Estadística en Física Experimental (1^{er} Cuatrimestre 2025)

Guía de Problemas N° 2 | Variables aleatorias discretas – Hipergeométrica, Binomial, Binomial Negativa, Poisson, etc. Función Generatriz.

1. Se lanzan diez dados. ¿Cuál es la probabilidad de obtener exactamente tres seis? [Rta: 0.155]

2. En una caja de resistencias hay 20 defectuosas y 80 buenas. Tomamos 10 al azar. ¿Cuál es la probabilidad de que la mitad sean buenas? [Rta: 0.0215].

3. Unos amigos juegan con un par de dados. Los tiran n veces y gana quien acierte cuantas veces van a sumar 7. ¿Cuántas veces deberían tirar los dados para que lo más probable fuera ganar jugándole a que sale 10 veces? [Rta: $60 \leq n \leq 64$]

4. Se lanza repetidamente una moneda cargada hasta que el primer resultado vuelve a aparecer. Muestre que:
   (a) La probabilidad de que el juego termine en el lanzamiento $n+2$ es $P_n = p^2q^n + q^2p^n$, donde p es la probabilidad de que salga cara y $q = 1 - p$.
   (b) P_n permite definir una función que a cada subconjunto de $A \subseteq \mathbb{N}$ le asigna una probabilidad. Defínala y muestre que satisface las tres propiedades de una medida de probabilidad (si el primer intento de definición no las satisface, modificar la definición para que sí lo haga).
   (c) La probabilidad de que el juego dure más de $n+2$ lanzamientos es $qp^{n+1} + pq^{n+1}$.
   (d) Si $p = 0$ o $p = 1$ habrá sólo dos tiradas, y en otro caso el número esperado de tiradas es siempre 3. ¿El resultado en este último caso depende del valor de p? [Rta: no].
   (e) Muestre que si bien la duración promedio del juego es siempre 3 tiradas (para todo valor de p, salvo p=0 o p=1), una moneda pareja (p=0.5) maximiza la probabilidad que el juego dure más de 4 tiradas.

5. Un jugador participa en el juego del ejercicio anterior apostando plata. La regla es la siguiente:
   (a) si el primer resultado es ceca (S), gana \$1 por cada cara (C) que salga hasta volver a obtener ceca. Si el primer resultado es cara, gana \$1 por cada ceca subsiguiente (por ejemplo, con CSSSSC gana \$4, con CC no gana nada). ¿Cuánta plata espera cobrar el jugador por participar? [Rta: si $p = 0$ o $p = 1$, no gana nada; gana \$1 en otro caso]
   (b) si ahora el juego paga \$1 por cada cara intermedia y \$2 por cada ceca intermedia ¿Cuánta plata espera ganar el jugador? [Rta: si $p = 0$ o $p = 1$, no gana nada; gana (\$2 - p×\$1) en otro caso]
   (c) el juego paga \$1 por la cantidad de caras totales que salieron ¿Cuánto espera ganar en este caso? [Rta: no gana si $p = 0$, \$2 si $p = 1$, y gana (p×\$3) en otro caso]

   (Sugerencia: le conviene calcular $E(X)$, siendo X la variable aleatoria monto ganado)

6. Al lanzarse un par de dados se espera que en promedio salga un doble seis cada 36 tiradas. Si se hacen sólo 25 tiradas, ¿conviene apostar a favor o en contra de la aparición de al menos un doble seis? [Rta: a favor, P=0.5055]

   Nota histórica: este problema se conoce como "El problema de los dados del caballero de Méré". Según consta en la carta que Pascal envió a Fermat el día 29 de julio de 1654, el problema había sido planteado a Pascal por su amigo Antoine Gombaud, un escritor y matemático aficionado quien se hacía llamar *Chevalier de Méré* y frecuentaba los salones literarios franceses. Gombaud no lograba entender cómo es posible que, siendo que conviene apostar a sacar al menos un seis tirando 4 veces un dado, no conviene apostar a sacar al menos un doble seis tirando 24 veces un par de dados (dado que 4 es a 6 como 24 es a 36). Varios "pioneros" de la teoría de probabilidades atacaron este problema, de formas distintas y más o menos acertadas.

7. Para determinar si un autroproclamado adivino es un trucho se lo somete a un test que consiste en disponer boca abajo sobre la mesa diez cartas de truco, 5 oros y 5 bastos, y pedirle que adivine cuáles son los cinco bastos. Encuentre la probabilidad de que acierte tres o más bastos por pura casualidad. [Rta: 0.5].

8. El 1% de una población tiene planeado votar un cierto candidato **A**. Se realiza una encuesta tomando $n = 500$ habitantes al azar. ¿Cuál es la probabilidad de que el resultado de intención de voto para **A** sea mayor o igual que el 2%? Dicho de otra manera, ¿cuál es la probabilidad de que la encuesta se equivoque por más del 100%?
   (a) Si el número de habitantes de la población es muchísimo mayor que n. [Rta: 0.0311]
   (b) Si hay $N = 5000$ habitantes en la población. [Rta: 0.0239]
   (c) ¿Cuántos habitantes N debería haber en la población para que los resultados entre (a) y (b) tuvieran una diferencia relativa menor al 0.1%? [Rta: $N \simeq 1.15 \times 10^6$]

9. Una fábrica produce integrados, de los cuales el 20% son defectuosos y los comercializa en cajas de 10. Un comprador quiere rechazar las cajas que contengan más de 2 chips defectuosos, es decir, más que la especificación del fabricante. Para ganar tiempo, en vez de probar todos los chips decide implementar el siguiente test. De cada caja toma 6 chips al azar: *(i)* si ninguno es malo, acepta la caja; *(ii)* si uno solo es malo, revisa el resto de la caja; *(iii)* si 2 o más son malos, devuelve la caja al fabricante.
   (a) ¿Qué fracción de las cajas tendrán más de 2 chips defectuosos? [Rta: 0.322]
   (b) ¿En qué fracción de las cajas deberá probar los 10 integrados? [Rta: 0.3932]
   (c) ¿Cuál es la probabilidad de que haya 3 chips malos en una caja aceptada? [Rta: 0.0114]

10. En promedio Messi mete un gol el 18.2% de las veces que patea al arco.
   (a) ¿Cuántas veces por partido debería patear para que la probabilidad de hacer al menos 2 goles sea mayor que el 90%? [Rta: 20 veces o más]
   (Sugerencia: explorá numéricamente la solución de la ecuación trascendente)
   (b) ¿Cuál es el número esperado de goles que haría si patea la cantidad de veces del ítem anterior? [Rta: 3.64 goles...como si existiera un gol fraccional!]
   (c) Si el técnico decide que va a sacarlo justo después de hacer el segundo gol, ¿cuál es el número esperado de veces que pateó? [Rta: 11 veces]

11. Una fuente radiactiva tiene una actividad de 4 Bq (Becquerel = 1 decaimiento s^{-1}). Calcule la probabilidad de observar al menos un decaimiento en (a) 1 segundo, (b) 2 segundos. [Rta: 0.9817 y 0.9997]

12. Para determinar la concentración c de glóbulos blancos en sangre de un paciente, se toma una muestra de $1\,\mathrm{mm}^3$ de sangre y se cuentan los glóbulos presentes.
   (a) Si, en promedio, se obtiene que $c = 4.05\,\mathrm{nL}^{-1}$, ¿cuál sería el error porcentual que se espera obtener en la medición? ¿Qué hipótesis emplea? [Rta: 1.6%]
   (b) Resulta ser que la muestra que se observa en el microscopio no es directamente la muestra extraída del paciente, sino que esta se diluye 10 veces para hacerla translúcida, y de esa nueva solución se extrae $1\,\mathrm{mm}^3$ para examinar bajo el microscopio. ¿Cuál es entonces el error porcentual correcto para c? ¿Siguen siendo válidas las hipótesis del inciso previo? [Rta: 4.9%]

13. Un experimento detecta en promedio dos neutrinos por día que son una señal de *fondo**.
   (a) El día en que telescopios en la superficie observaron la explosión de una supernova, el detector observó 8 neutrinos. ¿Cuál es la probabilidad de que se detecten 8 o más neutrinos debido a una fluctuación del fondo? Se puede publicar entonces que hay evidencia experimental de que las supernovas emiten neutrinos? [Rta: 0.0011]
   (b) Consideremos ahora que no hubo observación óptica, que se tomaron datos durante todo 1987, y que cierto día se observan 8 señales de neutrinos. ¿Cuál es la probabilidad de observar 8 o más neutrinos en al menos un día del año? Se puede concluir que por algún fenómeno cosmológico ese día aumentó el flujo de neutrinos sobre la Tierra? [Rta: 0.33]
   Analice conceptualmente la diferencia entre este caso y el anterior.

   Nota histórica: el detector gigante Irvine-Michigan-Brookhaven (IMB) que opera en una mina subterranea abandonada fue originalmente diseñado para medir el decaimiento del protón. El 23 de Febrero de 1987 el IMB detectó 8 neutrinos compatibles en tiempo con la explosión de la supernova SN 1987A ubicada a 168000 años luz de la Tierra. IMB jamás detectó el decaimiento de ningún protón.

14. Una fuente de luz emite en promedio 100 fotones s^{-1}. Durante cuánto tiempo deberá contarse fotones si se desea conocer la intensidad del haz con una precisión de $^1/_{1000}$. [Rta: 2.77 hs]
   Nota: use como 'precisión' el cociente $\sqrt{varianza}/media$.

15. Una astrónoma quiere mejorar una antigua determinación del período de un pulsar $\tau \pm \sigma_\tau/\sqrt{n}$, hecha a partir de $n = 25$ mediciones exitosas. Se sabe que por fluctuaciones en la polución de la atmósfera el 20% de las pulsaciones deberán ser descartadas luego de evaluar su calidad. Si la astrónoma quiere que la nueva determinación de τ sea el doble de precisa que la anterior,
   (a) ¿Cuántas observaciones exitosas (no afectadas por la atmósfera) son necesarias? [Rta: $n = 100$]
   (b) ¿Cuál es la probabilidad de que necesite realizar 130 o más observaciones en total (tanto exitosas como afectadas) para alcanzar el grado de precisión deseada? [Rta: $p = 0.206$]
   (c) ¿Cuántas observaciones totales debería planificar si quiere que la probabilidad de obtener la precisión deseada (o una incluso mejor) sea mayor al 99%? [Rta: por lo menos 139 mediciones]

*Se entiende por fondo a los eventos medidos que no provienen del proceso físico que se quiere estudiar. Por ejemplo eventos provenientes de decaimientos radiactivos en las cercanías del detector.

16. En cada uno de los siguientes casos, escriba la expresión (no se pide la cuenta!) que habría que mostrar para confirmar la afirmación hecha en itálica. Analice el ejemplo que ilustra cada caso.

   (a) *Suma de poissonianas es poissoniana.* Una fuente radiactiva emite λ partículas por unidad de tiempo. Si se ubica dicha fuente frente a un detector en presencia de un fondo radiactivo de μ emisiones por unidad de tiempo, el detector contará indistintamente las partículas del fondo y de la fuente. Si el número de partículas emitidas en un tiempo t por la fuente (i) y el fondo (j) tienen distribuciones poissonianas $P_i(\lambda t)$ y $P_j(\mu t)$ respectivamente, entonces: el número total de cuentas en el detector tiene distribución poissoniana con constante $(\lambda + \mu)t$.
   [Rta: Para observar n cuentas es necesario que la fuente haya emitido k partículas y el fondo $n-k$, con $0 \leq k \leq n$. Esto es, $P_n = \sum_{k=0}^{n} P_k(\lambda t) P_{n-k}(\mu t)$]

   (b) *Composición de poissoniana con una binomial es poissoniana.* Un detector tiene una eficiencia $\epsilon \leq 1$ para registrar partículas. Si le llegan n partículas, la probabilidad de detectar k será entonces la binomial $B_k(n, \epsilon)$. Si se coloca frente a él una fuente radiactiva que en promedio emite λ partículas en un tiempo t, entonces: el número de registros r obtenidos por el detector en un tiempo t, tiene una distribución poissoniana con constante $\epsilon\lambda t$.

   Nota: este problema describe la práctica de nuclear de Laboratorio 5, donde se obtiene un gráfico de Poisson midiendo una fuente radiactiva (poissoniana) con un detector que no tiene eficiencia 100% (binomial).

   (c) *Composición de binomial con hipergeométrica es binomial.* Una población tiene dos clases de elementos, los del tipo 'a' con probabilidad p y los del tipo 'b' con probabilidad $(1-p)$. Se selecciona una primera muestra de N_1 elementos. Luego se selecciona una submuestra de N_2 $(\leq N_1)$ elementos tomados al azar de la primera muestra. Es claro que la probabilidad de encontrar k_1 elementos de la clase 'a' en la primera muestra es una binomial $B(k_1; N_1, p)$. Bajo estas condiciones: la probabilidad de encontrar k elementos del tipo 'a' en la submuestra (de tamaño N_2) también es una binomial dada por $B(k; N_2, p)$. Es decir, que en estas condiciones tomar una muestra al azar a partir de otra muestra al azar es equivalente a haber tomado directamente una única muestra de tamaño N_2.
   [Rta: Cuando se observan k elementos en la segunda muestra, la primera muestra pudo tener $k_1 \geq k$ elementos. O sea basta con probar que:
   $Binomial(k; N_2, p) = \sum_{k_1=0}^{N_1-N_2+k} Hipergeometrica(k|N_1, N_2, k_1) Binomial(k_1|N_1, p)$.
   El límite superior de la suma proviene de la condición: $N_2 - k \leq N_1 - k_1$.]

Fórmulas útiles

Distribución	Fórmula	Esperanza	Varianza
Binomial	$B(X=k\|n,p) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$
Poisson	$P(X=k\|\lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$	λ	λ
Hipergeométrica	$H(X=k\|N,n,K) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$	$\frac{nK}{N}$	$n\frac{K(N-K)}{N^2}\left(\frac{N-n}{N-1}\right)$
Binomial negativa	$P(X=n\|k,p) = \binom{n-1}{k-1} p^k (1-p)^{n-k}$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$

La suma parcial $\sum_{n=0}^{N} p^n = \frac{p^{N+1}-1}{p-1}$, converge a una serie geométrica converge si $p < 1$ (y diverge en otro caso)

$$\sum_{n=0}^{\infty} p^n = \frac{1}{1-p}$$

La serie aritmético geométrica vale: $\sum_{n=0}^{\infty} (an+b)\, r^n = \frac{b}{1-r} + \frac{ar}{(1-r)^2}$.
La serie

$$\sum_{n=0}^{\infty} (an+b)\, p^n = \frac{b}{1-p} + \frac{ap}{(1-p)^2}$$

converge si $p < 1$ (y diverge en otro caso).

Ejercicio computacional

Simulación de la medición de una fuente luminosa con un detector de eficiencia $\epsilon < 1$. Realice la siguiente simulación, justificando cada uno de los pasos.

1. La mayor parte de los lenguajes de programación y muchos programas con orientación científica incluyen generadores de números pseudoaleatorios con distribución uniforme en $[0, 1)$. Cualquier conjunto de dichos números constituye una muestra de una variable aleatoria continua (tema de la guía siguiente) con dicha distribución, pero aquí nos interesa solamente el hecho de que pueden ser usados para simular experimentos de Bernoulli. Mostrar computacionalmente que la probabilidad de que tal generador entregue un número en $[0, p)$ $(0 \leq p \leq 1)$ es p (y en $[p, 1)$ es $1 - p$). Implemente, a partir de esta propiedad, un programa que genere el resultado de una serie de n experimentos de Bernoulli independientes y cuente el número de éxitos obtenido. Considere n y la probabilidad de éxito p como parámetros libres.

2. Suponga que inciden exactamente $n = 15$ fotones sobre un detector con una eficiencia $\epsilon = 0.75$. Use el programa escrito en el punto anterior para determinar cuántos fotones son detectados en un experimento particular. Repita el experimento 1000 veces y realice un histograma de los resultados (i.e., del número de fotones detectados en cada experimento). Compárelo con la distribución de probabilidad teórica para dicha variable aleatoria. No olvide normalizar correctamente el histograma para realizar la comparación.

3. Ahora considere una fuente de intensidad media $I = 15$ fot s^{-1}. Simule el número de fotones emitidos por la fuente en $\Delta t = 1$ s del siguiente modo. Subdivida el intervalo Δt en $m \gg 1$ subintervalos iguales dt (sugerencia: use $m = 1000$). Aproxime la probabilidad de que la fuente emita un fotón en dt como Idt, y desprecie la probabilidad de emitir más de 1 fotón en el mismo intervalo (¿cómo justifica estas hipótesis?). Simule entonces el número de fotones emitidos en cada dt usando el programa desarrollado en el punto (a) (¿por qué es válido usar un programa que simula experimentos de Bernoulli?), y sumando sobre todos los dt calcule el número total de fotones emitidos durante Δt. Repita el experimento 1000 veces, construya un histograma de los resultados y superponga a éste la distribución teórica correspondiente (no olvide normalizar). Discuta el procedimiento y los resultados en base a las hipótesis del proceso de Poisson.

4. Dado el resultado de cada uno de los 1000 experimentos del punto anterior, use el mismo programa para calcular el número de fotones *detectados* por el detector del punto (b), suponiendo que éste opera durante $\Delta t = 1$s, y que todos los fotones emitidos llegan a él (i.e., que el detector subtiende un ángulo sólido de 4π visto desde la fuente). Realice el histograma correspondiente y compárelo con la distribución teórica (no olvide normalizar). Discuta los resultados (vea el ejercicio 16b).

5. Para ahorrar tiempo, usted podría haber considerado la emisión y detección de cada fotón en forma conjunta, suponiendo una probabilidad *efectiva* (¿de qué valor?) para este proceso (¿cómo justifica esta hipótesis?). Realice la simulación de este modo, grafique los histogramas y sus correspondientes distribuciones teóricas. Muestre que el resultado es el mismo que el del punto anterior (en sentido estadístico). Discuta el procedimiento y los resultados a la luz de la composición de un proceso de Poisson con uno de Bernoulli.

6. ¿Qué distribución espera para que el número de datos en una determinada clase (bin) de un histograma y por qué? Use la desviación estándar como una estimación de la incerteza de dicha variable y grafique barras de error sobre **todos** los histogramas realizados. Discuta el sentido en que deben interpretarse las posibles discrepancias entre las distribuciones de probabilidad teóricas y los histogramas.