Unidad 4: Modelos de probabilidad discretos.

4.0 Contenidos y objetivos.

Contenidos:

Objetivos:

- *Aplicar correctamente los modelos de probabilidad en el cálculo de probabilidades en problemas de aplicación.*
- *Calcular percentiles.*
- *Reconocer la conveniencia de la aplicación de distribuciones límite cuando el tamaño n de la muestra es muy grande.*

Palabras clave:

Variable aleatoria.
Función de cuantía.
Función de distribución Acumulativa.
Percentil.
Independencia de Variables Aleatorias.
Esperanza.
Varianza.

4.1. Distribución de Bernoulli y distribución Binomial.

Palabras Clave: Distribución de probabilidad, función de cuantía, función de distribución acumulativa, esperanza matemática y varianza de una variable aleatoria.

Resumen de conceptos y propiedades.

- ✓ Una variable aleatoria X tiene distribución de Bernoulli de parámetro $p \in (0,1)$ si estudia la probabilidad de ocurrencia de un suceso E tal que $P(E) = p$, o de su contrario con $P(E^c) = 1 - p = q$, cuando una experiencia se realiza una vez. Si 1 designa la ocurrencia de E y 0 designa su no ocurrencia, entonces $P(X = 1) = P(E) = p$ y, en consecuencia, $P(X = 0) = 1 - P(E) = 1 - p$. Su valor esperado es $E(X) = p$ y su varianza está dada por $V(X) = p(1 - p) = pq$.

- ✓ Una variable aleatoria X tiene **distribución Binomial de parámetros** n y p si corresponde a una reiteración de la experiencia asociada a una variable Bernoulli cierta cantidad de n de veces y bajo condiciones que permiten identificar sus condiciones más importantes:

  C1. **Dicotomía**: Estudio de un suceso E y de su contrario, E^C, en un experimento que se efectúa n veces.

  C2. **Probabilidades constantes** de los sucesos en estudio: $P(E) = p$, $P(E^c) = 1 - p = q$ cada vez que se efectúa el experimento.

  C3. El experimento se realiza un <u>número fijo n</u> de veces.

  C4. Existe i**ndependencia** entre resultados de cada repetición del experimento ($n > 1$).

  C5. **Variable aleatoria**: queda definida como X : "número de veces en que ocurre E (el suceso de interés) en las n realizaciones del experimento".

- ✓ También se define a la variable aleatoria (v.a.) X con distribución Binomial de parámetros n y p como aquella cuya **función de cuantía** está dada por:

$$f_X(x) = \begin{cases} P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} & , \; x = 0,1,2,\ldots,n \\ 0 & , \; \textit{en otro caso.} \end{cases}$$

- Si la v.a. X tiene distribución binomial de parámetros n y p, se escribirá: $X \sim Binomial(n; p)$ o simplemente $X \sim B(n; p)$.

- Si $n = 1$ la distribución binomial es llamada distribución de Bernoulli.

- La función de distribución acumulativa (FDA) de probabilidades de X es:

$$F_X(a) = P[X \leq a] = \sum_{\substack{\forall x \leq a \\ x=0,1,\ldots,n}} f_X(x) = \sum_{\substack{\forall x \leq a \\ x=0,1,\ldots,n}} \binom{n}{x} p^x (1-p)^{n-x}, \; a \in \mathbb{R}.$$

- Si $\text{Rec}(X) = \{x_1, x_2, \ldots, x_{j-1}, x_j, \ldots, x_n\}$, donde x_i designa el resultado de la experiencia en la i-ésima repetición, entonces

$$P(X = x_{j+1}) = F_X(x_{j+1}) - F_X(x_j), \; j \in \mathbb{N}.$$

- **Si** $X \sim Binomial(n; p)$, entonces:

  1. El **Valor esperado** de X, **esperanza matemática** de X, **"media"** de X o, simplemente, **esperanza de** X es (sus unidades son las de la variable X):

  $$\boxed{E(X) = \mu_X = np}.$$

  2. La **Varianza** de X es (sus unidades son cuadrados de las unidades de la variable X):

  $$\boxed{V(X) = \sigma_X^2 = npq}.$$

  3. La **Desviación Estándar** de X es (sus unidades son las de la variable X):

  $$\boxed{\sigma_X = \sqrt{V(X)} = \sqrt{npq}}$$

4.1.1 Ejercicios resueltos, paso a paso:

Ejemplo 1.

(**Aplicación en Ciencias de la Educación**)

La última libro de un autor científico ha tenido un gran éxito, hasta el punto de que el 80% de los lectores ya la han leído. Un grupo de 4 amigos son aficionados a la lectura:

a. ¿Cuál es la probabilidad de que en el grupo de estos 4 amigos hayan leído la novela 2 personas?

b. ¿Y como máximo 2?

c. ¿Cuál es el número esperado de personas que han leído el libro en este grupo de cuatro?¿Cuál es su desviación estándar?

Esquema de solución:

Paso 1. Identificar los datos del problema, condiciones y la variable en estudio:

- ✓ **Suceso de interés**: Referido a la última novela de un autor,

  L: La persona **ha leído** la novela.

- ✓ **Probabilidad del suceso**:

  $P(L) = 0.80$ ("el 80% de las personas han leído la novela").

- ✓ **Experiencia base**: Se elige al azar una persona para registrar si ha leído o no la última novela del autor. El espacio muestral asociado es

$$\Omega_1 = \{l, l^c\}$$

  donde l significa "la persona **ha leído** la novela" y l^c significa "la persona **no ha leído** la novela"

- ✓ **Número de repeticiones de la experiencia base**: $n = 4$ (se eligen cuatro amigos para averiguar si han leído o no la novela). El espacio de muestras para este caso es

$$\Omega_4 = \left\{ (x_1, x_2, x_3, x_4) / \; x_i = \begin{cases} l \\ l^c \end{cases}, \; i = 1,2,3,4 \right\}$$

$$\Omega_4 = \Omega_4 \times \Omega_4 \times \Omega_4 \times \Omega_4 \,.$$

Se pide: "…la probabilidad de que en el grupo de estos 4 amigos hayan leído la novela 2 personas". Esto sugiere definir una variable contadora

X : Número de personas entre las 4 seleccionadas, que ha leído la novela

Con esto se tiene la las condiciones para comenzar a pensar en que X tiene distribución binomial, como se verá en los pasos siguientes.

Paso 2. Identificar el suceso en estudio y su suceso contrario:

E : $\{x / x \textit{ ha leído la novela}\}$, conjunto de personas que han leído la novela.

E^C: $\{x / x \text{ no ha leído la novela}\}$ conjunto de personas que no han leído la novela.

Paso 3. Identificar sus probabilidades (constantes):

P(E) = 0.80 (probabilidad de que una persona haya leído la novela).

$P(E^C)$ = 0.20 (probabilidad de que una persona no haya leído la novela).

Paso 4. Número fijo de veces en que se repite el experimento: n = 4.

Paso 5. Independencia: Para que exista independencia es necesario que las personas consultadas no se hayan ejercido influencias mutuamente en el sentido de leer o no la novela. Desde este punto de vista los amigos deben haber decidido en forma independiente la lectura o no del texto en cuestión como para garantizar el hecho de que una persona haya leído la novela no cambia la probabilidad de que la siguiente elegida la haya leído o no (0,80 o 0,20, respectivamente). Por lo tanto, la solución procede si hemos garantizado esta **independencia** entre resultados de repeticiones del experimento.

Paso 6. Identificación de la variable aleatoria en la concepción binomial:

X: Número de personas que han leído la novela entre las n = 4 elegidas.

Paso 7. Deducción de la distribución: Por los pasos 2 a 6, $X \sim Binomial(n=4; p=0.80)$.

Paso 8. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar.

a. $$P[X=2] = \binom{4}{2}(0.80)^2(0.20)^2 = 0.1536,$$

Resp.: La probabilidad de que entre las cuatro personas que forman el grupo de amigos haya 2 que han leído la novela es 0.1536 (15.36%).

b. $$P[X \leq 2] = 1 - P[X > 2] = 1 - \left(P[X=3] + P[X=4]\right)$$

$$P[X \leq 2] = 1 - \binom{4}{3}(0.80)^3(0.20) - \binom{4}{4}(0.80)^4(0.20)^0 = 0.1808$$

Resp.: La probabilidad de que como máximo 2 personas de las 4 seleccionadas, haya leído la novela es de 0.1808 (18.08%).

c. $$X \sim Binomial(4; 0.80) \quad \Rightarrow \quad E(X) = np = (4)(0.80) = 0.32$$

$$V(X) = 0.64$$

$$\sigma_X = \sqrt{V(X)} = 0.8$$

Resp.: El número esperado de personas que han leído el libro es de 0.32. La desviación estándar vale 0.8

Ejemplo 2.

(**Aplicación en Ciencias de la Ingeniería**)

Las líneas de teléfono de un sistema de reserva de vuelos están ocupadas el 40% del tiempo. Supongamos que los eventos en que las líneas estén ocupadas en las sucesivas llamadas son independientes. Si se realizan 10 llamadas a la compañía aérea:

a. ¿Cuál es la probabilidad de que en exactamente tres llamadas las líneas estén ocupadas?

b. ¿Cuál es la probabilidad de que en por lo menos una llamada las líneas no están ocupadas?

c. ¿Cuál es el número esperado de llamadas en las que las líneas están ocupadas?¿y la desviación estándar?

Esquema de solución

Paso 1. Identificar la experiencia: Verificar si la línea está ocupada cuando se efectúa una llamada de reserva de vuelo.

Paso 2. Identificar el suceso en estudio y su suceso contrario:

E : La línea está ocupada cuando se efectúa una llamada de reserva de vuelo.

E^C: La línea no está ocupada.

Paso 3. Identificar sus probabilidades (constantes):

P(E) = 0.40 (probabilidad de que la línea esté ocupada).

P(E^C) = 0.60 (probabilidad de que la línea no esté ocupada).

Paso 4. Número fijo de veces en que se repite la experiencia: n = 10.

Paso 5. Independencia: El hecho de que en un llamado la línea esté ocupada no cambia la probabilidad de que en el siguiente llamado la línea esté (o no) ocupada. Por lo tanto, existe independencia entre resultados de llamadas sucesivas.

Paso 6. Identificación de la variable aleatoria en la concepción binomial:

X : Número de llamadas en que la línea está ocupada entre las n = 10 llamadas realizadas.

Paso 7. Deducción de la distribución: $X \sim Binomial(n=10; p=0.40)$.

Paso 8. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar.

a. $P[X=3]=\binom{10}{3}(0.40)^3(0.60)^7=0.214990848\approx 0.2150$,

Resp.: La probabilidad de que entre 10 llamadas haya 3 en las que la línea esté ocupada es 0.2150 (21.50 %).

b. $P[X \leq 9] = 1 - P[X = 10] = 1 - \binom{10}{10}(0.40)^{10}(0.60)^{0}$

$$= 0.9998951424 \approx 1$$

Resp.: La probabilidad de que, en 10 llamadas, al menos en una de ellas la línea no esté ocupada es aproximadamente 1.

Este cálculo también pudo haber sido realizado en base a la definición de una variable nueva Y que designe el número de llamadas en que la línea **no está ocupada**. En este caso, si la probabilidad de que la línea esté ocupada cuando se realiza una llamada es de 0.40 entonces la probabilidad de que **no esté ocupada** es de 0.60. En base a la probabilidad de este evento se puede demostrar que Y tiene una distribución binomial de parámetros $n = 10$ y $p = 0.6$. Por lo tanto, "que, en 10 llamadas, al menos en 1 la línea no esté ocupada", significa calcular $P(Y \geq 1)$. Así,

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{10}{0}(0.6)^{0}(0.4)^{10} = 0.9998951424 \approx 1$$

c. $E[X] = \mu_X = np = (10)(0.40) = 4$ llamadas;

Resp.: El valor esperado de llamadas en que la línea está ocupada, entre 10 realizadas, es de 4 llamadas.

$V[X] = \sigma_X^2 = npq = (10)(0.4)(0.6) = 2.4 \Rightarrow \sigma_X \approx 1.55$ llamadas.

Resp.: La desviación estándar de llamadas que personas que han leído la novela entre los cuatro amigos es de 0.8 personas (no alcanza a ser 1).

4.1.2 Ejercicios Propuestos.

1. Para cada escenario descrito a continuación, indicar si la distribución binomial es o no (indicando por qué) un modelo razonable para la variable aleatoria mencionada. Explicite cualquier suposición que haga:

   a. Un proceso de producción produce miles de transductores de temperatura. Sea X el número de transductores disconformes en una muestra de tamaño 30 seleccionada al azar en el proceso.

   b. De un lote de 50 transductores de temperatura, una muestra de tamaño 30 es seleccionada sin reemplazo. Sea X el número de transductores disconformes en la muestra.

   c. Cuatro componentes electrónicos idénticos son conectados a un controlador que puede hacer un cambio desde un componente que ha fallado, hacia una de las piezas de repuesto restantes. Sea X el número de componentes que han fallado después de un período determinado de operación.

d. Sea X el número de accidentes que ocurren a lo largo de las carreteras federales en Arizona durante un período de un mes.

e. Sea X el número de respuestas correctas de un estudiante que toma un examen de opción múltiple en el que puede eliminar algunas o todas las opciones que ha respondido incorrectamente.

f. Los defectos se producen al azar sobre la superficie de un chip semiconductor. Sin embargo, sólo el 80% de los defectos pueden ser registrados por las pruebas. Se analiza una muestra de 40 fichas de chips con un defecto de cada uno. Sea X el número de fichas en las que el examen encuentra un defecto.

g. En el inciso anterior considere el caso en que ahora la muestra se compone de 40 chips con 1 o con 0 defectos.

h. En una operación de llenado de paquetes de detergente con el peso declarado, sea X el número de envases de detergente que son llenados con exceso.

i. Los errores en un canal de comunicación digital se producen en situaciones que afectan a varios bits consecutivos. Sea X el número de bits afectados en la transmisión de 100.000 bits.

2. La variable aleatoria X tiene una distribución binomial con $n=10$ y $p=0,5$. Dibuje la función de cuantía de probabilidad de X.

   a. ¿Qué valor de X es más probable? Resp.: $P(X=0)=0,59874$ aprox.
   b. ¿Qué valor (s) de X es (son) menos probable? Desde 6 con probabilidad $P(X=6)=2,6726\cdot 10^{-6}$ hasta 10 con probabilidad $P(X=10)=9,76563\cdot 10^{-14}$.

3. Determine la función de distribución acumulativa de una variable aleatoria con distribución binomial con $n=3$ and $p=0,1$, $p=0,2$.

   a. Resp.: $F_X(x)=\begin{cases} 0 & , x<0. \\ 0.729, & 0\le x<1. \\ 0.972, & 1\le x<2. \\ 0.999, & 2\le x<3. \\ 1 & , x\ge 3. \end{cases}$ para $p=0,1$;

$$F_X(x)=\begin{cases} 0 & , x<0. \\ 0.512, & 0\le x<1. \\ 0.896, & 1\le x<2. \\ 0.992, & 2\le x<3. \\ 1 & , x\ge 3. \end{cases} \text{ para } p=0,2$$

4. (Aplicación en Ciencias de la Ingeniería) Sea X el número de bits recibidos con error en un canal de comunicación digital. Suponga que X es una variable aleatoria binomial con $p = 0,001$. Si se transmiten 1000 bits, determine lo siguiente:

   a. $P[X = 1]$. Resp.: 0.3681. c. $P[X \leq 2]$. Resp. 0.9198.
   b. $P[X \geq 1]$. Resp.: 0.6323. d. Media y varianza de X. Resp.: $\mu_X = 1$; $\sigma_X^2 = 0.999$

5. (Aplicación en Ciencias de la Ingeniería) En lotes de 50 resortes en espiral de un proceso de producción se verifica su conformidad con los requisitos del cliente. El número medio de resortes no conforme en un lote es de 5 y que el número de resortes no conforme en un lote, denotado por X, es una variable aleatoria binomial.

   a. ¿Qué valores tienen n y p? Resp.: $n = 50$.
   b. ¿Cuál es la probabilidad de que en tres lotes se encuentren menos de 5 resortes disconformes? Resp.: $p = 0.1$.

6. (Aplicación en Ciencias de la Ingeniería) Un proceso de fabricación cuenta con 100 pedidos de clientes. Cada orden requiere un componente que se compra a un proveedor y un 2% de éstos ha resultado defectuoso. Se puede suponer que los componentes son independientes.

   a. Si el fabricante dispone de un stock de 100 componentes, ¿cuál es la probabilidad de que las 100 órdenes puedan llenarse sin necesitar comprar nuevas componentes?
   b. Si el fabricante cuenta con 102 componentes, ¿cuál es la probabilidad de que las 100 órdenes pueden llenarse sin necesitar comprar nuevas componentes?
   c. Si el fabricante cuenta con 105 componentes, ¿cuál es la probabilidad de que las 100 órdenes pueden llenarse sin necesitar comprar nuevas componentes?

7. (Aplicación en Ciencias de la Ingeniería) Ejemplo de control estadístico de proceso gráfico. Cada hora se seleccionan muestras de 20 partes de un proceso de troquelado de metal (el proceso consiste en cortar con precisión de medidas con un troquel). Normalmente, el 1% de las piezas requieren de revisión. Sea X el número de piezas en la muestra de 20 que requieren revisión. El problema del proceso es que se sospecha X supera su media en más de tres desviaciones estándar.

   a. Si el porcentaje de piezas que requieren rehacerse mantiene en el 1%, ¿cuál es la probabilidad de que X supere su media de más de tres desviaciones estándar?
   b. Si el porcentaje que debe rehacerse aumenta al 4%, ¿cuál es la probabilidad de que X sea superior a 1?
   c. Si el porcentaje que debe rehacerse aumenta al 4%, ¿cuál es la probabilidad de que X sea superior a 1 en al menos uno de los próximos cinco horas de muestras?

8. (Aplicación en Ciencias de la Educación) Un test de elección múltiple consta de 25 preguntas, cada una con cuatro respuestas. Supongamos que un estudiante sólo hace conjeturas sobre cada pregunta cuando va a responder.

   a. ¿Cuál es la probabilidad de que el estudiante responda correctamente más de 20 preguntas? Resp.: 0.
   b. ¿Cuál es la probabilidad de que el estudiante responda correctamente menos de 5 preguntas? Resp.: 0.2137.

9. (Aplicación en Ciencias de la Ingeniería) Un semáforo en particular a lo largo de su viaje matutino está de color verde el 20% de las veces en que usted se acerca a él. Suponga que cada mañana representa una situación independiente de las de otros días.

a. En la mañana de cinco días, ¿cuál es la probabilidad de que la luz esté en verde en exactamente en un día?
b. En la mañana de veinte días, ¿cuál es la probabilidad de que la luz esté en verde en exactamente cuatro días?
c. En la mañana de veinte días, ¿cuál es la probabilidad de que la luz es verde en más de cuatro días?

10. (Aplicación en Ciencias Sociales) En los robos que han habido en una ciudad determinada, el 75% ha sido causado por la necesidad de dinero para comprar alimentos. Si continúa esta tendencia, encuentre la probabilidad de que entre los próximos 5 casos de robo en esta ciudad

    a. exactamente en 2 el robo haya sido generado por la necesidad de dinero para comprar alimentos. Resp.: 0.0879.
    b. a lo más en 3 robos de los próximos 5 casos el móvil haya sido la necesidad de dinero para comprar alimentos. Resp.: 0.3672.

11. (Aplicación en Ciencias de la Ingeniería) Según una investigación del año 1990, aproximadamente el 30% de las fallas de las tuberías en plantas químicas son causadas por error del operador.

    a. ¿Cuál es la probabilidad de que en los próximos 20 fallos de tubería por lo menos 10 se deban a error del operador? Resp.: $P(X \geq 10) = 1 - P(X \leq 9) = 0.0480$.
    b. ¿Cuál es la probabilidad de que no más de 4 de cada 20 fallas se deban a errores del operador? Resp.: $P(X \leq 4) = 0.2375$.
    c. Supongamos, para una planta en particular, que al considerar una muestra al azar de 20 fallas de tubería, exactamente 5 se deban a error del operador. ¿Cree usted que la cifra 30% que se ha dicho es aplicable a esta planta? Comente. Resp.: $P(X = 5) = 0.1789$. Puesto que esta probabilidad puede no ser considerada pequeña, el valor p = 30% puede ser razonable.

12. (Aplicación en Ciencias de la Ingeniería) En la prueba de un cierto tipo de neumático de camión en terreno montañoso, se encontró que el 25% de los camiones no llegan a completar la prueba por causa de un reventón.

    a. De los 15 próximos camiones probados encuentre la probabilidad de que
        i. de 3 a 6 tengan un reventón. Resp.: 0.7073.
        ii. menos de 4 tengan un reventón. Resp.: 0.4613.
        iii. más de 5 tengan un reventón. Resp.: 0.1484.
    b. ¿Cuántos de los 15 camiones espera usted que tengan reventones?
    Resp.: $E(X) = \mu_X = np = (15)(0.25) = 3.75$, es decir, entre 3 y 4.

13. (Aplicación en Ciencias de la Ingeniería) Si la probabilidad de que una luz fluorescente para el árbol de navidad tiene una vida útil de al menos 800 horas es de 0,9, encuentre las probabilidades de que entre 20 de estas luces:

    a. exactamente 18 tengan una vida útil de al menos 800 horas. Resp.: 0.2852.
    b. al menos 15 tengan una vida útil de al menos 800 horas. Resp.: 0.9888.
    c. al menos 2 no tengan una vida útil de al menos 800 horas. Resp.: 0.6083.

14. (Aplicación en Ciencias de la Salud) Un prominente médico afirma que el 70% de los pacientes con cáncer de pulmón son fumadores empedernidos. Si su afirmación es correcta, determinar

a. la probabilidad de que de 10 pacientes recientemente ingresados en un hospital, menos de la mitad son fumadores empedernidos. Resp.: 0.0474.
b. la probabilidad de que de 20 pacientes recientemente ingresados en un hospital, menos de la mitad son fumadores empedernidos. Resp.: 0.0171.

15. (Aplicación en Ciencias de la Salud) La probabilidad de que un paciente se recupere de una delicada operación de corazón es de 0,9. ¿Cuál es la probabilidad de que sobrevivan exactamente 5 de los próximos 7 pacientes con esta operación? Resp.: 0.1240.

16. (Aplicación en Ciencias de la Salud) Un estudio examinó las actitudes nacionales sobre los antidepresivos. Este estudio reveló que aproximadamente el 70% cree que "los antidepresivos en realidad nada curan, sólo encubren el verdadero problema". Según a este estudio, ¿cuál es la probabilidad de que al menos 3 de las próximas 5 personas seleccionadas al azar tengan esta opinión? Resp.: 0.8369.

4.2. Distribución de Poisson.

Palabras Clave: Distribución de probabilidad, función de cuantía, función de distribución acumulativa, esperanza matemática y varianza de una variable aleatoria.

Resumen de conceptos y propiedades.

✓ Definición de una v.a. X con distribución de Poisson de parámetro $\lambda > 0,\ \lambda \in \mathbb{R}$:

Dado un intervalo I de números reales, se efectúan en él observaciones de la ocurrencia al azar de un suceso determinado E, contando las veces que E ocurre en ese intervalo.
Si el intervalo I se puede particionar en subintervalos de longitud lo sufientemente pequeña de manera que

a. La probabilidad de que E ocurra más de una vez en un subintervalo es cero,
b. La probabilidad de que E ocurra en un subintervalo es la misma en todos los subintervalos y proporcional a la longitud del subintervalo, y
c. La ocurrencia (o no) de E en cada subintervalo es independiente de otros subintervalos,

entonces, el experimento al azar se llama un **Proceso de Poisson**.
La variable aleatoria X que cuenta el número de veces en que E ocurre en un intervalo de números reales es una **variable aleatoria de Poisson** con parámetro $\lambda > 0$.

✓ La función de cuantía (o de masa) de una v.a. X con distribución de Poisson de parámetro $\lambda > 0$, lo cual es denotado por $X \sim Poisson(\lambda)$, está dada por

$$f_X(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \text{ para } x = 0,1,2,...$$

- ✓ La función de distribución acumulativa (FDA) de probabilidades de X es:

$$F_X(a) = P[X \le a] = \sum_{\substack{\forall x \le a \\ x=0,1,...,n}} f_X(x) = \sum_{\substack{\forall x \le a \\ x=0,1,2...}} \frac{\lambda^x e^{-\lambda}}{x!},\ a \in \mathbb{R}.$$

- ✓ Si $\text{Rec}(X) = \{0,1,...\} = \mathbb{N}_0$, entonces

$$P(X = j) = F_X(j+1) - F_X(j),\ j \in \mathbb{N}_0.$$

- ✓ Valor esperado de $X \sim Poisson(\lambda)$: $E(X) = \mu_X = \lambda$.
- ✓ Varianza de $X \sim Poisson(\lambda)$: $V(X) = E(X^2) - [E(X)]^2 = \sigma_X^2 = \lambda$.

  Desviación Estándar: $\sigma_X = \sqrt{\lambda}$.

OBSERVACIÓN:

1. El parámetro λ es llamado **tasa promedio** de ocurrencia del suceso E en el intervalo "patrón" I. Sus unidades se expresan en un a proporción de veces en que ocurre E en el intervalo. Por ejemplo, "llamadas telefónicas por minuto", "accidentes por tramo de 50 kilómetros", "número de defectuosos poer cajas de 10 artículos", etc.
2. Hay una relación entre la distribución de Poisson y la distribución binomial. Se puede demostrar con cierta complejidad que si p disminuye mientras n aumenta, de manera que el producto np permanezca constante, es decir, de manera que se cumpla $\lambda = np = E(X)$, entonces

$$\lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} = \frac{\lambda^x e^{-\lambda}}{x!} \text{ para } x = 0,1,2,...$$

   Con esto se puede estudiar situaciones en que los valores de probabilidad se pueden lograr con cálculos de **aproximación** entre estas distribuciones, cuando n sea lo suficientemente "grande" y p lo suficientemente "pequeño".

4.2.1 Ejercicios resueltos, paso a paso:

Ejemplo 1.

(**Aplicación en Ciencias de la Ingeniería)**

El número de llamadas telefónicas que llegan a una central telefónica es a menudo modelado como una variable aleatoria de Poisson. Suponga que en promedio hay 10 llamadas por hora.

a. ¿Cuál es la probabilidad de que haya exactamente cinco llamadas en una hora?
b. ¿Cuál es la probabilidad de que haya a lo menos 3 llamadas en una hora?
c. ¿Cuál es la probabilidad de que haya exactamente 15 llamadas en dos horas?
d. ¿Cuál es la probabilidad de que haya exactamente cinco llamadas en 30 minutos?
e. ¿Cuál es el valor esperado de llamadas en un lapso de 2 horas? ¿y la desviación estándar?

Esquema de solución

Paso 1. Identificar el experimento: Estudiar el **número de llamadas** que se realizan a una central telefónica.

Paso 2. Identificar el intervalo "patrón" en el cual se va a observar lo que se desea estudiar:

- ✓ Intervalos de 1 hora.

Paso 3. Identificar la tasa promedio de ocurrencia de lo que se desea estudiar:

$$\lambda = \frac{10\ llamadas}{1\ hora} = 10\ \frac{llamadas}{hora}.$$

Paso 4. Existe <u>independencia</u> entre los resultados del "número de llamadas" entre intervalos de 1 hora.

Paso 5. Identificación de la variable aleatoria:

X : **Número de llamadas** que se efectúan a la central telefónica **durante 1 hora**.

Paso 6. Deducción de la distribución:

Se cumplen las condiciones que definen una variable aleatoria con distribución de Poisson de intensidad 10 $\frac{llamadas}{hora}$. Luego, $X \sim Poisson(\lambda = 10)$.

Paso 7. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar:

a. $P[X=5] = \frac{(10)^5 e^{-10}}{5!} = 0.0378$.

Resp.: La probabilidad de que se efectúen 5 llamadas en una hora es 0.0378 (3.78%).

b. $P[X \geq 3] = 1 - P[X<3] = 1 - \left(P[X=0] + P[X=1] + P[X=2]\right)$

$$= 1 - \frac{(10)^0 e^{-10}}{0!} - \frac{(10)^1 e^{-10}}{1!} - \frac{(10)^2 e^{-10}}{2!} = 0.9972$$

Resp.: La probabilidad de que, en una hora, haya a lo menos 3 llamadas es 0.9972 (99.72%).

c. $P[X_{2\ horas} = 15] =$

Para calcular el número de llamadas en intervalos de 2 horas se debe cambiar el intervalo "patrón" de 1 hora por uno de 2 horas. Esto se hace aprovechando la proporcionalidad definida en el proceso de Poisson:

$$\lambda = \frac{10\ llamadas}{1\ hora} = \frac{20\ llamadas}{2\ horas} = 20\ \frac{llamadas}{2\ horas}$$

Luego este nuevo proceso tiene una tasa promedio de 20 llamadas **cada dos horas**. Con este dato podemos ahora calcular:

$$P\left[X_{2\ horas} = 15\right] = \frac{(20)^{15} e^{-20}}{15!} = 0.0516$$

Resp.: La probabilidad de que, en 2 horas, haya 15 llamadas es 0.0516 (5.16%).

d. Procedemos de manera semejante al caso anterior y logramos ahora

$$\lambda = \frac{10\ llamadas}{1\ hora} = \frac{10\ llamadas}{60\ \text{minutos}} = 5\ \frac{llamadas}{30\ \text{minutos}}.$$

Así,

$$P\left[X_{30\ \text{minutos}} = 5\right] = \frac{(5)^{5} e^{-5}}{5!} = 0.1755$$

Resp.: La probabilidad de que, en 30 minutos, haya 5 llamadas es 0.1755 (17.55%).

e. $E\left[X_{2\ horas}\right] = \lambda_{2\ horas} = 20$ llamadas.

Resp.: El promedio esperado de llamadas cada dos horas es de 20 llamadas.

$V\left[X_{2\ horas}\right] = \sigma^2_{X_{2\ horas}} = \lambda_{2\ horas} = 20 \Rightarrow \sigma_{X_{2\ horas}} = \sqrt{20} = 4.47$ llamadas.

Resp.: La variabilidad de llamadas expresada en términos de desviación estándar de es 4.47 llamadas cada dos horas (entre 4 y 5).

Ejemplo 2.

(**Aplicación en Ciencias Medioambientales)**

Cierto área geográfica del planeta es, en promedio, golpeado por seis huracanes por año. Encontrar la probabilidad de que en un año dado la zona se vea afectada por

a. menos de cuatro huracanes;
b. entre 6 y 8 huracanes.

Esquema de solución

Paso 1. Identificar el experimento:

- ✓ Estudiar el **número de huracanes** que se producen **en cierta zona geográfica del planeta**.

Paso 2. Identificar el intervalo "patrón" en el cual se va a observar lo que se desea estudiar:

- ✓ Intervalos de 1 año.

Paso 3. Paso 3. Identificar la tasa promedio de ocurrencia de lo que se desea estudiar:

$$\lambda = \frac{6\ huracanes}{1\ año} = 6\ \frac{huracanes}{año}.$$

Paso 4. Existe <u>independencia</u> entre los resultados del "número de huracanes" entre intervalos de 1 año.

Paso 5. Identificación de la variable aleatoria:

X: Número de huracanes que se producen en cierta zona geográfica del planeta durante 1 año.

Paso 6. Deducción de la distribución:

Se cumplen las condiciones que definen una variable aleatoria con distribución de Poisson de intensidad $6\ \frac{huracanes}{año}$. Luego, $X \sim Poisson(\lambda = 6)$.

Paso 7. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar:

Paso 1. $P[X < 4] = P[X \leq 3] = \sum_{i=0}^{3} \frac{(6)^i e^{-6}}{i!} = \frac{(6)^0 e^{-6}}{0!} + \frac{(6)^1 e^{-6}}{1!} + \frac{(6)^2 e^{-6}}{2!} + \frac{(6)^3 e^{-6}}{3!} = 0.1512$

Resp.: La probabilidad de que se produzcan menos de 4 huracanes en un año es de 0.1512 (15.12%).

Paso 2. $P[6 \leq X \leq 8] = P[5 < X \leq 8] = F_X(8) - F_X(5) =$

$$= P[X \leq 8] - P[X \leq 5] = 0.8473 - 0.4457 = 0.4016$$

Resp.: La probabilidad de que, en un año, se produzcan entre 6 y 8 huracanes es 0.8493 (84.93%). Bastante alta. Obsérvese que se trata de valores en torno al promedio anual ($\lambda = 6$).

Ejemplo 3.

(**Aplicación en Ciencias de la Salud)**

Un chef restaurante prepara una ensalada que contiene, en promedio, 5 verduras cada día. Encontrar la probabilidad de que la ensalada contenga más de 5 verduras

a. en un día determinado;
b. en 3 de los 4 días siguientes;

Esquema de solución

Paso 1. Identificar el experimento:

- ✓ Estudiar el **número de verduras** que tiene **la ensalada preparada por el chef del restaurante.**

Paso 2. Identificar el intervalo "patrón" en el cual se va a observar lo que se desea estudiar:

- ✓ Intervalos de 1 día.

Paso 3. Identificar la tasa promedio de ocurrencia de lo que se desea estudiar:

$\lambda = \frac{5\ verduras}{1\ día} = 5\ \frac{verduras}{día}$.

Paso 4. Existe <u>independencia</u> entre los resultados del "número de verduras" entre intervalos de 1 día.

Paso 5. Identificación de la variable aleatoria:

X : Número de verduras que tiene la ensalada preparada por el chef en 1 día.

Paso 6. Deducción de la distribución:

Se cumplen las condiciones que definen una variable aleatoria con distribución de Poisson de intensidad $5\ \frac{verduras}{día}$. Luego, $X \sim Poisson(\lambda = 5)$.

Paso 7. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar:

a. $P[X > 5] = 1 - P[X \leq 5] = \sum_{i=0}^{3} \frac{(5)^i e^{-5}}{i!} = 1 - \frac{(5)^0 e^{-5}}{0!} - \frac{(5)^1 e^{-5}}{1!} - \ldots - \frac{(5)^5 e^{-5}}{5!} = 0.3840$

Resp.: La probabilidad de que en un día cualquiera la ensalada tenga más de 5 verduras es de 0.3840 (38.40%).

b. Si $p = P(X > 5) = 0.3840$ es la probabilidad de que en un día cualquiera la ensalada tenga más de 5 verduras (suceso E), podemos anotar $P(E) = p = 0.3840$. Estudiar este suceso en $n = 4$ días en forma **independiente** y contar en cuántos de ellos E va a ocurrir nos conduce a definir una variable aleatoria Y tal que $Y \sim Binomial(4; p)$. Por lo tanto lo que se pide es

$$P[Y = 3] = \binom{4}{3} p^3 (1-p)^{4-3} = \binom{4}{3} (0.3840)^3 (0.6160)^{4-3} = 0.1396$$

Resp.: La probabilidad de que la ensalada contenga más de 5 verduras en 3 de los 4 días siguientes es 0.1396 (13.96%).

4.2.2 Ejercicios Propuestos.

1. (Teoría de Inventario) Un estudio de inventario en un almacén determina que, en promedio, la demanda de un artículo particular se hace 5 veces al día. ¿Cuál es la probabilidad de que en un día determinado este artículo sea solicitado
   a. más de 5 veces al día? Resp.: 0.3840.

   b. no en todos los días? Resp.: 0.0067.

2. (Ingeniería de Tránsito) En promedio se producen, en un cruce determinado, tres accidentes de tráfico al mes. ¿Cuál es la probabilidad de que en un mes determinado en esta intersección
   a. exactamente 5 accidentes se producen? Resp.: 0.1008.
   b. menos de 3 accidentes se producen? Resp.: 0.4232.
   c. por lo menos 2 accidentes se producen? Resp.: 0.8009.

3. (Control de Calidad) Una secretaria comete 2 errores por página, en promedio. ¿Cuál es la probabilidad de que en la siguiente página cometerá
   a. 4 o más errores? Resp.: 0.1429.
   b. 0 errores? Resp.: 0.1353.

4. (Ciencias del Medio Ambiente)El promedio de ratones de campo por acre en un campo de trigo de 5 acres se estima en 12. Encuentre la probabilidad de que menos de 7 ratones de campo se encuentran
   a. en un acre determinado. Resp.: 0.0907.
   b. en 2 de los próximos 3 acres inspeccionadas. Sugerencia: Distribución binomial (¿por qué?) Resp.: 0.2725.

5. (Nutrición y Dietética) Un chef restaurante prepara una ensalada que contiene, en promedio, 5 verduras. Encuentre la probabilidad de que la ensalada contenga más de 5 verduras
   i. en un día determinado. Resp.: 0.3840.
   ii. b. en 3 de los 4 días siguientes. Sugerencia: Distribución binomial (¿por qué?) Resp.: 0.1395.

6. (Control de Calidad) Un fabricante de automóviles está preocupado por una falla en el mecanismo de frenado de un modelo en particular. La culpa puede, en raras ocasiones, causar un accidente a gran velocidad. La distribución del número de automóviles por año en el que experimentará la falla es una variable aleatoria de Poisson con intensidad $\lambda = 5$ por año.
   a. ¿Cuál es la probabilidad de que más de 3 autos por año experimenten un accidente por la razón citada? Resp.: 0.7350.
   b. ¿Cuál es la probabilidad de que más de un vehículo por año experimente un accidente? Resp.: 0.9933.

7. (Teoría de Colas) Los cambios en los procedimientos de planificación de aeropuertos requieren considerables planificaciones. Las tasas de llegada de las aeronaves son factores importantes que deben tenerse en cuenta. Supongamos que las aeronaves pequeñas llegan a un aeropuerto determinado, de acuerdo con un proceso de Poisson, a razón de 6 por hora.
   a. ¿Cuál es la probabilidad de que exactamente cuatro pequeños aviones lleguen durante un período de 1 hora? Resp.: 0.1339.
   b. ¿Cuál es la probabilidad de que a lo más lleguen 4 durante un período de 1 hora? Resp.: 0.2851.
   c. Si definimos una jornada de trabajo de 12 horas, ¿cuál es la probabilidad de que al menos 75 aviones pequeños llegan en un día? 0.3773.

8. (Teoría de Colas) El número de clientes que llegan por hora en un centro de servicio para automóviles determinados se supone que siguen una distribución de Poisson con una media de $\lambda = 7 \frac{clientes}{hora}$.
   a. Calcular la probabilidad de que más de 10 clientes llegan en un periodo de 2 horas. Resp.: 0.8243.
   b. ¿Cuál es el número promedio de llegadas durante un periodo de 2 horas? Resp.: 14.

9. El servicio de llamadas que entran a un centro de mantenimiento sigue un proceso de Poisson y, en promedio, 2.7 llamadas entran por minuto. Encuentre la probabilidad de que
   a. no más de 4 llamadas entren en un momento cualquiera;
   b. menos de 2 llamadas se produzcan en un cualquiera;
   c. más de 10 llamadas lleguen en un período de 5 minutos.

10. El titular de una farmacia local sabe que, en promedio, 100 personas por hora pasan por la tienda.
    a. Determinar la probabilidad de que en un período dado de 3 minutos nadie entre a la tienda.
    b. Determine la probabilidad de que en un determinado período de 3 minutos más de 5 personas entren a la tienda.

4.3. Distribución Geométrica.

Palabras Clave: Distribución de probabilidad, función de cuantía, función de distribución acumulativa, esperanza matemática y varianza de una variable aleatoria.

Resumen de conceptos y propiedades.

- ✓ Definición de una v.a. X con distribución geométrica de parámetro p:
  En una serie de ensayos Bernoulli (ensayos independientes con probabilidad constante $p \in (0,1)$ de un suceso E), se define una variable X como "número de repeticiones de la experiencia de Bernoulli hasta la primera ocurrencia de E", entonces X es llamada **variable aleatoria con distribución geométrica de parámetro** p.

- ✓ La función de cuantía (o de masa) de una v.a. X con distribución distribución geométrica de parámetro p, lo cual es denotado por $X \sim Geométrica(p)$, está dada por

$$f_X(x) = (1-p)^{x-1} p \text{, para } x = 1, 2, \ldots$$

- ✓ La función de distribución acumulativa (FDA) de probabilidades de X es:

$$F_X(a) = P[X \le a] = \sum_{\substack{\forall x \le a \\ x=1,2,\ldots}} f_X(x) = \sum_{\substack{\forall x \le a \\ x=1,2\ldots}} (1-p)^{x-1} p,\ a \in \mathbb{R}.$$

- ✓ Si $\text{Rec}(X) = \{1, 2, \ldots\} = \mathbb{N}$, entonces

$$P(X = j+1) = F_X(j+1) - F_X(j),\ j \in \mathbb{N};\ F_X(1) = P(X = 1) = p$$

- ✓ Valor esperado de $X \sim Geométrica(p)$: $E(X) = \mu_X = \frac{1}{p}$.

✓ Varianza de $X \sim Geométrica(p)$: $V(X) = E(X^2) - \left[E(X)\right]^2 = \sigma_X^2 = \frac{1-p}{p^2}$.

Desviación Estándar: $\sigma_X = \sqrt{\frac{1-p}{p^2}}$.

4.3.1 Ejercicios resueltos, paso a paso:

Ejemplo 1.

(**Aplicación en Ciencias de la Ingeniería**)

La probabilidad de una alineación óptica con éxito en el montaje de un producto óptico, según especificaciones para su almacenamiento, es de 0.8. Suponga que los intentos de montaje son independientes.

a. ¿Cuál es la probabilidad de que la alineación con éxito se produzca por primera vez en el quinto ensayo?
b. ¿Cuál es la probabilidad de que la primera alineación con éxito requiera como mínimo cuatro intentos?
c. ¿Cuál es la probabilidad de que la primera alineación con éxito requiera a lo menos cuatro y a lo más 6 intentos?
d. ¿Cuántos intentos se espera realizar para lograr el primer éxito?
e. Calcular la desviación estándar del número de intentos hasta lograr éxito en la alineación.

Esquema de solución:

Paso 1. Identificar el experimento:

Estudiar el número de intentos hasta lograr realizar con éxito la alineación de un producto óptico en su montaje.

Paso 2. Identificar el suceso E que se desea estudiar y que determina el éxito en la experiencia, y su contrario, E^C:

E : Se realiza con éxito la alineación de un producto óptico en el montaje.
E^C : No se realiza con éxito la alineación de un producto óptico en el montaje.

Paso 3. Identificar las probabilidades de E y E^C:

$$P(E) = p = 0.8 \qquad P(E^C) = p = 0.2.$$

Paso 4. Existe <u>independencia</u> entre los intentos de "alineación del producto óptico en el montaje" .

Paso 5. Identificación de la variable aleatoria:

X: Número de intentos hasta lograr realizar con éxito la alineación de un producto óptico en su montaje.

Paso 6. Deducción de la distribución: $X \sim Geométrica(0.8)$.

Paso 7. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar:

a. $P[X=5]=(0.2)^4(0.8)=0.0013$

Resp.: La probabilidad de que la alineación se produzca con éxito por primera vez en el quinto ensayo es de 0.0013 (1.3%).

b. $P[X \geq 4]=1-P[X \leq 3]=1-P[X=1]-P[X=2]-P[X=3]$
$$=1-(0.2)^0(0.8)-(0.2)^1(0.8)-(0.2)^2(0.8)=1-0.992=0.08$$

Resp.: La probabilidad de que la primera alineación con éxito requiera como mínimo cuatro intentos es de 0.08 (8.00%).

c. $P[4 \leq X \leq 6]=P[X \leq 6]-P[X \leq 3]=F_X(6)-F_X(3)$
$$=0.999936-0.992=0.007936$$

Esta forma de solución necesita de la Función de distribución acumulativa. Otra forma más simple es:

$$P[4 \leq X \leq 6]=P[X=4]+P[X=5]+P[X=6]=$$
$$=0.9984+0.99968+0.999936=0.007936$$

Resp.: La probabilidad de que la primera alineación con éxito requiera a lo menos cuatro y a lo más 6 intentos 0.0079 aprox. (7.9%, casi 8%).

d. $E(X)=\dfrac{1}{p}=\dfrac{1}{0.8}=1.25$ intentos.

Resp.: Se espera realizar entre 1 y 2 intentos para lograr realizar con éxito la alineación del producto óptico en el montaje. Esto explica las bajas probabilidades logradas en los cálculos de las preguntas anteriores.

e. $V(X)=\sigma_X^2=\dfrac{1-p}{p^2}=\dfrac{1-0.2}{(0.8)^2}=0.3125 \Rightarrow \sigma_X=0.5590$ intentos.

Resp.: La desviación estándar del número de intentos hasta realizar con éxito la alineación del producto óptico es de 0.559 intentos. Ni siquiera alcanza a ser 1.

Ejemplo 2.

(**Aplicación en Ciencias de la Salud**)
Un científico inocula varias ratas, una a la vez, con un germen de cierta enfermedad hasta lograr una que la haya contraído. Si la probabilidad de contraer la enfermedad es 1/6,

a. ¿Cuál es la probabilidad de que se requieran 8 inoculaciones hasta encontrar una rata que haya contraído la enfermedad?
b. ¿Cuál es la probabilidad de que por primera vez a lo más en la cuarta inoculación se encuentre una rata que haya contraído la enfermedad?
c. ¿Cuál es la probabilidad de que se requiera como mínimo cinco inoculaciones para encontrar la primera rata que haya contraído la enfermedad?
d. ¿Cuál es la probabilidad de que se requiera a lo menos cuatro y a lo más 6 inoculaciones para encontrar la primera rata que haya contraído la enfermedad?
e. ¿Cuántos intentos se espera realizar para encontrar la primera rata que haya contraído la enfermedad?
f. Calcular la desviación estándar del número de intentos hasta encontrar la primera rata que haya contraído la enfermedad?

Esquema de Solución.

Paso 1. Identificar el experimento:

Estudiar el número de inoculaciones hasta encontrar una rata que haya contraído la enfermedad.

Paso 2. Identificar el suceso E que se desea estudiar y que determina el éxito en la experiencia, y su contrario, E^C:

$E:$ Se encuentra una rata que haya contraído la enfermedad.
$E^C:$ No se encuentra una rata que haya contraído la enfermedad.

Paso 3. Identificar las probabilidades de E y E^C:

$$P(E) = p = \tfrac{1}{6} \quad P(E^C) = p = \tfrac{5}{6}.$$

Paso 4. Existe independencia entre los intentos de "inoculación hasta encontrar una rata que haya contraído la enfermedad".

Paso 5. Identificación de la variable aleatoria:

X: Número de inoculaciones hasta encontrar una rata que haya contraído la enfermedad.

Paso 6. Deducción de la distribución: $X \sim Geométrica(\tfrac{1}{6})$.

Paso 7. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar:

a. $P[X=8] = (\tfrac{5}{6})^8(\tfrac{1}{6}) = 0.0465$

Resp.: La probabilidad de que se requieran 8 inoculaciones hasta encontrar una rata que haya contraído la enfermedad es de 0.0465 (4.65%).

b. $P[X \le 4] = P[X=1] + P[X=2] + P[X=3] + P[X=4]$
$= (\tfrac{5}{6})^0(\tfrac{1}{6}) + (\tfrac{5}{6})^1(\tfrac{1}{6}) + (\tfrac{5}{6})^2(\tfrac{1}{6}) + (\tfrac{5}{6})^3(\tfrac{1}{6}) = 0.5178$

Resp.: La probabilidad de que se requieran a lo más 4 inoculaciones hasta encontrar un rata que haya contraído la enfermedad es de 0.5178 (51.78%).

c. $P[X \geq 5] = 1 - P[X \leq 4] = 1 - F_X(4) = 1 - 0.5178 = 0.4822$

Esta forma de solución necesita de la Función de Distribución Acumulativa (posiblemente una tabla de valores). Otra forma es:

$$P[X \geq 5] = 1 - P[X \leq 4] = 1 - P[X = 1] - P[X = 2] - P[X = 3] - P[X = 4]$$
$$= (\tfrac{5}{6})^0(\tfrac{1}{6}) + (\tfrac{5}{6})^1(\tfrac{1}{6}) + (\tfrac{5}{6})^2(\tfrac{1}{6}) + (\tfrac{5}{6})^3(\tfrac{1}{6})$$
$$= 1 - 0.1667 - 0.1389 - 0.1157 - 0.0965 = 1 - 0.5178 = 0.4822$$

Resp.: La probabilidad de que se requiera como mínimo cinco inoculaciones para encontrar la primera rata que haya contraído la enfermedad es de 0.4822 aprox. (48.22%).

d. $E(X) = \dfrac{1}{p} = \dfrac{1}{(1/6)} = 6$ intentos.

Resp.: Se espera realizar entre 6 inoculaciones para encontrar la primera rata que haya contraído la enfermedad. Esto explica la probabilidad en (a) y (b)es logradas. Calculando $1 - F_X(x)$ para los distintos valores de x compare los valores de $F_X(x)$ y de $1 - F_X(x)$, compare y emita un juicio que permita entender mejor lo que está pasando.

e. $V(X) = \sigma_X^2 = \dfrac{1-p}{p^2} = \dfrac{5/6}{(1/6)^2} = 30 \Rightarrow \sigma_X = 5.4772$ inoculaciones.

La desviación estándar del número de inoculaciones hasta obtener una rata infectada es de 5.48 inoculaciones (entre 5 y 6).

4.3.2 Ejercicios Propuestos.

1. Supongamos que la variable aleatoria X tiene una distribución geométrica con $p = 0,5$. Determinar las siguientes probabilidades:

   (a) $P[X = 1]$ Resp.: 0.5. (b) $P[X = 4]$ Resp.: 0.0625.

   (c) $P[X = 8]$ Resp.: 0.0039. (d) $P[X \leq 2]$ Resp.: 0.75. (e) $P[X > 2]$ Resp.: 0.25.

   Si X tiene una distribución geométrica, con una media de 2,5 determinar las siguientes probabilidades:

   (a) $P[X = 1]$ Resp.: 0.4. (b) $P[X = 4]$ Resp.: 0.0864.
   (c) $P[X = 5]$ Resp.: 0.05184. (d) $P[X \leq 3]$ Resp.: 0.784. (e) $P[X > 3]$ Resp.: 0.216.

2. (Ciencias de la Salud) En un estudio clínico, los voluntarios son puestos a prueba de un gen que se ha encontrado que aumenta el riesgo de contraer una enfermedad. La probabilidad de que una persona sea portadora del gen es de 0,1.
   a. ¿Cuál es la probabilidad de tener que probar 4 o más personas para detectar un portador del gen? Resp.: 0.729.
   b. ¿Cuántas personas se espera probar para detectar un portador del gen? Resp.: 10.

3. (Ciencias de las Comunicaciones) Supongamos que cada una de las llamadas a una popular emisora de radio tiene una probabilidad de 0,02 de obtener conexión (es decir, de no obtener una señal de ***ocupado***). Suponga que las llamadas son independientes.
   a. ¿Cuál es la probabilidad de que la primera llamada que se conecta la décima que se realiza? Resp.: 0.01668.
   b. ¿Cuál es la probabilidad de que se requiera más de cinco llamadas para que usted pueda conectarse? Resp.: 0.9039.
   c. ¿Cuál es el número medio de llamadas necesarias para conectarse? Resp.: 50.

4. (Ingeniería en Tránsito) En su viaje de cada mañana al trabajo, un particular semáforo está de color verde el 20% de las veces en usted se acerca en su vehículo a él. Supongamos que cada mañana representa un evento independiente.
   a. ¿Cuál es la probabilidad de que hayan pasado más de cinco mañanas hasta que usted se encuentre con el semáforo en verde? Resp.: 0.4096.
   b. ¿Cuál es la probabilidad de que hayan pasado veinte mañanas hasta que usted se encuentre con el semáforo en verde? Resp.: 0.9885.
   c. ¿Cuál es la probabilidad de que hayan pasado más de veinte mañanas hasta que usted se encuentre con el semáforo en verde? Resp.: 0.01153 (1.2% aprox.).
   d. ¿Cuál es la probabilidad de que la primera mañana en que la luz esté en verde sea la cuarta mañana en que usted se aproxima al semáforo? Resp.: 0.1024.

5. (Ciencias de la Ingeniería) Una empresa comercial tiene un equipo computacional que utiliza para el comercio al exterior de su región. La probabilidad de que el equipo falle en un día es de 0.005, y cuando falla el equipo es reparado inmediatamente por la noche. Cada día es un evento independiente.
   a. ¿Cuál es la probabilidad de que el equipo falle el primer día en que es puesto en funcionamiento? Resp.: 0.005.
   b. ¿Cuál es la probabilidad de que el equipo falle después del tercer día de funcionamiento? ¿Se puede estar tranquilo en la empresa al observar este resultado? Resp.: 0.9851. Se puede estar tranquilo porque la probabilidad de falla es baja y esto provoca que la probabilidad de que falle después de dos años de funcionamiento ininterrumpido siga siendo alta (0.902, 90.2% aprox.).
   c. ¿Cuál es el número medio de días de buen funcionamiento hasta que el equipo falle? Resp.: 200.

6. (Ciencias de la Ingeniería) En una operación de llenado automático, una báscula electrónica detiene la línea de fabricación después de detectar un paquete de peso inferior al normal. Supongamos que la probabilidad de que se produzca un paquete de bajo peso es de 0.001 y que el relleno de cada uno es independiente.
   a. ¿Cuál es el número medio de rellenos antes de que se detenga la línea de producción? Resp.: 1000.
   b. ¿Cuál es la desviación estándar del número de rellenos antes de que se detenga la línea de producción? Resp.: 999.5 aprox.

7. (Educación y Ciencias Humanas) La probabilidad de que una persona, que vive en un determinado lugar de la ciudad, sea propietaria de un perro se estima en 0,3. Encuentre la probabilidad de que una persona entrevistada al azar en décimo lugar sea la primera propietaria de un perro en esa zona de la ciudad. Resp.: 0.0121.

8. (Ciencias Biológicas) Un científico inocula varias ratas, una a la vez, con un germen de cierta enfermedad hasta encontrar una que la haya contraído. Si la probabilidad de contraer la enfermedad es

1/9, ¿cuál es la probabilidad de que se deba inocular 8 ratas hasta encontrar una que haya contraído la enfermedad? Resp.: 0.0487.

9. (Ciencias de la Ingeniería) La probabilidad de que un alumno pase la prueba escrita para una licencia de piloto privado, después de haber seguido un curso, es de 0,7. Encuentre la probabilidad de que el estudiante pase la prueba
   a. en el tercer intento. Resp.: 0.063.
   b. antes del cuarto intento. Resp.: 0.973.

10. (Transporte) En el control de equipaje de un aeropuerto se sabe que el 3% de las personas revisadas tienen objetos cuestionables en su equipaje.
    a. ¿Cuál es la probabilidad de que una cadena de 15 personas pase el control con éxito antes de que pase un individuo con un objeto cuestionable? Resp.: 0.019 aprox.
    b. ¿Cuál es el número esperado de individuos que pasan el control en una fila antes de que un individuo detenga el proceso? Resp.: 33.3333.

4.4. Distribución de Pascal o binomial negativa.

Palabras Clave: Distribución de probabilidad, función de cuantía, función de distribución acumulativa, esperanza matemática y varianza de una variable aleatoria.

Resumen de conceptos y propiedades.

✓ Definición de una v.a. X con distribución de Pascal (o binomial negativa):

Si en una serie de ensayos Bernoulli (ensayos independientes con probabilidad constante $p \in (0,1)$ de un suceso E), se define una variable X como "número de repeticiones de la experiencia de Bernoulli hasta la r-ésima ocurrencia de E", entonces X es llamada **variable aleatoria con distribución de Pascal de parámetros** r y p.

✓ La función de cuantía (o de masa) de una v.a. X con distribución de Pascal de parámetros r y p, lo cual es denotado por $X \sim Pascal(r,p)$, está dada por

$$f_X(x) = \binom{x-1}{r-1}(1-p)^{x-r} p^r \text{, para } r = 1,2,\ldots \text{ y } x = r, r+1, r+2, \ldots$$

✓ La función de distribución acumulativa (FDA) de probabilidades de $X \sim Pascal(r,p)$ es:

$$F_X(a) = P[X \le a] = \sum_{\substack{\forall x \le a \\ x=1,2,\ldots}} f_X(x) = \sum_{\substack{\forall x \le a \\ x=1,2\ldots}} \binom{x-1}{r-1}(1-p)^{x-r} p^r p \text{ , } a \in \mathbb{R}., r = 1,2,\ldots$$

✓ Si $\text{Rec}(X) = \{1,2,\ldots\} = \mathbb{N}$, entonces

$$P(X=j+1)=F_X(j+1)-F_X(j),\ j\in\mathbb{N};\ F_X(1)=P(X=1)=p.$$

- ✓ Valor esperado de $X \sim Pascal(r,p)$: $\boxed{E(X)=\mu_X=\frac{r}{p}}$.
- ✓ Varianza de $X \sim Poisson(\lambda)$: $\boxed{V(X)=\sigma_X^2=\frac{r(1-p)}{p^2}}$.
- ✓ Desviación Estándar: $\boxed{\sigma_X=\sqrt{\frac{r(1-p)}{p^2}}}$.

4.4.1 Ejercicios resueltos, paso a paso:

Ejemplo 1.

(**Aplicación en Ciencias de la Ingeniería)**

Un científico inocula varias ratas, una a la vez, con un germen de cierta enfermedad hasta encontrar 3 que la hayan contraído. Si la probabilidad de contraer la enfermedad es 0.25,

a. ¿Cuál es la probabilidad de que se requieran 8 inoculaciones?
b. ¿Cuál es la probabilidad de que a lo más en la cuarta inoculación se encuentre la tercera rata que haya contraído la enfermedad?
c. ¿Cuál es la probabilidad de que se requiera como mínimo cinco inoculaciones para encontrar la tercera rata que haya contraído la enfermedad?
d. ¿Cuál es la probabilidad de que se requiera a lo menos cuatro y a lo más 6 inoculaciones para encontrar la tercera rata que haya contraído la enfermedad?
e. ¿Cuántas inoculaciones se espera realizar para encontrar tres ratas que hayan contraído la enfermedad?
f. Calcular la desviación estándar del número de inoculaciones hasta encontrar la tercera rata que haya contraído la enfermedad.

Esquema de solución:

Paso 1. Identificar el experimento:

Estudiar el número de inoculaciones hasta encontrar 3 ratas que hayan contraído la enfermedad.

Paso 2. Identificar el suceso E que se desea estudiar y que determina el éxito en la experiencia, y su contrario, E^C:

E : Se encuentra una rata que haya contraído la enfermedad.
E^C : No se encuentra una rata que haya contraído la enfermedad.

Paso 3. **Identificar las probabilidades de E y E^C:**

$$P(E) = p = 0.25 \qquad P(E^C) = p = 0.75\,.$$

Paso 4. **Existe independencia entre los intentos de "inoculación hasta encontrar 3 ratas que hayan contraído la enfermedad".**

Paso 5. **Identificación de la variable aleatoria:**

X: Número de inoculaciones hasta encontrar 3 ratas que hayan contraído la enfermedad.

Paso 6. **Deducción de la distribución**: $X \sim Pascal(3;0.25)$.

Paso 7. **Traducir las preguntas a una expresión matemática adecuada, calcular y responder**:

a. $P[X=8] = (0.75)^{8-3}(0.25)^3 = 0.0779$

Resp.: La probabilidad de que se requieran 8 inoculaciones hasta encontrar 3 ratas que hayan contraído la enfermedad es de 0.0779 (7.79%).

b. $P[X \le 4] = P[X=3] + P[X=4]$

$$= (0.75)^2(0.25) + (0.75)^3(0.25) = 0.0508$$

Resp.: La probabilidad de que se requieran a lo más 4 inoculaciones hasta encontrar 3 ratas que hayan contraído la enfermedad es de 0.0508 (5.08%).

c. $P[X \ge 5] = 1 - P[X \le 4] = 1 - F_X(4) = 1 - 0.0508 = 0.9492$

Esta forma de solución necesita de la Función de Distribución Acumulativa (posiblemente una tabla de valores). Otra forma es:

$$\begin{aligned} P[X \ge 5] &= 1 - P[X \le 4] = 1 - P[X=3] - P[X=4] \\ &= (0.75)^2(0.25) + (0.75)^3(0.25) \\ &= 1 - 0.0156 - 0.0352 = 1 - 0.0508 = 0.9492 \end{aligned}$$

Resp.: La probabilidad de que se requiera como mínimo cinco inoculaciones para encontrar la tercera rata que haya contraído la enfermedad es de 0.9492 aprox. (94.92%).

d. $E(X) = \dfrac{r}{p} = \dfrac{3}{0.25} = 12$ intentos.

Resp.: Se espera realizar 12 inoculaciones para encontrar la tercera rata que haya contraído la enfermedad. ¿Podría usted explicar este valor?

e. $V(X)=\sigma_X^2=\dfrac{r(1-p)}{p^2}=\dfrac{3(0.75)}{(0.25)^2}=36 \Rightarrow \sigma_X=6$ inoculaciones.

Resp.: La desviación estándar del número de inoculaciones hasta encontrar la tercera rata que haya contraído la enfermedad es de 6 inoculaciones.

Ejemplo 2.

(**Aplicación en Ciencias de la Ingeniería)**

Suponga que en el viaje que usted realiza cada mañana a su trabajo, un particular semáforo está de color verde el 20% de las veces en que usted, en su vehículo, se acerca a él. Suponga que cada mañana representa un evento independiente.

a. ¿Cuál es la probabilidad de que la luz esté en verde en exactamente un día en a lo más cinco mañanas?
b. ¿Cuál es la probabilidad de que la luz esté en verde en cuatro días en más de 20 mañanas?
c. ¿Cuál es la probabilidad de que la segunda mañana en que la luz esté en verde sea la cuarta mañana en que usted se aproxima al semáforo?
d. ¿Cuál es el número esperado de días en que usted espera llegar al semáforo para que esté en verde por cuarta vez? ¿Cuál es la desviación estándar?

Esquema de solución:

Paso 1. Identificar el experimento:

Estudiar el número de días en que cierto semáforo está en verde cuando la persona se desplaza a su trabajo.

Paso 2. Identificar el suceso E que se desea estudiar y que determina el éxito en la experiencia, y su contrario, E^C:

E : El semáforo se encuentra con luz verde.
E^C : El semáforo no se encuentra con luz verde.

Paso 3. Identificar las probabilidades de E y E^C:

$$P(E)=p=0.2 \qquad P(E^C)=p=0.8.$$

Paso 4. Existe <u>independencia</u> entre los días en que "el semáforo está en verde cuando se va al trabajo cada día".

Paso 5. Identificación de la variable aleatoria:

X: Número de días en que cierto semáforo está en verde cuando la persona se desplaza a su trabajo.

Paso 6. Deducción de la distribución: $X \sim Pascal(r;0.2)$.

Paso 7. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar:

a. En este caso lo que se observa es si el semáforo está en verde en **exactamente una mañana**, por lo tanto $r=1$. Así,

$$P[X \leq 5] = F_X(5) = \sum_{x=1}^{5} \binom{x-1}{1-1} (0.8)^{x-1} (0.2)^1$$
$$= (0.8)^{1-1}(0.2)^1 + (0.8)^{2-1}(0.2)^1 + \ldots + (0.8)^{5-1}(0.2)^1 = 0.6723$$

Resp.: La probabilidad de que la luz esté en verde en exactamente un día en a lo más cinco mañanas es de 0.6723 (67.23%).

b. **Se pide calcular**: Probabilidad de que la luz esté en verde en **cuatro días** en más de ***20 mañanas***. En este caso $r=4$ y $X>20$, por tanto:

$$P[X > 20] = 1 - P[X \leq 20] = 1 - F_X(20) = 1 - \sum_{x=4}^{20} \binom{x-1}{4-1} (0.8)^{x-4} (0.2)^4$$

$$= 1 - \binom{4-1}{4-1}(0.8)^{4-4}(0.2)^4 - \binom{5-1}{4-1}(0.8)^{5-4}(0.2)^4 - \ldots - \binom{20-1}{4-1}(0.8)^{20-4}(0.2)^4$$

$$= 1 - 0.5886 = 0.4115$$

Para este cálculo conviene usar la planilla Excel de Microsoft Office.

Resp.: La probabilidad de que la luz esté en verde cuatro días en más de 20 mañanas es de 0.4115 (41.15%).

c. **Se pide calcular**: Probabilidad de que la ***segunda mañana en que la luz esté en verde*** sea la ***cuarta mañana en que usted se aproxima al semáforo***. En este caso $r=2$ y $X=4$, por tanto:

$$P[X = 4] = \binom{4-1}{2-1}(0.8)^{4-2}(0.2)^2 = 0.0768$$

Resp.: La probabilidad de que la luz esté en verde en cuatro días en más de 20 mañanas es de 0.0768 (7.68%).

d. $E(X) = \frac{r}{p} = \frac{4}{0.20} = 20$ días.

Se espera un promedio de 20 llegando al semáforo para que éste esté en verde por cuarta vez.

$V(X) = \sigma_X^2 = \frac{r(1-p)}{p^2} = \frac{4(0.8)}{(0.2)^2} = 40 \Rightarrow \sigma_X = 6.3246$ días.

Resp.: El número esperado de días en que usted espera llegar al semáforo para que esté en verde por cuarta vez es de 20 días y la desviación estándar es de 6,3246 días.

4.4.2 Ejercicios Propuestos.

1. Demostrar que la función de densidad de probabilidad de una variable aleatoria binomial negativa es igual a la función de densidad de probabilidad de una variable aleatoria geométrica cuando $r = 1$. Mostrar que las fórmulas para la media y la varianza de una variable aleatoria binomial negativa equivale a la de los resultados correspondientes cuando a la variable aleatoria geométrica cuando $r = 1$.

2. Considere la posibilidad de una secuencia de ensayos independientes de Bernoulli con $p = 0.2$.

   a. ¿Cuál es el número esperado de ensayos para obtener el primer éxito? Resp.: 5.
   b. ¿Cuál es la probabilidad de detener el proceso después del octavo intento? Resp.: 0.1678.

3. Supongamos que X es una variable aleatoria binomial negativa con $p = 0.2$ y $r = 4$. Determine lo siguiente:

   a. $E(X)$. Resp.: 20.
   b. $P[X = 20]$. Resp.: 0.0436.
   c. $P[X = 19]$. Resp.: 0.046.
   d. $P[X = 21]$. Resp.: 0.0411.
   e. El valor más probable de X. Resp.: 15 o 16.

4. En un estudio clínico, los voluntarios son puestos a prueba de un gen que se ha encontrado que aumenta el riesgo de contraer una enfermedad. La probabilidad de que una persona sea portadora del gen es de 0,1.
   a. ¿Cuál es la probabilidad de tener que probar 4 o más personas para detectar 2 portadores del gen? Resp.: 0.972.
   b. ¿Cuántas personas se espera probar para detectar 2 portadores del gen? Resp.: 20.

5. Demostrar que la función de densidad de probabilidad de una variable aleatoria binomial negativa es igual a la función de densidad de probabilidad de una variable aleatoria geométrica cuando $r = 1$. Mostrar que las fórmulas para la media y la varianza de una variable aleatoria binomial negativa equivale a la de los resultados correspondientes cuando a la variable aleatoria geométrica cuando $r = 1$.

6. En una operación de llenado automático, una báscula electrónica detiene la línea de fabricación después de detectar tres paquetes de peso inferior al normal. Supongamos que la probabilidad de un paquete de bajo peso es de 0.001 y que el relleno de cada uno es independiente.
   a. ¿Cuál es el número medio de rellenos antes de que se detenga la línea de producción? Resp.: 3000.
   b. ¿Cuál es la desviación estándar del número de rellenos antes de que se detenga la línea de producción? Resp.: 1731.18 (interprete este resultado).

7. Un sistema tolerante a errores que procesa transacciones para una empresa de servicios financieros utiliza tres computadores distintos. Si el computador en funcionamiento falla, una de los dos de reemplazo puede ponerse de inmediato en línea. Después de fallar el segundo computador, el último equipo puede ser puesto de inmediato en línea. Suponga que la probabilidad de un fallo en cualquier transacción es 10^{-8} y que las transacciones pueden considerarse eventos independientes. ¿Cuál es el número medio de transacciones antes de que los tres computadores hayan fracasado?
   Resp.: 300 000 000.

8. La probabilidad de que una persona, que vive en un determinado lugar de la ciudad, sea propietaria de un perro se estima en 0,3. Encuentre la probabilidad de que la persona entrevistada al azar en décimo lugar sea la quinta de tener un perro en ese lugar de la ciudad. Resp.: 0.0515.

9. Según un estudio publicado por un grupo de sociólogos de una prestigiosa universidad, cerca de dos tercios de los 20 millones de personas en este país que toman Valium son mujeres. Suponiendo que esta cifra sea una estimación válida, encontrar la probabilidad de que en un día determinado la quinta prescripción escrita por un médico para el Valium sea
   a. una prescripción de Valium por primera vez, en ese día, para una mujer. Resp.: 0.0082.
   b. una prescripción de Valium sea la tercero para una mujer. Resp.: 0.1975.

10. Suponga que hay una probabilidad de 0.8 de que una determinada persona creerá una historia sobre las aventuras de una famosa actriz. ¿Cuál es la probabilidad de que
   a. la sexta persona que escucha esta historia sea la cuarta en creerla? Resp.: 0.04096.
   b. la tercera persona que conozca esta historia sea la primera a creerla? Resp.: 0.032.

4.5. Distribución hipergeométrica.

Palabras Clave: Distribución de probabilidad, función de cuantía, función de distribución acumulativa, esperanza matemática y varianza de una variable aleatoria.

Resumen de conceptos y propiedades.

✓ Definición de una v.a. X con distribución Hipergeométrica:

Se tiene un conjunto A con N elementos. Se particiona A en dos subconjuntos E y E^C, con N_1 y $N_2 = N - N_1$ elementos respectivamente.

En A se elige una muestra aleatoria de n elementos sim remplazamiento de manera que $N_1 < N$ y $n < N$.

Se define una variable X como "número de elementos de E que están en la muestra aleatoria de tamaño n, entonces X es llamada **variable aleatoria con distribución Hipergeométrica** .

- ✓ La función de cuantía (o de masa) de una v.a. X con distribución Hipergeométrica, lo cual es denotado por $X \sim Hipergeométrica(N, N_1, n)$, está dada por

$$f_X(x) = \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}, \text{ para } x = Max\{0, n+N_1-N\} \text{ hasta } Min\{N_1, n\}$$

- ✓ La función de distribución acumulativa (FDA) de probabilidades de X es:

$$F_X(a) = P[X \le a] = \sum_{\substack{\forall x \le a \\ x=1,2,\ldots}} f_X(x) = \sum_{\substack{\forall x \le a \\ x=1,2\ldots}} \frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}, \; a \in \mathbb{R}.$$

- ✓ Si $\text{Rec}(X) = \{1, 2, \ldots, n\}$, entonces

$$P(X = j+1) = F_X(j+1) - F_X(j), \; j \in \mathbb{N}.$$

- ✓ Valor esperado de $X \sim Hipergeométrica(N, N_1, n)$: $\boxed{E(X) = \mu_X = np = \frac{nN_1}{N}}$.

- ✓ Varianza de $X \sim Hipergeométrica(N, N_1, n)$: $\boxed{V(X) = \sigma_X^2 = np(1-p)\left(\frac{N-n}{N-1}\right) \; ; \; p = \frac{N_1}{N}}$.

- ✓ Desviación Estándar de $X \sim Hipergeométrica(N, N_1, n)$: $\boxed{\sigma_X = \sqrt{np(1-p)\left(\frac{N-n}{N-1}\right)}}$.

4.5.1 Ejercicios resueltos, paso a paso:

Ejemplo 1.

(**Aplicación en Ciencias)**

Suponga que se venden 500 boletos de lotería y, entre ellos, 200 pagan al menos el costo del boleto. Suponga que usted compra 5 boletos. Encuentre la probabilidad de que

a. usted gane al menos el costo de 3 boletos. Sugerencia: pruebe que si se usa distribución binomial el resultado es 0.3174 (31.74%); pero si se usa distribución hipergeométrica se obtiene 0.3167 (31.67%). Comente estos resultados.
b. ¿Cuál es el número esperado de boletos que hay que comprar para obtener uno que pague al menos su valor? ¿cuál es el valor de la desviación estándar?

Esquema de solución:

Paso 1. Identificar el experimento:

Estudiar el número de boletos que paguen al menos su costo en una compra de 5 boletos de lotería.

Paso 2. Identificar el tamaño del conjunto A, el número elementos de E, E^C y el tamaño muestral n:

Tamaño de A: $N = 500$ boletos de lotería.
Tamaño de E: $N_1 = 200$ boletos que pagan al menos su costo.
Tamaño de E^C: $N_2 = N - N_1 = 300$.
Tamaño muestral: $n = 5$.

Paso 3. No existe independencia entre las elecciones de boletos: Las probabilidades de obtener un boleto premiado van cambiando de elección en elección.

Paso 4. Identificación de la variable aleatoria:

X: Número de boletos premiados entre los comprados.

Paso 5. Deducción de la distribución: $X \sim Hipergeom(500; 200; 5)$.

Paso 6. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar:

a. **Se pide calcular**: Probabilidad de que se gane ***al menos*** el costo de 3 boletos.

Si se soluciona mediante una distribución hipergeométrica, se tendrá:

$$P\left[X \geq 3\right] = P(X = 3) + P(X = 4) + P(X = 5) = \frac{\binom{200}{3}\binom{300}{2}}{\binom{500}{5}} + \frac{\binom{200}{4}\binom{300}{1}}{\binom{500}{5}} + \frac{\binom{200}{5}\binom{300}{0}}{\binom{500}{5}}$$
$$= 0.2308 + 0.0760 + 0.0099 = 0.3167$$

Puede suceder que un estudiante "resuelva" el problema razonando de la siguiente manera:

Si 200 boletos del total de 500 tienen premio, es decir, permiten "ganar al menos el costo del boleto", entonces la probabilidad de que un boleto tenga premio es $p = \frac{200}{500} = 0.4$. En este caso se define A: el boleto paga al menos su costo y se verificará $P(A) = 0.4$.

Si se compran 5 boletos, la probabilidad de que " al menos 3 ganen su costo" se expresa por $P(X \geq 3)$ y esta variable podría considerarse con distribución Binomial:

a. Es dicotómica: A, A^C.
b. Sus probabilidades son constantes: $P(A) = p = 0.4$,
$$P(A^C) = 1 - p = 1 - 0.4 = 0.6$$
c. $n = 5$ (fijo).
d. X : número de boletos que ganan al menos su costo entre $n = 5$ comprados.

Por lo tanto, si se calcula se obtiene:

$$P\left[X \geq 3\right] = P(X=3) + P(X=4) + P(X=5)$$
$$= \binom{5}{3}(0.4)^3(0.6)^2 + \binom{5}{4}(0.4)^4(0.6)^1 + \binom{5}{5}(0.4)^5(0.6)^0 = 0.3174$$

Comentario: Este es sólo un cálculo de **aproximación**. Mientras más grande sea la cantidad de boletos comprados mayor será la aproximación. Esta aproximación es útil cuando algunos de los cálculos de probabilidad en la distribución hipergeométrica (que es la que corresponde porque al comprar números la compra se hace sin reemplazo y por tanto la **independencia** exigida en la probabilidad binomial NO SE VERIFICA) puedan ser complicados de lograr o simplemente los medios de cálculo (calculadoras de bolsillo o computadores) no están capacitados para lograr los resultados de combinaciones.

Se debe recalcar que el valor que corresponde calcular es el de una distribución hipergeométrica.

Resp.: La probabilidad de que se gane al menos el costo de 3 boletos, al comprar 5 es de 0.3167.

b. $E(X) = \mu_X = \dfrac{nN_1}{N} = \dfrac{(5)(200)}{500} = 2$ boletos.

Resp.: Se espera en promedio obtener 2 boletos que paguen al menos su costo cuando se compran 5.

$$V(X) = \sigma_X^2 = np(1-p)\left(\tfrac{N-n}{N-1}\right) = \tfrac{(5)(200)}{(500)}\left(1-\tfrac{(200)}{(500)}\right)\left(\tfrac{(500)-(5)}{(500)-1}\right) = 1.1904$$

$$\Rightarrow \sigma_X = 1.0911 \text{ boletos.}$$

Resp.: La desviación estándar del número de boletos que pagan al menos su costo, cuando se compran 5 es de 1.0911 días (aprox. 1 boleto).

Ejemplo 2.

(**Aplicación en Administración Educacional**)

Existen dos vacantes en el Departamento de Estadística de cierta prestigiosa universidad. A un llamado a concurso se presentan cinco postulantes: tres tienen experiencia en modelos lineales y dos en probabilidad aplicada. El comité de búsqueda se encarga de elegir tres miembros en forma aleatoria.

a. ¿Cuál es la probabilidad de que dos tengan experiencia en modelos lineales?
b. ¿Cuál es la probabilidad de que a lo más dos tengan experiencia en modelos lineales?
c. ¿Cuál es el valor esperado de postulantes, en la muestra, que tienen experiencia en modelos lineales?

Esquema de solución:

Paso 1. Identificar el experimento:
Estudiar el número de postulantes que tienen experiencia en modelos lineales 3 que se eligen aleatoriamente de entre los que se presentan a un llamado a concurso.

Paso 2. Identificar el tamaño del conjunto A, el número elementos de E, E^C y el tamaño muestral n:

Tamaño de A: $N = 5$ postulantes al concurso.
Tamaño de E: $N_1 = 3$ postulantes tienen experiencia en modelos lineales.
Tamaño de E^C: $N_2 = N - N_1 = 2$ tienen experiencia en estadística aplicada.
Tamaño muestral: $n = 3$.

Paso 3. <u>No existe independencia</u> entre las elecciones de postulantes:

Las probabilidades de seleccionar un postulante que tenga experiencia en modelos lineales van cambiando de elección en elección.

Paso 4. Identificación de la variable aleatoria:

X: Número de postulantes que tienen experiencia en modelos lineales entre los tres seleccionados en forma aleatoria.

Paso 5. Deducción de la distribución: $X \sim Hipergeom(5;2;2)$.

Paso 6. Traducir las preguntas a una expresión matemática adecuada, calcular y contestar:

a. $P\left[X = 2\right] = \dfrac{\binom{3}{2}\binom{2}{1}}{\binom{5}{3}} = 0.6$

Resp.: La probabilidad de que tres seleccionados, de entre los cinco postulantes, tengan experiencia en modelos lineales es de 0.6 (60%).

b. En este caso lo que se observa que los valores de X son $x = 1, 2, 3$ (¿por qué?). por lo tanto se pide

$$P\left[X \le 2\right] = F_X(2) = P(X = 1) + P(X = 2) = \frac{\binom{3}{1}\binom{2}{2}}{\binom{5}{3}} + \frac{\binom{3}{2}\binom{2}{1}}{\binom{5}{3}} = 0.3 + 0.6 = 0.9$$

Resp.: La probabilidad de que a lo más dos tengan experiencia en modelos lineales es de 0.90 (90.00%).

c. $E(X) = \mu_X = \dfrac{nN_1}{N} = \dfrac{(3)(3)}{5} = 1.8$ postulantes.

Resp.: Se espera en promedio que 2 postulantes tengan experiencia en modelos lineales, entre los 3 elegidos aleatoriamente de un total de cinco que se presentaron al concurso.

Como un dato adicional podemos, sin compromiso de responder a pregunta alguna, calcular la desviación estándar de X:

$$V(X) = \sigma_X^2 = np(1-p)\left(\tfrac{N-n}{N-1}\right) = \tfrac{(3)(3)}{(5)}\left(1 - \tfrac{(3)}{(5)}\right)\left(\tfrac{(5)-(3)}{(5)-1}\right) = 0.36$$

$\Rightarrow \sigma_X = 0.6$ postulantes.

La desviación estándar del número de postulantes que, en la muestra, tendrán experiencia en modelos lineales es de 0.6 postulantes.

Intente interpretar este resultado.

4.5.2 Ejercicios Propuestos.

1. Suponga que X tiene una distribución hipergeométrica con N = 100, n = 4 y K = 20. Determine lo siguiente:
   a. $P[X=1]$.
   b. $P[X=6]$.
   c. $P[X=4]$.
   d. Determine la media y la varianza de X.

   Suponga ahora que X tiene una distribución hipergeométrica con N = 20, n = 4 y K = 4. Determine lo siguiente:
   a. $P[X=1]$. Resp.: 0.4623.
   b. $P[X=4]$. Resp.: 0.0002.
   c. $P[X\le 2]$. Resp.: 0.9866.
   d. Determine la media y la varianza de X. Resp.: E(X) = 0.8; V(X) = 0.539.

2. Supongamos que X tiene una distribución hipergeométrica con $N=10$, $n=3$ y $k=4$.
   a. Grafique la función de masa de probabilidad de X.
   Resp.: La función de cuantía (o de masa) de una v.a. con distribución hipergeométrica está dada por

$$f_X(x)=\frac{\binom{N_1}{x}\binom{N-N_1}{n-x}}{\binom{N}{n}}, \text{ con } x=máx\{0,n+N_1-N\} \text{ hasta } mín\{N_1,n\}.$$

   Por lo tanto, se dibujará un gráfico de líneas con valores dados por

$$f_X(x)=\frac{\binom{4}{x}\binom{10-4}{3-x}}{\binom{10}{3}}, \text{ con } x=máx\{0,3+4-10\}=0 \text{ hasta } mín\{4,3\}=3.$$

   b. Determinar la función de distribución acumulada de X.

   Resp.: De (a) se tiene

   i. $f_X(x)=\frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}}=\frac{1}{6}$

   ii. $f_X(x)=\frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}}=\frac{1}{2}$

   iii. $f_X(x)=\frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}}=\frac{3}{10}$

   iv. $f_X(x)=\frac{\binom{4}{3}\binom{6}{0}}{\binom{10}{3}}=0.0333333333$

   La función de distribución acumulada de X será, entonces

$$F_X(x) = \begin{cases} 0 & , si\ x < 0. \\ 1/6 & , si\ 0 \le x < 1. \\ 2/3 & , si\ 1 \le x < 2. \\ 29/30 & , si\ 2 \le x < 3. \\ 1 & , si\ x \ge 3. \end{cases}$$

3. Un lote de 75 lavadoras contiene 5 en las que la variabilidad en el grosor de la circunferencia de la lavadora es inaceptable. Una muestra de 10 lavadoras se selecciona al azar, sin reemplazo.
   a. ¿Cuál es la probabilidad de que ninguna de las lavadoras inaceptables esté en la muestra? Resp.: Si X es el número de lavadoras inaceptables, entonces
   $$f_X(0) = P(X = 0) = \frac{\binom{5}{0}\binom{70}{10}}{\binom{75}{10}} = 0.4786.$$
   b. ¿Cuál es la probabilidad de que al menos una lavadora inaceptable esté en la muestra?

   Resp.: $P(X \ge 1) = 1 - P(X = 0) = 1 - \frac{\binom{5}{0}\binom{70}{10}}{\binom{75}{10}} = 0.5214.$
   c. ¿Cuál es la probabilidad de que exactamente una lavadora inaceptable esté en la muestra?

   Resp.: $f_X(1) = P(X = 1) = \frac{\binom{5}{1}\binom{70}{9}}{\binom{75}{10}} = 0.8709\cdot$
   d. ¿Cuál es el número medio de lavadoras inaceptables en la muestra?
   Resp.: $\mu_X = E(X) = np = (10)(\frac{5}{75}) = \frac{2}{3}$;
   $$\sigma_X^2 = V(X) = np(1-p)(\tfrac{N-n}{N-1}) = (10)(\tfrac{5}{75})(\tfrac{70}{75})(\tfrac{65}{74}) = 0.5466.$$

5. (Aplicación en Ciencias de la Salud) Una empresa emplea a 800 hombres menores de 55 años. Supongamos que el 30% lleva un marcador en el cromosoma masculino que indica un mayor riesgo para la presión arterial alta. Si 10 hombres en la empresa se analizan para determinar el marcador en este cromosoma,
   a. ¿cuál es la probabilidad de que exactamente un hombre tenga el marcador? Resp.: 0.1201.
   b. ¿cuál es la probabilidad de que más de uno tenga el marcador? Resp.: 0.8523.

6. (Aplicación en Ciencias de la Ingeniería) Tarjetas de circuitos impresos se colocan en prueba de funcionamiento, después de haber sido dotadas con chips semiconductores. Un lote contiene 140 tarjetas, y 20 son seleccionadas sin sustitución, para la prueba de funcionamiento.
   a. Si 20 tarjetas son defectuosas, ¿cuál es la probabilidad de que al menos una tarjeta defectuosa aparezca en la muestra? Resp.: 0.9644.
   b. Si 5 tarjetas son defectuosas, ¿cuál es la probabilidad de que al menos una tarjeta defectuosa aparezca en la muestra? Resp.: 0.5429.

7. (Teoría de juegos) Un estado crea una lotería en la que 6 números son seleccionados al azar de 40, sin reemplazo. Un jugador elige seis números antes de que, también por elección al azar, se realice el sorteo oficial.
   a. ¿Cuál es la probabilidad de que los 6 números elegidos por el jugador concuerden con los 6 números del resultado oficial? Resp.: $2.6053 \cdot 10^{-7}$.
   b. ¿Cuál es la probabilidad de que 5 de los 6 números elegidos por el jugador aparezcan en el resultado oficial? Resp.: $5.3147 \cdot 10^{-5}$.
   c. ¿Cuál es la probabilidad de que 4 de los 6 números elegidos por un jugador aparezcan en el resultado oficial? Resp.: 0.02192.

d. Si un jugador entra en un sorteo cada semana, ¿cuál es el número esperado de semanas hasta las que el jugador acierta los seis números del resultado oficial? Resp.: La probabilidad de acertar los seis números en el sorteo de alguna semana es $P(X = 6) = 2.6053 \cdot 10^{-7} = p$. Si Y designa el número de semanas hasta obtener el pleno de seis números acertados, entonces Y se distribuye geométrica con $p = 2.6053 \cdot 10^{-7}$. Por lo tanto, $E(Y) = 1/p = 3\ 838\ 380$ semanas (algo así como 7 382 años).

8. (Ciencias Biológicas y Medioambientales) En los estudios de población biológica y medio ambiente a menudo se etiquetan y liberan especies con el fin de estimar tamaño y grado de ciertas características de la población.

   Diez animales de una determinada población que se cree extinta (o en extinción) son capturados, marcados y luego liberados en una región determinada. Información fidedigna afirma que hay 25 animales de este tipo en la región. Después de un período de tiempo se selecciona una muestra aleatoria de 15 de animales de este tipo en esa región. ¿Cuál es la probabilidad de que 5 de los animales seleccionados estén marcados? Resp.: 0.2315.

9. (Teoría de Probabilidades) Para evitar la detección en la aduana, un viajero lugares 6 comprimidos de estupefacientes en un frasco con 9 pastillas de vitaminas que son similares en apariencia. Si el funcionario de aduanas selecciona a 3 de los comprimidos al azar para su análisis, lo que es la probabilidad de que el viajero será detenido por posesión ilegal de narcóticos?

10. (Ciencias Biológicas) Un dueño de casa 6 plantas de bulbos seleccionados al azar de una caja que contiene 5 bulbos de tulipanes y narcisos 4 bulbos. ¿Cuál es la probabilidad de que él sembró 2 bulbos de narcisos y 4 bulbos de tulipán?

13. (Teoría de Muestreo) Una compañía manufacturera utiliza un plan de aceptación en los artículos de producción antes de su envío al comprador. El plan consta de dos etapas: Se prepara para el envío una caja de 25 artículos de la cual se toma una muestra de 3 y se prueba estas unidades para encontrar defectuosas. Si se encuentra alguno defectuoso, toda la caja es enviada de vuelta para el control del 100%. Si no se encuentran defectuosos, la caja es enviada al comprador.
    a. ¿Cuál es la probabilidad de que una caja que contiene 3 defectuosos sea enviada? Resp.: 0.3013.
    b. ¿Cuál es la probabilidad de que una caja que contiene sólo 1 defectuoso sea devuelta para su revisión total? Resp.: 0.12.

14. (Ciencias Sociales) Se estima que 4.000 de los 10.000 votantes residentes en una ciudad están en contra de un nuevo impuesto de ventas. Si 15 los votantes elegibles son seleccionados al azar y se les pidió su opinión, ¿cuál es la probabilidad de que más de 7 estén a favor el nuevo impuesto? Resp.: 0.2129.

15. (Control medioambiental) Un organismo gubernamental sospecha que algunas empresas están violando las leyes sobre regulación de la contaminación con respecto al dumping de cierto tipo de producto. Veinte empresas están bajo sospecha, pero todas no pueden ser inspeccionadas. Supongamos que 3 de las empresas están violando las leyes.
    a. ¿Cuál es la probabilidad de que en la inspección de 5 empresas no encuentre violaciones? Resp.: 0.3991.
    b. ¿Cuál es la probabilidad de que el plan permita encontrar dos violaciones? Resp.: 0.1316.