Probabilidad

Tomás Ricardo Basile Álvarez
316617194

10 de diciembre de 2020

1. Probabilidad Elemental

Definición 1.1. Experimento Aleatorio: Un experimento repetible y que da resultados variables del espacio muestral.
Espacio Muestral: Ω Todos los posibles resultados de un experimento.
Probabilidad Clásica: La probabilidad de que suceda un evento en $A \subset \Omega$ es $P(A) = \frac{\#A}{\#\Omega}$
(Siempre y cuando todos los eventos sean equiprobables)

Definición 1.2. Probabilidad Axiomática : La probabilidad es una función $P : \wp(\Omega) \longrightarrow [0,1]$ tal que:
1) $P(A) \geq 0$ 2) $P(\Omega) = 1$ 3) $P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$ (Con A_k disjuntos)

Teorema 1.1. Propiedades de la probabilidad:
1) $P(\emptyset) = 0$
2) $P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$ (Con A_k disjuntos)
3) $P(A^c) = 1 - P(A)$
4)$A \subset B \Rightarrow P(A) \leq P(B)$
5)$A \subset B \Rightarrow P(B - A) = P(B) - P(A)$
6) Si A es un evento $\Rightarrow$ $0 \leq P(A) \leq 1$
7) Para cualquier A y B tenemos que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Definición 1.3. Álgebra: Es una colección $F \subset \wp(\Omega)$ con:
1) $\Omega \in F$ 2) Si $A \in F$ $\Rightarrow A^c \in F$

3) La unión finita de elementos de F está en F.

Sigma Álgebra: Igual que un álgebra pero con uniones numerables.
Sigma Álgebra de Borel: Es la Sigma álgebra de $\mathbb{R}$ más chica posible que contiene a $\{(-\infty, x] : x \in \mathbb{R}\}\ \forall x \in \mathbb{R}$

Con estos tres elementos: Un Espacio Muestral (posibles resultados de los experimentos), una función de probabilidad (que a varios subconjuntos de Ω les asigna una probabilidad entre 0 y 1 cumpliendo los 3 axiomas) y una σ- álgebra de Ω que consiste de los subconjuntos de Ω que son 'medibles'.

2. Conteo y Combinatoria

Teorema de Multiplicación: Si un evento A se puede realizar de n formas distitnas y un evento B de m formas distintas, entonces ambos eventos a la vez se pueden realizar de nm formas distintas.

Ordenaciones Y Combinaciones:

1) **Ordenación con Repetición:** De cuántas maneras puedo escoger una cadena ordenada de k elementos de entre n elementos totales, repitiendo. Cuántas cadenas de $K = 10$ letras $n = 26$ hay?
n^k

2) Ordenaciones sin Repetición: De cúantas formas puedo formar una cadena ordenada de k elementos distintos de un conjunto de n elementos. Cuáles son los posibles resultados del podio $k = 3$ de una carrera de n coches (dichos en orden)
$\frac{n!}{(n-k)!}$

3) Combinación (Sin repetición): Cuantos subconjuntos de longitud k hay de un conjunto de n elementos totales (no me importa el orden). Cuántos equipos de futbol $k = 11$ puedo formar en mi salón $n = 30$.
$\frac{n!}{k!(n-k)!}$

4) Combinaciones Con Repetición: Cuáuntos conjuntos de k elementos puedo formar (si puedo repetir) a partir de un conjunto de n elementos. Cuántos helados distintos puedo hacer con 3 bolas de helado. De cuántas maneras puedo poner 4 pelotas en 10 cajas (tengo que escoger 4 de las diez cajas y puedo repetir). Estrellas y barras:
$\binom{n+k-1}{k}$

Con estás operaciones, podemos calcular la probabilidad de muchos eventos, sólo hace falta contar la cantidad de casos favorables y dividir entre la cantidad de casos totales.

3. Probabilidad Condicional:

Nos interesa ver la probabilidad de un evento A dado que ya sucedió otro evento B anteriormente.

Probabilidad de A dado que sucedió B: $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$
Intuición: Como sucedió B, podemos reducir el espacio muestral a B y entonces el evento A se reduce a $A \cap B$.

De hecho, si dejamos B fijo y definimos $P_*(A) = P(A|B) = \dfrac{P(A \cap B)}{P(B)}$ entonces esta P_* es una función de probabilidad (cumple los 3 axiomas de Kolmogorov)

Teorema 3.1.
Teorema de Probabilidad Total: Tomamos una partición disjunta de Ω $\{B_1, ..., B_n\}$. Entonces, se cumple que:

$$P(A) = \sum P(A|B_i)P(B_i) \quad \forall A$$

Viendólo como conjuntos y recordando la def. de probabilidad condicional, queda claro.

Un ejemplo particularmente común es cuando la partición es $\{B, B^c\}$, en este caso:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$
$$P(A) = P(A \cap B) + P(A \cap B^c)$$

Teorema 3.2.
Teorema de Bayes: Dada $\{B_1, ...B_2\}$ una partición de Ω y A un evento, tenemos que:

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum P(A|B_i)P(B_i)}$$

O, en particular: $P(B|A) = \dfrac{P(A|B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$

4. Variable Aleatoria:

Los resultados de un experimento aleatorio se encuentran en el espacio muestral. Sin embargo, sería mejor tener números en vez de estos datos abstractos. Para eso, definimos una **Variable Aleatoria** que es una función que toma elementos del espacio muestral y los convierte en reales, $X : \Omega \to \mathbb{R}$.

Dado un Ω podemos usar varias variables aleatorias dependiendo de cuál sea el dato que nos interese. Por ejemplo, si el espacio muestral es un dardo cayendo en un tablero, varias posibles variables aleatorias son: 1) tomar la coordenada x del dardo , 2) tomar la coordenada y , 3) tomar la distancia del dardo al centro , 4) tomar el ángulo del dardo con respecto al eje x, 5) cualquier otra cosa.
Estos ejemplos nos muestran como X es simplemente una forma de asociarle números a Ω. Claro que las probabilidades como tal sólo las conocemos para el conjunto Ω (es el conjunto dominio de la función de probabilidad). Esto nos permite mover la responsabilidad del cálculo de probabilidad para en vez de medirlo en Ω lo medimos en los reales.
Entonces, para calcular la probabilidad de un evento depueés de aplicar la variable aleatoria, necesitamos calcular la probabilidad de su preimagen (que está en Ω)

Definición 4.1.
Preimagen: El conjunto de eventos de Ω tales que aplicarles X da como resultado un valor entre a y b se denota como $a \leq X \leq b$. Es decir, si quiero calcular la probabilidad de que un evento (tras aplicar la variable aleatoria) caiga entre a y b, sólo me fijo en todos los elementos del espacio muestral que lo llevan ahí y calculo su probabilidad.

Para espacios muestrales discretos (finitos)

PMF (Probability Mass Funcion): $P_X(x) = P(X = x)$
O más generalmente, para $A \subset \mathbb{R}$, $\quad P_X(A) = P(X^{-1}(A))$
Donde P es la función de probabilidad de Ω (Dado un evento de Ω, le asigna una proba). Y entonces P_X es una función de probabilidad de los reales (es decir, ahora, a los reales les asigna una probabilidad).

Propiedades: 1) $P(A) \geq 0$ 2) $P(\mathbb{R}) = 1$ 3) $P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$
(Con A_k disjuntos) Es decir, es una función de probabilidad (de los reales)

PMF da la probabilidad de que seleccionemos un evento de Ω y tras aplicar la variable aleatoria, caiga en x.

CDF (Cumulative Distribution Function): $F_X(x) = P(X \leq x)$
Es decir, nos da el valor de la probabilidad de que obtengamos un elemento de Ω tal que la variable aleatoria lo manda a un real menor que x.

Propiedades: 1) $F_X(x) \to 0$ si $x \to -\infty$ 2)$F_X(x) \to 1$ si $x \to \infty$

Nota: A una variable aleatoria X la podemos componer con una función $g : \mathbb{R} \to \mathbb{R}$ para obtener una nueva variable aleatoria. Es decir, toma un resultado del espacio muestral, le aplica X y luego a eso le aplica g para obtener finalmente un real. Por ejemplo, para el dardo, digamos que X nos da la distancia al origen y g nos da el cuadrado, entonces la variable aleatoria $g \circ X$ nos da la distancia al cuadrado.

Definición 4.2.
Esperanza: La esperanza de una variable aleatoria $X : \Omega \to \{x_1, ...x_n\}$ es:

$$E(X) = \sum_i x_i P(X = x_i)$$

Es decir, es como un promedio ponderado de los valores de x_i (porque los multiplicamos por la probabilidad de que salgan). Es como el centro de masa de los puntos x_i donde $P(X = x_i)$ es la masa de cada uno de los puntos.

Satisface: 1)Linealidad: $E(aX + bY + c) = aE(X) + bE(Y) + c$
2) $E(g \circ X) = \sum_i g(x_i) P_X(X = x_i)$
Esto es porque da el centro de masa de los puntos. La masa de cada punto es la misma, pero ahora están movidos por g.

Definición 4.3.
Varianza: La varianza de una variable aleatoria $X : \Omega \to \{x_1, ...x_n\}$ es:

$$V(X) = E((X - E(X))^2) \;=\; E(X^2) - (E(X))^2$$

$E(X)$ es el centro de masa, entonces $(X - E(X))^2$ es la distancia cuadrada de un punto al centro de masa (es decir el momento de inercia de X). En general, usando la definición de esperanza, y definiendo $E(X) = \mu$, tenemos que $V(X) = \sum_i \; P(X = x_i)(x_i - \mu)^2$
Es decir, suma las probabilidades por la distancia cuadrada al centro.

Satisface: 1) $V(X) \geq 0$
2) $Var(c) = 0$
3) $Var(cX) = c^2 Var(X)$
4) No es lineal.
5) $Var(X + a) = Var(X)$

Para espacios muestrales continuos

PDF (Probability Density Funcion): Es una función $f(x) \geq 0$ tal que se cumple que: $P(X \in [a,b]) = \int_a^b f(x)dx$

Es decir, es una función tal que nos permite calcular la probabilidad de que la imagen de X esté en un conjunto $[a,b]$. O bien, $f(x)$ es más o menos una densidad, si queremos la probabilidad de que X caiga en $[a,b]$ (la masa dentro de $[a,b]$) hay que integrar esta densidad entre a y b. Nota: que el espacio muestral sea continuo implica que no tiene sentido cosas como $P(X = x)$

Propiedades: $1) f(x) \geq 0$ $\quad 2) \int_{-\infty}^{\infty} f = 1$

3) Si $Y = g(X)$, entonces, $f_Y(y) = f_X(g^{-1}(y))|\frac{d}{dy}g^{-1}(y)|$

PDF da la densidad de probabilidad de que seleccionemos un evento de Ω y tras aplicar la variable aleatoria, caiga en $[a,b]$.

CDF (Cumulative Distribution Function): $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\,du$

Es decir, nos da el valor de la probabilidad de que obtengamos un elemento de Ω tal que la variable aleatoria lo manda a un real menor que x.

Propiedades: 1) $F_X(x) \to 0$ si $x \to -\infty$ $\quad 2) F_X(x) \to 1$ si $x \to \infty$ $\quad 3) F'(x) = f(x)$

Nota: A una variable aleatoria X la podemos componer con una función $g : \mathbb{R} \to \mathbb{R}$ para obtener una nueva variable aleatoria. Es decir, toma un resultado del espacio muestral, le aplica X y luego a eso le aplica g para obtener finalmente un real. Por ejemplo, para el dardo, digamos que X nos da la distancia al origen y g nos da el cuadrado, entonces la variable aleatoria $g \circ X$ nos da la distancia al cuadrado.

Definición 4.4.

Esperanza: La esperanza de una variable aleatoria $X : \Omega \to \{x_1, ...x_n\}$ es:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Recordando que $E(X)$ es el centro de masa, tiene sentido la operación definida, multiplicamos cada pedacito de masa $f(x)dx$ por la distancia x. Con esto, se obtiene el promedio ponderado o centro de masa.

Satisface: 1)Linealidad: $E(aX + bY + c) = aE(X) + bE(Y) + c$
2) $E(g \circ X) = \int_{\infty}^{\infty} g(x)f(x)dx$
Esto es porque da el centro de masa de los puntos. La masa de cada punto es la misma, pero ahora están movidos por g.

Definición 4.5.
Varianza: La varianza de una variable aleatoria $X : \Omega \to \mathbb{R}$ es:

$$V(X) = E((X - E(X))^2) \; = \; E(X^2) - (E(X))^2$$

$E(X)$ es el centro de masa, entonces $(X - E(X))^2$ es la distancia cuadrada de un punto al centro de masa (es decir el momento de inercia de X). En general, usando la definición de esperanza, y definiendo $E(X) = \mu$, tenemos que $V(X) = \int_{-\infty}^{\infty} f(x)(x_i - \mu)^2 dx$
Es decir, suma las probabilidades por la distancia cuadrada al centro.

Satisface: 1) $V(X) \geq 0$
2) $Var(c) = 0$
3) $Var(cX) = c^2 Var(X)$
4) No es lineal.
5) $Var(X + a) = Var(X)$

Resumen:
Explicaré dos ejemplos a la vez, uno dentro de []. Digamos que hacemos un experimento que consiste en decirle algo a cruz y esperar su respuesta [O en lanzar un dardo a un blanco]. El espacio muestral del experimento son todas las posibles respuestas de Cruz [El espacio muestral son todas las posibles posiciones en las que cae el dardo].

X es la función que convierte dichos posibles resultados en números reales según qué sea lo que queramos medir. Por ejemplo, podemos decir que $X : \Omega \to \mathbb{R}$ sea el número de palabras en la respuesta de Cruz, o a lo mejor podría ser el tiempo que tardó en responder medido en segundos (nos quedaremos con la primera opción) [Algunas posibles variables aleatorias son dar como resultado el puntaje qué generó el dardo o podría ser la distancia del dardo con respecto al centro, nos quedaremos con la segunda]. En cualquier caso, X convierte los eventos de Ω en números reales de alguna forma que nos parezca interesante. Esto nos permite hablar de una 'nueva' probabilidad, ya no es la probabilidad de que suceda cierto evento de Ω sino que es la probabilidad de que obtengamos cierto real (o cierto intervalo de reales) al aplicar X.

Esta es una distribución discreta, por lo que nos interesa su PMF, el PMF es la probabilidad de que X de un cierto valor $x \in \mathbb{R}$ y lo hace midiendo la probabilidad de que suceda un evento de Ω que bajo X se mapee a x. [Esta es una distribución continua, por lo que nos interesa el PDF, que es la densidad de probabilidad).
En cualquier caso, podemos pensar en los posibles resultados del experimento pasados por X (la imagen de X) como puntos [intervalos continuos] sobre la recta real. Podemos pensar en estos puntos como un conjunto de masas puntuales [masas distribuidas] tales que la masa total (probabilidad total) es 1. En este caso, podemos definir la masa de cada punto [podemos definir la densidad en cada punto, tal que al integrarla en un intervalo, nos da su masa]. La masa de un punto (o intervalo) es la probabilidad de que suceda un evento de Ω que tras mapearlo con X caiga en ese punto [o en ese intervalo]

La esperanza es en cierto sentido el promedio de estos números, o bien el centro de masa de la imagen de X (pesada con sus correspondientes probabilidades). En el caso de cruz, nos dará el número de palabras promedio con las que responderá Cruz [La esperanza nos da la distancia promedio del dardo al centro]. Esto es análogo a calcular el centro de masa.

La varianza es una medida de qué tan distribuida está la imagen de X en la recta real. Para esto sumamos las distancias de cada punto con respecto al centro de masa multiplicada por la masa del punto [lo mismo pero integramos]. Esto es equivalente al momento de inercia que se estudia en mecánica.

Momentos: El momento $n-$ ésimo de una variable aleatoria X es:

$\mu_n = E(X^k)$
Así, el primer momento es simplemente la esperanza y la varianza es $V(X) = \mu_2 - \mu_1^2$

Definición 4.6.
Moment Generating Function:
La moment generating function de una variable aleatoria X es la función de t:

$$M_X(t) = E(e^{tX}) = E\left(\sum_{i=0}^{\infty} \frac{(tX)^i}{i!}\right) = \sum_{i=0}^{\infty} \frac{t^i}{i!} E(X^i)$$

Es decir, es un polinomio infinito en la variable t, donde cada uno de los coeficientes son $\frac{E(X^i)}{i!}$ (es decir, los momentos divididos entre su orden factorial). Es decir, este polinomio tiene 'guardada' la información de los momentos en sus coeficientes.
Se puede definir para variables aleatorias discretas o continuas. Es común usar $E(e^{tX}) = \sum x_i e^{tx} P(X = x) \simeq \int e^{xt} f(x) dx$ (para continuas) para obtener la MGF y luego usar lo de los coeficientes para a partir de eso calcular todos los momentos.

Propiedades:
1) $\mu_k = E(X^k) = M_X^{(k)}(0)$ (Lo cual describe el hecho de que M_X guarda información de los momentos)

2) Es única y una MGF determina de forma única a una variable aleatoria.

3) Si X y Y son independientes, entonces: $M_{X+Y}(t) = M_X(t)\ M_Y(t)$

4.1. Para Variables Aleatorias de Dos Dimensiones:

Dado un espacio muestral Ω, podemos definir varias variables aleatorias. Por ejemplo, digamos que nuestro experimento es lanzar un dardo a un tablero, podemos definir las variables aleatorias $X : \Omega \to \mathbb{R}$ que nos da la coordenada x del dardo y $Y : \Omega \to \mathbb{R}$ que nos da la coordenada y del dardo.

Entonces, podemos pensar en una variable aleatoria conjunta, X, Y que a cada evento $\omega \in \Omega$

le asocia el punto $(X(\omega), Y(\omega))$ del plano.

Si las variables aleatorias son discretas, podemos definir una PMF como $P_{X,Y}(x,y) = P(X = x, Y = y)$
Es decir, es la probabilidad de que suceda un evento en Ω que al aplicarle X, Y, sea mandado a (x, y)

Si las variables aleatorias son continuas, entonces también podemos definir una PDF conjunta como la función $f_{X,Y}$ tal que cumple que: $P(X, Y \in [a,b]\times[c,d]) = \int_a^b \int_c^d f_{X,Y}(x,y)\ dydx$.
Es decir, $f_{X,Y}$ mide la densidad de probabilidad en el plano tras aplicar la variable aleatoria X, Y.

Indep.: Con esto, decimos que dos v.a son independientes si se cumple que $f_{X,Y} = f_X \cdot f_Y$

Para estas variables conjuntas, también se puede definir una **CDF** como
$F_{X,Y} = P(X \leq x, y \leq y) = \displaystyle\int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u,v) dv\ du$ (esto último para v.a. continuas)

Esperanza: Nuevamente, la esperanza es el promedio ponderado de las imágenes de X, Y , entonces:
$E(X,Y) = \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dxdy\ (x,y)$ que es el centro de masa.
Propiedad: (para $g : \mathbb{R}^2 \to \mathbb{R}^2$): $E(g(X,Y)) = \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\ f(x,y) dxdy\ (x,y)$

Varianza: La varianza es nuevamente el momento de inercia de la imágen de X, Y, es decir:
$V(X,Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\ \ |(x,y) - E(X,Y)|^2\ dxdy$
porque primero que nada g realocates las masas a $g(x,y)$

Dada una distribución bidimensional, también se pueden definir distribuciones marginales como sigue:
$P(X = x) = \sum_y P(X = x, Y = y)$
$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy$
Es decir, éstas son probabilidades para la variable X, lo único que nos interesa es la masa (o densidad) con respecto al eje x. Por eso, para todo x, definimos su masa (densidad) como la suma de todas las masas (densidades) que tienen como cordenada x el valor x. Como si

proyectáramos la distribución X, Y al eje X.

Covarianza: $COV(X,Y) = E((X-E(X))(Y-E(Y))) = E(XY) - E(X)E(Y)$ (Nota que $E(XY)$ no tiene la coma, es la esperanza de una v.a de una dimensión).

Correlación: Es la covarianza estandarizada: $Corr(X,Y) = Cov(X,Y)/\sqrt{Var(X)Var(Y)}$
Propiedad: Si X e Y son indep. entonces su $Cov = 0$

5. Algunas Distribuciones:

5.1. Distribuciones Discretas:

1) Distribución de Bernouilli:

Se dice que $X \sim Ber(p)$ si $X : \Omega \to \{0,1\}$
Y la variable aleatoria sólo regresa un verdadero o falso. Es útil para cualquier experimento en el que nos interese saber si sucedió algo o no.

PMF: $P(X=1) = p \;\; P(X=0) = 1-p$
Esperanza: $E(X) = p$
Varianza: $V(X) = p(1-p)$
MGF: $M_X(t) = (1-p) + pe^t$
Por ejemplo: 1) $\Omega = \{$ Lanzamientos de un a moneda cargada $\}$, $X : \Omega \to \{0,1\}$, $X(\omega) = 1$ si es cara, 0 si es cruz.
2) $\Omega = \{$ Posibles respuestas de Cruz al decirle algo $\}$, $X : \Omega \to \{0,1\}$ con $X = 1$ si Cruz contesta, $X = 0$ si Cruz no contesta. O bien, $X = 1$ si Cruz dice ehhhhhh, $X = 0$ si no dice ehhh.

La esperanza y la varianza se pueden calcular directamente de las definiciones. o también, se puede calcular el MGF como $M_X(t) = E(e^{tX}) = \sum g(x_i)P(X = x_i) = \sum e^{tx_i}P(X = x_i) = e^{0t}(1-p) + e^{1t}p = (1-p) + pe^t$ Luego calculasmos los momentos a partir de este polinomio.

2) Distribución Binomial:

Se dice que $X \sim Bin(n, p)$ si $X : \Omega \to \{0, 1, ..., n\}$
Y la variable aleatoria X cuenta el número de veces que sucede un evento (que tiene probabilidad p) al realizar n intentos.

PMF: $P(X = k) \ = \ \begin{pmatrix} n \\ k \end{pmatrix} \ \ p^k \ (1-p)^{n-k}$
Esperanza: $E(X) = np$
Varianza: $V(X) = np(1-p)$
MGF: $M_X(t) = ((1-p) + pe^t)^n$
Propiedad: Si $Y \sim Bin(n, p) \ = \ X_1 + ... + X_n$ (donde $X_i \sim Ber(p)$).
Por ejemplo: Repetimos n veces un Bernouilli (Por ejemplo, dados n tiros de una moneda, contar el número de veces que cae cruz ($p = 1/2$). O bien, dado n tiros de dados, contar el número de veces que sale 3 ($p = 1/6$)).

La PMF sale de que suceden k eventos positivos (proba p), luego $n-k$ eventos negativos (proba $1-p$). Finalmente, sabiendo el número de eventos positivos, lo multiplicamos por todas las formas distintas de que suceda dicho evento.
Podemos calcular las cosas directamente con las formulitas. O bien, usamos la propiedad junto con la linealidad de E para calcularlo, o la propiedad de MGF de una suma para calcular el MGF y luego usar sus derivadas para calcular los momentos.

3) Distribución Geométrica:

Se dice que $X \sim Geom(p)$ si $X : \Omega \to \mathbb{N}$
Y la variable aleatoria X cuenta el número de veces que sucede un evento fallido (que tiene probabilidad $1-p$) antes de que suceda un evento exitoso (con proba p)

PMF: $P(X = k) = p \ (1-p)^k$

Esperanza: $E(X) = \dfrac{1-p}{p}$

Varianza: $V(X) = \dfrac{1-p}{p^2}$

MGF: $M_X(t) = \dfrac{p}{1-(1-p)e^t}$

Por ejemplo: Lanzamos una moneda varias veces y contamos el número de caras antes de la primera cruz. O bien, si p es la proba de ganar la lotería, $P(X = k)$ me da la proba de que necesite comprar k boletos fallidos antes de ganar y en promedio esperaría comprar unos $\frac{1-p}{p}$ boletos.

La PMF sale de que necesitamos que sucedan k eventos fallidos, con proba: $(1-p)^k$ y luego suceda uno exitoso (con proba p).
La esperanza se puede calcular haciendo la suma $\sum k(1-p)^k p = p(1-p)\sum k(1-p)^{k-1} = p(1-p)\frac{d}{dp}(-\sum(1-p)^k) = p(1-p)\frac{d}{dp}(-1/p) = \frac{1-p}{p}$
El MGF y la varianza se puede calcular de forma similar.

4) Distribución Negativa Binomial:

Se dice que $X \sim NegBin(r,p)$ si $X : \Omega \to \mathbb{N}$
Y la variable aleatoria X cuenta el número de veces que sucede un evento fallido (que tiene probabilidad $1-p$) antes de que suceda el r-ésimo éxito (con probabilidad p).

PMF: $P(X = k) = \begin{pmatrix} r+k-1 \\ r-1 \end{pmatrix} p^r\ (1-p)^k$

Esperanza: $E(X) = \dfrac{r(1-p)}{p}$

Varianza: $V(X) = \dfrac{r(1-p)}{p^2}$

MGF: $\left(\dfrac{p}{1-(1-p)e^t}\right)^r$

Propiedad: Si $Y \sim NegBin(r,p) = X_1 + ... + X_r$ (donde $X_i \sim Geom(p)$).
2)$NegBin(1,p) \sim Geom(p)$

Por ejemplo: Queremos calcular el número de veces que perderemos la lotería antes de ganarla r veces. O bien, cuántas veces debo de perder a los dados antes de hacer r generalas.

La propiedad 1 sale directamente de la definición de la distribución y nos es muy útil para calcular las cosas.
La PMF nos dice la probabilidad de que el resultado de X sea k (la probabilidad de que se requieran k fallos antes del r-ésimo éxito). Para esto, k fallos tiene la proba $(1-p)^k$ y r éxitos la forma p^r. Luego, para calcular el número de maneras distintas, podemos pensar que tenemos r cajas y k pelotas, donde la i-ésima caja representa los intentos que sucedieron antes del i-ésimo éxito pero después del (i-1)-ésimo. Luego, tenemos k pelotas que representan los fallos. El problema entonces se reduce a la forma de colocar estas k pelotas en las r cajas, que como sabemos es una combinación sin repetición y se calcula como dice ahí.

Para calcular las cosas, podemos calcular la MGF usando la propiedad 1 y a partir de la MGF de la distribución geom. Con la MGF podemos calcular ya todos los momentos y así la esperanza y la varianza.
Sino, la esperanza se puede calcular con la misma propiedad y con linealidad de E.

5) Hipergeométrica:

Se dice que $X \sim HipGeom(w, b, n)$ si $X : \Omega \to \{0, 1, ..., min(w, n)\}$
Dada una población de w buenos objetos y b malos, X cuenta el número de veces que obtenemos un elemento tipo b al realizar n tomas SIN reemplazo ($n \leq w + b$ para que tenga sentido y $k \leq w, n$).

PMF: $P(X = k) = \binom{w}{k}\binom{b}{n-k} / \binom{w+b}{n}$
Esperanza: $E(X) = \dfrac{nw}{b+w}$
Varianza: $\dfrac{w+b-n}{w+b-1}\dfrac{\mu}{n}\left(1-\dfrac{\mu}{n}\right)$
MGF: Fea expresión

Ejemplos: 1) Al tomar 5 (n) cartas de un conjunto de 52 cartas, la probabilidad de que k de ellas sean aces $w = 4$, $b = 48$.
2) En un bosque con N alces, n de ellos están marcados. Capturamos m alces y queremos

ver cómo se distribuye la probabilidad de que k de ellos ya estubieran marcados. Entonces es $HipGeom(n, N-n, m)$

La PMF sale de que calculamos el número de formas de obtener k objetos buenos (con el w choose k) y lo multiplicamos por el número de formas de que el resto de los objetos sean malos (b choose $n-k$) y finalmente dividimos entre el número total de formas de tomar n elementos.
Podemos calcular las cosas directamente con las formulitas, realizando sumas complicadas. Sin embargo, la esperanza tiene sentido, pues al tomar 100 objetos, nos imaginamos que la en promedio algo así como $w/b + w * 100$ serán buenos (la proporción de objetos buenos del total)

6) Poisson

Se dice que $X \sim Pois(\lambda)$ si $X : \Omega \to \mathbb{N}$
Tenemos eventos raros que suceden a un ritmo de λ veces por unidad de tiempo, $X : \Omega \to \mathbb{N}$ nos dice el número de veces que sucedió este tipo de evento en este intervalo de tiempo

PMF: $P(X = k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$
Esperanza: λ
Varianza: λ
MGF: $e^{\lambda(e^t-1)}$
Propiedad: 1) Si $X \sim Pois(\lambda_1)$ y $Y \sim Pois(\lambda_2)$ con X y Y indep. entonces $X + Y \sim Pois(\lambda_1 + \lambda_2)$
2) Es como $Bin(n, p)$ pero con $p << 1$ y $n >> 1$

Por ejemplo: λ es el número de errores tipográficos en una página de texto y queremos calcular la distribución de proba de tener k errores en una pag.
2) Hay 2 accidentes por mes en una calle, entonces el número de accidentes en un mes sigue $Pois(2)$ y en dos meses $Pois(4)$.

Las propiedades 1 y 2 están claras de la definición. El PMF medio tiene sentido. La esperanza tiene mucho sentido y se puede calcular directo del PMF o bien intuitivamente tiene sentido.

6. Algunas Distribuciones Continuas

1) Uniforme

Decimos que $X \sim Unif(a, b)$ si $X : \Omega \to [a, b]$

f da el mismo valor de densidad de probabilidad para todo $x \in X(\Omega)$ y tiene valor $\frac{1}{b-a}$ (Todo es equiprobable)

PDF: $f(x) = \dfrac{1}{b-a} \;\; \forall\, x \in [a, b]$

Esperanza: $E(X) = \dfrac{a+b}{2}$

Varianza: $V(X) = \dfrac{(b-a)^2}{12}$

MGF: $M_X(t) = \dfrac{e^{tb} - e^{ta}}{t(b-a)}$

Por ejemplo: Un random number generator entre a y b.

La esperanza, varianza y MGF se puede calcular directamente de las fórmulas.

2) Distribución Normal: Decimos que $X \sim N(\mu, \sigma^2)$ si $X : \Omega \to \mathbb{R}$

Que tiene la siguiente PDF:

PDF: $f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{\dfrac{-(x-\mu)^2}{2\sigma^2}}$

Esperanza: $E(X) = \mu$

Varianza: $V(X) = \sigma^2$

MGF: $M_X(t) = e^{t\mu + \dfrac{\sigma^2 t^2}{2}}$

Estandard: La distribución Estandard tiene centro 0 y var 1.

Ejemplo: La distribución de los puntajes IQ. $\Omega = \{$ personas en el mundo $\}, X(\omega) =$ punraje IQ de ω. Y además, la densidad de probabilidad es la f mencionada, con $\mu = 100, \sigma = 15$

Relacuón con binomial:

$$f(x) = \binom{n}{x} p^x q^{n-x} \sim \frac{1}{\sqrt{2\pi npq}} e^{-(x-np)^2/(2npq)}$$

Que es una binomial con $\mu = np$, $\sigma = \sqrt{npq}$

Distribución Normal Stándar: Es aquélla con $\mu = 0, \sigma = 1$. Por tanto, tiene pdf:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

3) Exponencial: Decimos que $X \sim Expo(\lambda)$ si $X : \Omega \to \mathbb{R}$
Que tiene la siguiente PDF:
PDF: $f(x) = \lambda e^{-\lambda x}$
Esperanza: $E(X) = \dfrac{1}{\lambda}$
Varianza: $V(X) = \dfrac{1}{\lambda^2}$
MGF: $M_X(t) = \dfrac{\lambda}{\lambda - t}$
Ejemplo: Es parecido a la distribución poisson. Por ejemplo, si una estrella fugaz se presenta $\lambda = 4$ veces por hora, entonces el tiempo esperado para ver uno es $\dfrac{1}{\lambda}$ horas. $\Omega = \{$ Noches viendo las estrellas $\}$ $X : \Omega \to \mathbb{R}$ es el tiempo antes de que se vea la primera estrella fugaz. Y entonces $f(x)$ es la densidad de probabilidad de ver la primera estrella en un tiempo x.

Dado $f(x)$, podemos calcular $M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \lambda e^{-\lambda x}\ dx = \frac{\lambda}{\lambda - t}$. Con esto podemos calcular los momentos, y así tenemos E y V.

Propiedades: No tiene memoria. $P(X < s + t | X > s) = P(X > t)$

4) Gamma Decimos que $X \sim Gamma(a, \lambda)$ si $X : \Omega \to \mathbb{R}$
Que tiene la siguiente PDF:
PDF: $f(x) = \dfrac{1}{\Gamma(a)} (\lambda x)^a\ e^{-\lambda x} \dfrac{1}{x}$
Esperanza: $E(X) = \dfrac{a}{\lambda}$

Varianza: $V(X) = \dfrac{a}{\lambda^2}$

MGF: $M_X(t) = \left(\dfrac{\lambda}{\lambda - t}\right)^a$

Ejemplo: $\Omega = \{$ noches viendo las estrellas $\}$ y definimos $X : \Omega \to \mathbb{R}$ con $X(\omega) =$ tiempo de espera antes de ver a estrellas. Entonces $f(x)$ da la densidad de probabilidad de que el tiempo de espera sea x.
Es decir, se distribuye como la suma de n v.as. exponenciales. Con eso, podemos calcular fácilmete M_X a partir del exponencial y luego usar las derivadas para calcular los momentos.

Ejemplo: Hay 3 personas adelante en una fila, el promedio de tiempo de espera para que una persona salga de la fila es 2 min. Mi tiempo de espera se distribuirá como $Gamma(3, \frac{1}{2})$. Es decir, con esto podemos calcular la densidad de probabilidad de que el evento suceda a veces en un tiempo x

6.1. Estadística y Medidas Experimentales

Tenemos una secuencia de datos x_i.
No sabemos todos los datos reales, esto es sólo una muestra aleatoria. Entonces queremos aproximar las medidas.

Estimación del Promedio:

$$\mu \simeq \overline{x} = (1/n) \sum_{i=1}^{n} x_i$$

Estimación de la Variancia:

$$\sigma^2 \simeq \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2$$