éxitos (N_1) y fracasos $(N_2 = N - N_1)$ en la población es grande en comparación con el número n de observaciones y además $n/N < 0.1 \Rightarrow$ la diferencia entre que se tomen las observaciones <u>sin</u> reemplazo (modelo hipergeométrico) o <u>con</u> reemplazo (modelo binomial) será insignificante.

En tal caso puede aproximarse la distribución hipergeométrica por una binomial y por ello suele decirse que

El modelo binomial es el límite del modelo hipergeométrico cuando N tiende a infinito

Ejemplo 8

En un vivero hay 1000 plantines, de los cuales 400 están atacados por cierta plaga. Una persona elige al azar 10 plantines y los compra. ¿Cuál es la probabilidad que haya comprado 10 plantines que no estén infectados? ¿ Y que entre esos 10 haya 3 plantines infectados?

Distribución de Poisson

Existen eventos que ocurren en puntos aleatorios de tiempo, espacio, volumen, superficie, etc:

- Número de bacterias en un determinado cultivo.
- Número de partículas emitidas por una fuente radioactiva en un intervalo de tiempo.
- Número de clientes que usan un cajero automático en un intervalo de tiempo.
- Número de accidentes automovilísticos que ocurren en la ciudad de Corrientes por semana.
- Número total de llamadas telefónicas que llegan a una central telefónica entre las 10 y las 12 hrs de los días hábiles.

Este tipo de eventos puede modelarse utilizando el modelo de Poisson

Distribución de Poisson

Supuestos:

- El número de eventos que ocurre en un intervalo de tiempo (área, volumen, etc) es independiente del número de eventos que ocurre en cualquier otro intervalo disjunto de tiempo (área, volumen, etc)
- La probabilidad de n ocurrencias del evento en intervalos de igual longitud (área, volumen) es la misma (probabilidad depende de la amplitud del intervalo, no del origen)
- μ_P: número promedio de ocurrencias del evento por unidad de tiempo (espacio/volumen/etc.).
- λ: la tasa media de ocurrencia del evento en un intervalo de tiempo, (espacio, volumen, etc.) Δ_t, es constante. $\lambda = \mu_p \Delta_t$

Distribución de Poisson

Si X es la v.a que cuenta el número de eventos que ocurre en un intervalo de tiempo (espacio, volumen) $\Rightarrow X$ tiene distribución de Poisson de parámetro λ

$$P(X = r) = \frac{\lambda^r}{r!} e^{-\lambda}, r = 0, 1, 2, \ldots$$

$$E(X) = \lambda, \quad Var(X) = \lambda$$

Ejemplo 9

En promedio, 25 personas por hora utilizan el cajero automático que está en la entrada del campus.

1. ¿Cuál es la probabilidad que entre las 15 y las 16 concurran a lo sumo 10 personas al cajero?
2. ¿Cuál es la probabilidad que entre las 15 y las 17 concurran exactamente 20 personas?

Relación entre modelos Binomial y Poisson

Supongamos $X \sim \mathcal{B}(n, p)$.

Si n es grande ($n \to \infty$) y p pequeño ($p \to 0$), entonces
$\left(\begin{array}{c} n \\ x \end{array} \right) p^x (1-p)^{n-x} \cong e^{-np} \frac{(np)^r}{r!}$

Regla práctica para aproximar binomial por Poisson:

1. $n > 100$, $p < 0.05$ y $np \leq 20$.
2. Dada p pequeña, cuanto más grande sea n mejor la aproximación
3. Dada n, cuanto más pequeña p mejor será la aproximación

Ejemplo 10

En una editorial se hace un gran esfuerzo en la corrección de los textos para asegurar que los libros que edita estén libres de errores tipográficos. La probabilidad de encontrar al menos un error de tipeo en una página es 0.005, y los errores son independientes página a página. Supongamos que un libro tiene 400 páginas, ¿cuál es la probabilidad de encontrar dos páginas con errores?

¿Qué modelo de distribución sería conveniente para modelar a la variable aleatoria X definida en cada una de estas situaciones?

1. Un lote de 10 celulares contiene 3 con problemas de cámara. Se extraen 4 celulares de ese lote y se define $X =' $ número de celulares con problemas en la cámara.
2. El recuento de glóbulos blancos de un individuo sano puede presentar un promedio de hasta 6000 por milímetro cúbico de sangre. Para detectar una deficiencia de glóbulos blancos se determina su número en una gota de sangre de 0.001 milímetros cúbicos. Sea $X=$ número de glóbulos blancos en 0.001 mm^3 de sangre
3. Un producto electrónico contiene 40 circuitos integrados, que funcionan independientemente uno de otro. La probabilidad que alguno de ellos falle es 0.01. Sea $X =$ número de circuitos integrados defectuosos.
4. El 30% de los habitantes de una ciudad numerosa tienen grupo sanguíneo A^+. Se define $X =$ número de donantes (en esa ciudad) que ingresan a un hospital hasta conseguir el primer donante con grupo A^+.