# Introducción a la Estadística (TD)

Ejercicios que resolveremos en la Clase Teórica

## Guía 1

**Ej. 1** En una ciudad hay 20 familias, 4 de ellas tienen un hijo, 8 tienen 2 hijos, 5 familias tienen 3 hijos, hay 2 familias con 4 hijos y una con 5 hijos.

*a*) Se selecciona al azar una familia, cual es la probabilidad de que tenga $i$ hijos con $i = 1, 2, 3, 4, 5$.

*b*) Se se selecciona al azar un niño, cual es la probabilidad de que pertenezca a una familia de $i$ hijos con $i = 1, 2, 3, 4, 5$.

**Ej. 2** En una urna hay 3 bolitas rojas y 4 bolitas blancas. Si se seleccionan dos bolitas sin reposición, ¿Cuál es la probabilidad de que la primera sea roja y la segunda blanca? Cual es la probabilidad de que la primera sea roja? Cual es la probabilidad de que la segunda sea roja?

**Ej. 3** **El Problema Del Cumpleaños**:

*a*) Calcular la probabilidad de que entre los 50 invitados a la fiesta hayan al menos dos que cumplan años en la misma fecha.

*b*) Si en lugar de 50 asisten 20, como cambia la probabilidad anterior?

*c*) Y si en lugar de 20 asisten $k$ con $k$ un numero entre 2 y 120, como cambia esa probabilidad con el valor de $k$. Grafique.

*d*) Identificar la cantidad mínima de invitados que deben ingresar para que la probabilidad de que se produzca al menos una repeticón en las fechas de cumpleaños sea superior a 0,6.

*e*) Utilizando la función `sample()` simular las fechas de cumpleaños (con enteros de 1 a 365) de los primeros 50 invitados que ingresan a la fiesta. Usar la función `duplicated()` para chequear si se produce alguna repeticón en las fechas de cumpleaños. Observación: usando `sum(duplicated())` podés contar cuántas repeticiones se producen.

*f*) Utilizar la función `sample()` para realizar 10 simulaciones de 50 invitados ingresando a la fiesta y calcular la frecuencia relativa con la que en ese grupo de 50 personas hayan al menos dos que cumplan años en la misma fecha. Repetir para 1000 y 100000 simulaciones. Comparar cada una de las 3 frecuencias relativas con la probabilidad calculada en el inciso (a).

## Guía 2

**Ej. 1** Se lanzan dos dados *justos*. Consideramos el evento

$$S = \{\text{la suma de los dos resultados obtenidos es } 7\}$$

Calcular la probabilidad de $S$ condicional a (dado que):

*a*) La suma sea 7.

*b*) La suma es mayor que 9.

*c*) El 2do dado salió par.

*d*) Al menos un dado salió impar.

*e*) Ambos dados salieron iguales.

*f*) Ambos dados salieron distintos.

**Ej. 2** En una urna hay 3 bolitas rojas y 4 bolitas blancas. Si se seleccionan dos bolitas sin reposición, Cual es la probabilidad de que la primera sea blanca y la segunda sea roja? Si saco tres sin repo, Cual es la probabilidad de que la primera sea blanca y la segunda sea roja y la tercera blanca? Cual es la probabilidad de que la segunda sea roja?

**Ej. 3** **Monty Hall**: Un certamen televisivo hay tres puertas cerradas. Detrás de dos de ellas hay una cabra y detrás de la otra un auto. El concursante elige una puerta pero no la abre. El animador (que sabe en qué puerta está el premio) abre una puerta que tiene una cabra entre las que no fueron elegidas por el concursante. Finalmente el concursante puede elegir abrir la puerta que había elegido inicialmente o la puerta que dejó sin abrir el animador.

El concursante sabe que el animador nunca abrirá la puerta con el auto.

*a*) Calcula la probabilidad de ganar el auto si el concursante abre la puerta que eligió inicialmente. (no cambia)

*b*) Calcula la probabilidad de ganar el auto si el concursante abre la ventana que dejó sin abrir el conductor. (cambia)

**Ej. 4** Una aseguradora divide a sus clientes en riesgo alto, medio y bajo. Dichos clientes tienen probabilidades de cometer un accidente de 0,25, 0,16 y 0,1 respectivamente.

*a*) Si el porcentaje de clientes de alto riesgo es 20 %, de riesgo medio es 30 % y de bajo riesgo es 50 %, entonces ¿cuál es la probabilidad de que un cliente seleccionado al azar cometa un accidente?

*b*) ¿'Si el cliente seleccionado cometio un accidente, cual es la probabilidad de que sea un cliente clasificado de bajo riesgo?

**Ej. 5** Una enfermedad afecta al 0.1 % de la población. Un test tiene una sensibilidad de 0.95 y una especificidad de 0.99. Es decir la probabilidad de que el test resulte positivo si la persona esta enferma es 0.95. Mientras que la probablidad de que el test de negativo si la persona es sana es 0.99.

*a*) Cual es la probabilidad de que una persona esté realmente enferma.

*b*) Si la persona le dió su examen positivo cuál es la probabilidad de que la persona este enferma?

*c*) Suponiendo que la enfermedad afecta al p % de la población, graficar usando R la probabilidad de estar enfermo dado un examen positivo vs $p$.

**Ej. 6** En un bolillero hay 20 bolillas rojas y 8 verdes, de las cuales 12 están marcadas (6 de cada color). Se extra una bolita. Considere los eventos:

- $A =$ "La bolilla extraída es roja".
- $B =$ "La bolilla extraída está marcada".

Calcular $P(A)$ y $P(A \mid B)$. Realizar los siguientes cálculos utilizando R:

*a*) Calcular la frecuencia relativa ($f_A$) de bolillas rojas en 1000 extracciones.

*b*) Usando las mismas 1000 extracciones, calcular la frecuencia relativa ($f_{A|B}$) de bolillas rojas entre las extracciones en que la bolilla obtenida está marcada.

*c*) Comparar $f_A$ y $f_{A|B}$ respecto a cuál se aproxima mejor a $P(A)$ y $P(A \mid B)$, respectivamente.

*d*) Calcular 100 frecuencias relativas de cada tipo e identificar cuál toma un valor más cercano a la probabilidad que aproxima. Contar en cuántos casos se verifica la respuesta del inciso anterior.

**Ej. 7** **Un Método de Sondeo:** Un investigador conducirá un sondeo con el objetivo de estimar el porcentaje de gente en la población que consume drogas ilícitas.

Llamemos $p$ a la probabilidad de que una persona elegida al azar consuma drogas. Elegimos una persona al azar y le preguntamos si consumió drogas ilícitas.

*a*) Supongamos que la persona no miente. Cuál es la probabilidad de que responda que sí consumió.

*b*) Supongamos que si la persona no consumio no miente, pero si consumió la probabilidad de que conteste que no consumio es $r$. Cuál es la probabilidad de que responda que sí consumió.

*c*) Preocupado de que mucha gente mentirá si se le pregunta directamente Ha consumido drogas ilícitas alguna vez? el investigador usará un método de sondeo conocido como de respuesta aleatoria. El método consiste en lo siguiente, se elige una persona al azar y se le da un dado y una moneda. Se le dice que tire un dado, si sale 1 ó 2, a continuación se le pide que tire una moneda. Si sale cara entonces la persona deberá responder que consume drogas y su sale ceca la persona deberá responder que no consume drogas. En cambio si el resultado del dado es 3, 4, 5, ó 6 se le pide a la persona que conteste a la verdad. (Obviamente que el encuestador no mira el resultado del dado ni la moneda). Calcular la probabilidad que la persona responda que si consumió.

**Ej. 8** Si $A$ y $B$ son dos eventos independientes, probar que $A^c$ y $B$, $A$ y $B^c$ y $A^c$ y $B^c$ también son independientes.

**Ej. 9** **Paradoja de Simpson** Para una campaña publicitaria de blanqueadores dentales, se dice que tener dientes amarillos aumenta la probabilidad de desarrollar enfisema basada en la siguiente informacíon.

| | Enfisema | No Enfisema |
|---|---|---|
| Dientes amarillos | 22 | 23 |
| Dientes Blancos | 10 | 45 |

*a*) Pruebe que tener dientes amarillos y tener enfisema no son eventos independientes.

*b*) La competencia dice que esta información es tramposa. Para ello muestra la siguiente información

**Fumadores**

| | Enfisema | No Enfisema |
|---|---|---|
| D. amarillos | 21 | 9 |
| D. Blancos | 7 | 3 |

**No Fumadores**

| | Enfisema | No Enfisema |
|---|---|---|
| D. amarillos | 1 | 14 |
| D. Blancos | 3 | 42 |

Esta de acuerdo con la competencia?

**Ej. 10** **Ensayos de Bernoulli** Consideremos $n$ Ensayos de Bernoulli independientes y todos con igual probabilidad de éxito. Calcular la probabilidad de:

*a*) ocurra al menos un éxito

*b*) ocurra exactamente un éxito

*c*) ocurra exactamente $k$ éxitos.

## Guía 3

**Ej. 1** El número de mails que llegan a una casilla en un día sigue una distribución de Poisson de parámetro $\lambda = 5$. Calcular la probabilidad de:

- *a*) lleguen exactamente 3 mails.
- *b*) lleguen más de 3 mails.
- *c*) lleguen a lo sumo 2 mails.
- *d*) lleguen al menos 2 mails.
- *e*) llegue algún mail.

**Ej. 2** Un local que vende heladeras, tiene en stock 2 heladeras de marca Masfrio. Si la demanda por día de heladeras Masfrio es una variable con distribución de Poisson de parámetro $\lambda = 3{,}5$. Calcular la probabilidad puntual del número de heladeras vendidas en un día.

**Ej. 3** **Sobreventa:** Supongamos que un avión tiene 100 asientos. La aerolínea vende 110 pasajes. La probabilidad de que un pasajero se presente el dia del vuelo a tomar el avión es 0,9. Y supongamos que cada pasajero se presenta o no a volar independientemente del resto. S

- *a*) Calcule la probabilidad de que todos los pasajeros que se presentan a volar puedan hacerlo.
- *b*) Sea $Y$ = "Número de pasajeros que se presentan a volar". Cuál es la distribución de $Y$.
- *c*) Si la aerolíneas esta evaluando vender un número $M$ pasajes en lugar de 110. Grafique la probabilidad anterior en función de $M$.
- *d*) Sea $Z$ = "Número de pasajeros que no pueden volar". Que distribución tiene $Z = max\{\ 0, Y - 100\}$.
- *e*) Supongamos que cada pasaje vendido le significa 500 dolares de ganancia a la aerolínea, cada pasajero que no puede volar una perdida de 600 dolares y cada vuelto un costo operatico de 4000 dolares. Sea la variable aleatoria $X$ = "Ganancia Neta", definida como: $X = 500 \times 110 - 600 \times \text{máx}\{0, Y - 100\} - 40000$, cual es la distribución de $X$.
- *f*) Simula 100 vuelos del avión (es decir considera 100 instancias distintas donde la compañía vende 110 asientos tiendo a disposición para los pasajeros solo 100) y grafica un histograma con las realizaciones de $X$ (las ganancia neta en cada uno de los 100 vuelos).
- *g*) Ejecuta varias veces tus líneas de código y actualiza la grafica con la distribución empírica (simulada) de la ganancia neta en cada actualización. Que puedes notar? A que se debe el comportamiento cambiante del histograma de la ganancia neta
- *h*) Repite el ejercicio anterior, pero ahora simula $r = 10{,}000$ vuelos del avión. Que diferencias notas con respecto al inciso anterior.
- *i*) Que ocurre con la distribución de la ganancia neta si mantenemos la política de sobre–vender 10 asientos del avión pero la probabilidad de que un pasajero se presente a volar pasa del 0,9 al 0,95?
- *j*) Mediante simulaciones numéricas optimiza la estrategia de sobre–venta de pasajes (elige el valor óptimo de $M$) de tal forma que la probabilidad de obtener beneficios positivos sea de como mínimo el 90 %.
- *k*) Calcule la $E(Y)$ y $E(X)$.
- *l*) Como podrías aproximar la media y la varianza de la variable aleatoria $X$ utilizando tus simulaciones numéricas?

**Ej. 4** **Clasificación**: Las aves parásitas de cría ponen huevos en nidos de otras especies (hospedador), las cuales incuban los huevos y crían al pichón parásito. En un bosque de talas de la provincia de Buenos Aires hay dos especies hospederas que son indistinguibles a simple vista. Una de las principales diferencias entre estas especies radica en el grado de discriminación y remoción de huevos paráitos de sus nidos. Una de las especies es aceptadora de huevos parásitos, ya que remueve del nido sólo el 30 % de los huevos parásitos, mientras que la otra especie es rechazadora ya que remueve el 80 % de los huevos parásitos presentes en su nido. Además, se sabe que el 90 % de los nidos del bosque corresponden a la especie aceptadora , mientras que apenas el 10 % restante son nidos de la especie rechazadora.

Se elige al azar un nido del bosque y se colocan $n = 8$ huevos parásitos. Denotemos con $X$ a la variable aleatoria que indica el número de huevos removidos del nido. Asuma que, en cada nido, la remoción (o no) de los diferentes huevos se realiza de manera independiente.

*a*) Calcule la probabilidad de que 5 de los huevos parásitos sean removidos del nido si se sabe que el nido parasitado es de la especie aceptadora.

*b*) Calcule la probabilidad de que $x$ de los huevos ajenos sean removidos del nido si se sabe que el nido parasitado es de la especie aceptadora, para $0 \leq x \leq 8$.,

*c*) Calcule la probabilidad de que 5 de los huevos parásitos sean removidos del nido si se sabe que el nido parasitado es de la especie rechazadora.

*d*) Calcule la probabilidad de que $x$ de los huevos parásitos sean removidos del nido si se sabe que el nido parasitado es de la especie rechazadora, para $0 \leq x \leq 8$.

*e*) Calcule la probabilidad de que 5 de los huevos parásitos sean removidos del nido.

*f*) Calcule la probabilidad de que $x$ de los huevos parásitos sean removidos del nido, para $0 \leq x \leq 8$.

*g*) Hallar $E(X)$ y $V(X)$.

**Ej. 5** **La Catadora de Té** : Una señora afirma que al probar una taza de té con leche puede distinguir qué fue lo primero que se echó en la taza: el té o la leche. Cómo podríamos hacer un experimento para determinar si la señora sabe distinguir?

Procuramos realizar un experimento que indique evindencia a favor de la señora. Para ello, procedemos de la siguiente manera: le damos $n = 12$ tazas preparadas de diferentes maneras y contamos el número de veces que la mujer acierta el orden en el que fue preparado el té. Sea $X$ la variable aleatoria que cuenta el número de aciertos en los $n = 12$ ensayos.

a) ¿Qué distribución propone para la variable aleatoria $X$? ¿Cuáles son los parámetros de esta distribución?

b) ¿Cuál es la probabilidad de que acierte las 12 tazas si responde al azar?

¿Le daremos la oportunidad de equivocarse? ¿Cuántas tazas serán suficientes para que confiemos en que la señora tiene la habilidad que ella afirma?

Propongamos diferentes reglas que podemos utilizar para decidir que la señora es una entendida en el tema.

- Criterio 0: Determinamos que sabe si acierta las 12 veces.
- Criterio 1: Determinamos que sabe si acierta 11 o más veces.
- Criterio 2: Determinamos que sabe si acierta 10 o más veces.
- Criterio 3: Determinamos que sabe si acierta 9 o más veces.

c) Supongamos que la señora responde al azar. Para cada uno de los criterios propuestos, calcule la probabilidad de que consideremos que la señora sabe.

d) Que criterio debemos utilizar si queremos que la probabildad de equivocarnos al decidir que sabe cuando en realidad responde al azar no sea superior el valor 0,02? Es decir, que número mínimo de aciertos serían suficientes para que confiaras en que la señora tiene la habilidad si queremos que la probabilidad de equivocarnos cuando en realidad responde al azar no sea superior el valor 0,02?

e) Supongamos que la señora responde bien en 10 ocaciones. ¿Cuál es el criterio más exigente que nos permite decir que la señora sabe? ¿Cúal es la probabilidad de error de dicho criterio cuando en realidad responde al azar? ¿Cuál es la probabilidad de observar un valor tan o más grande que el observado, asumiendo que responde al azar?

f) Si en realidad la señora tiene un verdadero don y acierta al orden en que se pone la leche y el té con probabilidad igual a 0,8. Para cada uno de los criterios propuestos, calcule la probabilidad de que consideremos que la señora sabe. Y si la probabilidad de acertar con el orden es 0,9, ¿como cambia la probabilidad de decidir que la señora sabe con cada uno de los criterios?

g) Si en realidad la señora tiene un verdadero don y acierta al orden en que se pone la leche y el té con probabilidad igual a 0,8. Para cada uno de los criterios propuestos, calcule la probabilidad de que consideremos que la señora NO sabe. Y si la probabilidad de acertar con el orden es 0,9, ¿como cambia la probabilidad de decidir que la señora NO sabe con cada uno de los criterios?

## Guía 4

**Ej. 1** Una empresa vende dos variedades de soja. La variedad I tiene un rendimiento por ha. que puede considerarse una variable aleatoria con distribución $\mathcal{N}(27, 25)$; y la variedad II tiene un rendimiento por ha. que puede considerarse $\mathcal{N}(40, 25)$. Un cliente realizó una compra de semillas de la variedad II y antes de continuar comprando a esta empresa, quiere asegurarse de que las semillas que le enviaron realmente pertenecen a esa variedad y no a la variedad I.

Sea $X$ la variable aleatoria que mide el rendimiento de soja en una hectárea cultivada.

*a*) ¿Qué distribución tiene $X$ si se utiliza la variedad I?

*b*) ¿Qué distribución tiene $X$ si se utiliza la variedad II?

*c*) ¿Cuán alto debe ser el valor de $X$ para considerar que, efectivamente, le enviaron la variedad II?

Propongamos diferentes reglas que podemos utilizar para determinar que las semillas plantadas son de la variedad II

- Criterio 0: Determinamos que son de alto rendimiento si obtenemos 30 o más tn./ha.
- Criterio 1: Determinamos que son de alto rendimiento si obtenemos 33 o más tn./ha.
- Criterio 2: Determinamos que son de alto rendimiento si obtenemos 35 o más tn./ha.
- Criterio 3: Determinamos que son de alto rendimiento si obtenemos 38 o más tn./ha.

1) Supongamos que las semillas son de la variedad I. Para cada uno de los criterios propuestos, calcule la probabilidad de que consideremos que las semillas son de alto rendimiento.
2) Que criterio debemos utilizar si queremos que la probabildad de equivocarnos al decidir que las semillas son de alto rendimiento cuando en realidad son de la variedad I no sea superior el valor 0,05?
3) Supongamos al probar las semillas obtenemos 34 toneladas en una hectárea. ¿Cuál es el criterio más exigente que nos permite decir que se utilizó la variedad II? ¿Cúal es la probabilidad de error de dicho criterio cuando en realidad se utiliza la variedad I? ¿Cuál es la probabilidad de observar un valor tan o más grande que el observado, asumiendo que se utiliza la variedad I?
4) Supongamos que las semillas son de la variedad II, para cada uno de los criterios propuestos, calcule la probabilidad de que consideremos que las semillas son de la variedad I.
5) Construye un criterio de modo tal que la probabilidad de decidir que son de la variedad II cuando en realidad son de la variedad I sea 0.01.

## Guía 5

**Ej. 1** Sea $X$ el retorno de un portfolio conservador e $Y$ el retorno de un portfolio agresivo. Ambas variables estan medidas en los últmis 48 meses y cada 100 dolares. Se sabe que la probabilidad puntual conjunta de $(X, Y)$ es

| Y \X | -2 | -1 | 1 | 2 |
|---|---|---|---|---|
| -4 | 0.06 | 0.05 | 0.02 | 0.01 |
| -2 | 0.04 | 0.17 | 0.10 | 0.04 |
| 2 | 0.03 | 0.09 | 0.19 | 0.05 |
| 4 | 0.01 | 0.02 | 0.05 | 0.07 |

*a*) Hallar las marginales

*b*) Hallar la distribución de $Z = 3X + 2Y$.

**Ej. 2** En una jaula hay un ratón y dos pollitos. La probabilidad de que en la próxima hora cada animal salga de la jaula 0.3 e independiente de los otros animales. Pasa una hora y se definen las siguientes variables, $X$ el número de patas que quedan en la jaula e $Y$ el número de animales que quedan en la jaula. Hallar la probabiliad puntual conjunta de $(X, Y)$. Hallar $E(X + Y)$.

**Ej. 3** Continuando el ejercicio 5.1. Hallar $E(3X + 2Y)$. Si el retorno conservador será de 100 dolares ($X = 1$). Cual es la probabilidad puntual del retorno agresivo.

**Ej. 4** **Clasificación**: Las aves parásitas de cría ponen huevos en nidos de otras especies (hospedador), las cuales incuban los huevos y crían al pichón parásito. En un bosque de talas de la provincia de Buenos Aires hay dos especies hospederas que son indistinguibles a simple vista. Una de las principales diferencias entre estas especies radica en el grado de discriminación y remoción de huevos paráitos de sus nidos. Una de las especies es aceptadora de huevos parásitos, ya que remueve del nido sólo el 30 % de los huevos parásitos, mientras que la otra especie es rechazadora ya que remueve el 80 % de los huevos parásitos presentes en su nido. Además, se sabe que el 90 % de los nidos del bosque corresponden a la especie aceptadora , mientras que apenas el 10 % restante son nidos de la especie rechazadora. Se elige al azar un nido del bosque y se colocan $n = 8$ huevos parásitos. Denotemos con $X$

a la variable aleatoria que indica el número de huevos removidos del nido. Asuma que, en cada nido, la remoción (o no) de los diferentes huevos se realiza de manera independiente. Y sea $Y$ la variable que vale 1 si la especie es aceptadora y 0 si es rechazadora

*a*) Hallar la función de probabilidad conjunta del vector $(X, Y)$.

*b*) Considere la regla de clasificación que determina que un nido corresponde a la especie rechazadora si la cantidad de huevos removidos es mayor o igual a $t = 5$. Caso contrario, determina que el nido es de la especie hospedadora aceptadora. Calcule el error de medio de clasificación, es decir, calcule la probabilidad de que la regla de clasificación, clasifique mal.

*c*) Considere la regla de clasificación que determina que un nido corresponde a la especie rechazadora si la cantidad de huevos removidos es par. Caso contrario, determina que el nido es de un hospedador aceptador. Calcule el error de medio de clasificación, es decir, calcule la probabilidad de que la regla de clasificación, clasifique mal.

*d*) Cuál de las reglas consideradas en los items anteriores prefiere?

*e*) Para cada $t \in \{0, \ldots, 8\}$, considere la regla de clasificación que determina que un nido corresponde a la variedad rechazadora si la cantidad de huevos removidos es mayor o igual a $t$. Caso contrario, determina que el nido es del hospedador aceptador. Calcule el error de medio de clasificación correspondiente a cada regla con punto de corte en $t$ para $0 \leq t \leq 8$.

*f*) Con qué punto de corte obtiene la regla con menor error de clasificación?

**Ej. 5** Sea $N$ el número de clientes que realizan una compra en un supermercado en un día de la semana. Supongamos que $N$ tiene distribución de Poisson con parámetro 3000. Sea $Y_i$ el consumo en pesos del cliente $i$ y supongamos que los clientes consumen en forma independientes unos de otros. Sabemos que $E(Y_i) = 500$ y $V(Y_i) = 900$. Cúal es la esperanza y la varianza total del consumo de clientes en una semana?

## Guía 6

**Ej. 1** Sean $X$ e $Y$ el retorno de 2 activos si se invierten 100 dólares. Se sabe que $(X, Y)$ tienen distribución normal multivariada con media $(0{,}15, 0{,}25)$, desvios $(0{,}2, 0{,}5)$ y $cov(X, Y) = 0{,}01$. Si invertis 50 dólares en cada portfolio

*a*) Cuál es la esperanza y la varianza del retorno total.

*b*) Como se distribuye el retorno total.

*c*) Calcular la probabilidad de tener un retorno total positivo.

*d*) Obtener el cuantil 0.9 del retorno total si $X = 0{,}18$

## Guía 7

**Ej. 1** Sea $U_1, \ldots, U_n$ una muestra aleatoria con distribución $\mathcal{U}[a, b]$ y sea $h$ una función continua. Consideremos la integral

$$I = \int_a^b h(x)\, dx$$

Proponga una forma de calcular numericamente el valor de $I$ basado en $U_1, \ldots, U_n$.

**Ej. 2** Un investigador conducirá un sondeo con el objetivo de estimar el porcentaje de gente en la población que consume drogas ilícitas. Supongamos que la probabilidad de que una persona elegida al azar consuma drogas ilícitas es 0,4.

Sea $X$ la variable aleatoria que cuanta el número de personas entre n encuestas que responden que consume drogas. Si se encuestan a 200 personas calcular de dos formas distintas $P\left(0{,}38 \leq \frac{X}{n} \leq 0{,}42\right)$. Comparar con la cota de Tchebychev.

**Ej. 3** Cuando se realiza una encuesta con preguntas delicadas, donde la gente tiene cierta resistencia a responder, se suele usar una método indirecto para hacer la pregunta. Supongamos que estamos interesados en conocer (o estimar) la proporción de personas que consumen drogas. Un modo de realizar la encuesta sería el siguiente: a cada persona le damos un dado y una moneda y le decimos que tire el dado. Si el resultado del dado es 3, 4, 5, o 6 la persona responde la verdad y si sale 1 o 2 tira la moneda; si sale cara responde SI consume y si sale ceca responde NO consume.

Obviamente, el encuestador no ve el resultado del dado ni el de la moneda y por lo tanto no sabría si la respuesta del encuestado es la verdad o es producto del azar.

A partir de este mecanismo quisieramos poder estimar la verdadera probabilidad $p$ de que una persona elegida al azar consuma drogas.

*a*) Calcular (en función de $p$) la probabilidad de que una persona elegida al azar responda afirmativamente a la pregunta ($P(SI)$).

*b*) A partir de encuestar a $n$ personas con este mecanismo, como estimaría la probabilidad de que alguien responda que SI, es decir $P(SI)$? Mas precisamente, si pensamos en $n$ variables i.i.d $W_1, \ldots, W_n$ donde $W_i = 1$ si la persona responde SI y 0 en caso contario, que cuenta deberá hacer para estimar la proporción de gente que responde $SI$, P($SI$).

*c*) Usando a) y b), proponga una cuenta para estimar $p$, la probabilidad de que una persona elegida al azar consuma drogas. . Llamemos $\hat{p}$ a esta cuenta.

*d*) Para convencernos que este mecanismo resulta efectivo, realice la siguiente simulación. Supongamos que el verdadero valor de $p = 0{,}1$ es decir la probabilidad de que una persona elegida al azar consuma drogas es 0.1. Simule $n = 100$ respuestas, $W_1, \ldots, W_n$ según el mecánismo de respuesta propuesto y compare con la verdadera proporción $p$. Repita esto $Nrep = 1000$ veces.

*e*) Si $p = 0{,}1$, ¿ qué distribución tiene $W_i$? Usando el Teorema Central del Límite, aproxime la probabilidad $P(0 < \hat{p} < 0{,}2)$ cuando $n = 100$.

*f*) Supongamos como en el ejercicio anterior que $p = 0{,}1$ pero hacemos la encuenta directa y nadie miente. En ese caso, estimaríamos a $p$ simplemente contando el porcentaje de gente que contesta si. Es decir, tendríamos $X_1, \ldots, X_n$ variables aleatorias i.i.d $Bi(1, p)$ y estimaríamos $p$ con $\bar{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Usando el Teorema Central del Límite, aproxime la probabilidad $P(0 < \bar{X_n} < 0{,}2)$.

*g*) Explique la relación o las diferencias entre los dos item anteriores.