Universidad de la República **Probabilidad y Estadística**
Facultad de Ingeniería - IMERL **Primer Semestre 2017**

Examen Julio
Sábado 22 de julio 2017.

Número de Examen	Cédula	Nombre y Apellido

PARA USO DOCENTE				
Ej. 1	Ej. 2	Ej. 3	Ej. 4	TOTAL

Ejercicio 1. [**15 puntos**]

De los correos que llegan a una cuenta de mail, se sabe que el 20 % son spam (correos no deseados). Un detector automático sencillo de spam busca frases que se repitan en dichos correos, en particular se encontró que la frase“dinero gratis” aparece en el 10 % de los correos no deseados, mientras que aparece en apenas el 1 % de los correos normales (es decir que no son considerados spam).

1. Calcular la probabilidad de que un correo entrante contenga la frase “dinero gratis”.
2. Un correo entrante contiene la frase “dinero gratis¿‘cuál es la probabilidad de que sea spam? ¿y de que no lo sea?

Ejercicio 2. [**30 puntos**] Los siguientes datos representan las alturas de una muestra de 10 hombres y 10 mujeres elegidos al azar del curso de PyE 2017 (primer semestre).

Mujeres (cm)	168	170	167	168	159	170	173	167	160	165
Hombres (cm)	188	165	193	173	179	173	168	174	179	183

Denotamos por X la variable aleatoria que mide la altura de las mujeres e Y la de los hombres. Suponemos que X e Y tienen distribución normal.

1. *a*) Hallar un intervalo de confianza exacto, al nivel de confianza 0,9, para la esperanza de X.
   b) Hallar un intervalo de confianza exacto, al nivel de confianza 0,9, para la varianza de Y.
2. Se sabe que en dicho curso la proporción de mujeres es de 0,2. Si se sabe que la altura de un estudiante elegido al azar del curso está entre 168 cm y 169 cm. ¿Es más probable que sea un hombre o una mujer? Asumir que el valor esperado y la varianza de X e Y son los estimados a partir de la muestra.

Ejercicio 3. [**30 puntos**] Una moneda tiene probabilidad p de salir cara. Sea X la variable aleatoria que cuenta el número de lanzamientos de la moneda hasta que salga cara por segunda vez (si sale cara en el primer y en el segundo lanzamiento X vale 2).

1. Sea $X_1, \ldots, X_n$ una muestra i.i.d. de X. Hallar el estimador de máxima verosimilitud de p.
2. La tabla siguiente resume $n = 500$ valores observados de X.

Valor observado de X	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Frecuencia	74	98	96	66	47	42	26	20	11	7	2	6	0	3	1	1

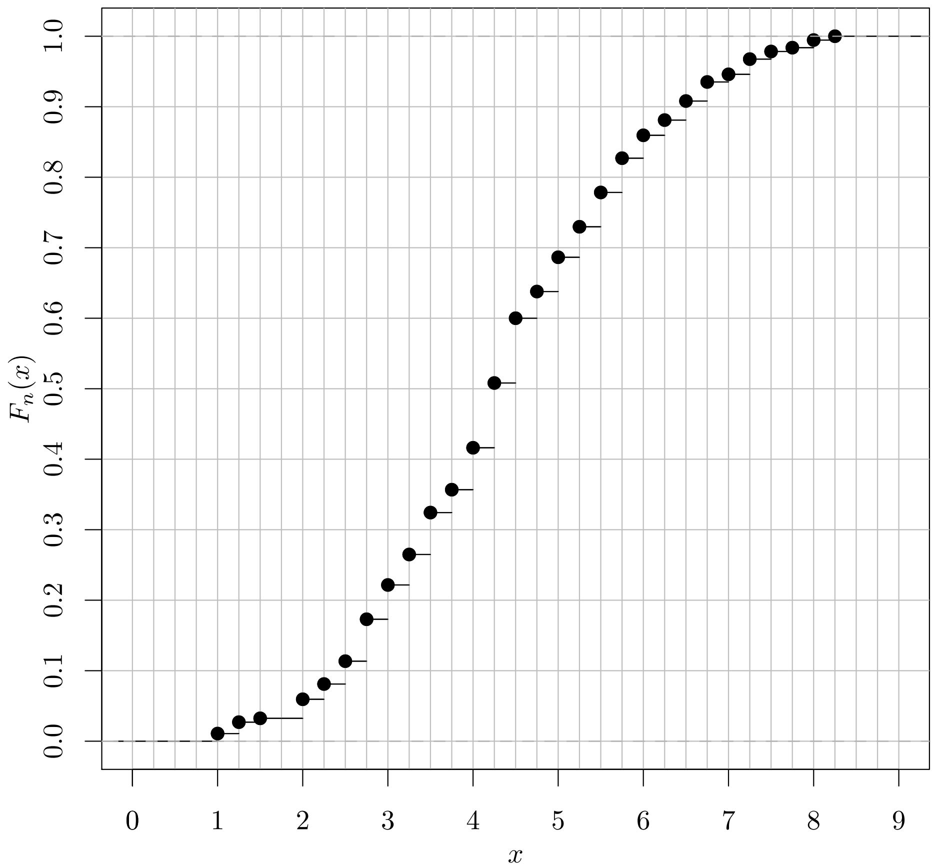
   Estimar p a partir de estos datos.

3. ¿Se le ocurre otro estimador de p? En caso afirmativo calcularlo.

Ejercicio 4. [**25 puntos**]

Se consideran las 4 últimas cifras de un número de teléfono celular. Se asume que cada una de estas cifras puede ser modelada por una variable aleatoria uniforme en el conjunto $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Se asume además que las cifras son independientes entre sí. Para $i = 1, 2, 3, 4$ se define X_i como la i-ésima cifra de las últimas cuatro: si el número es 099123456, entonces $X_1 = 3$, $X_2 = 4$, $X_3 = 5$ y $X_4 = 6$.

1. Se define $\bar{X}_4 = \frac{1}{4} \sum_{i=1}^{4} X_i$. Si utilizamos el teorema central del límite (TCL) para aproximar la distribución de $\bar{X}_4$ ¿qué distribución obtendríamos? ¿cuáles son sus parámetros?
2. Sea X una variable aleatoria con la distribución hallada en la parte anterior. Hallar los cuantiles x_p de X para $p \in \{0{,}2; 0{,}4; 0{,}5; 0{,}6; 0{,}8; 0{,}9\}$.
3. A continuación se muestra la función de distibución empírica $F_n(x)$ de los promedios de las últimas cuatro cifras de $n = 185$ números de teléfono recolectados durante el último curso de PyE. Hallar los cuantiles empíricos $\hat{x}_p$ de dicha muestra para los mismos valores de p que antes.



4. Graficar $(x_p, \hat{x}_p)$ para los valores de p anteriores (graficar x_p en el eje de las x y $\hat{x}_p$ en el eje de las y). Si los puntos están cerca de la recta $y = x$ entonces podemos decir que hay un buen ajuste entre la distribución del promedio y su aproximación. ¿Cuál es su conclusión sobre el ajuste en este caso?