



Tarea #3

Variable Aleatoria Discreta, Continua y Distribución Multivariada

Ayudante encargado: Adolfo Rojas

Entrega: 08/11/2024

Usted y su amigo, un agrónomo, están trabajando juntos en IntelliPaltas, la start-up que fundaron hace menos de dos años mientras aún cursaban la universidad. La misión de su empresa es innovar en el ámbito de los cultivos hidropónicos, introduciendo el cultivo de paltas mediante este método. Sin embargo, uno de los principales desafíos ha sido la rentabilidad, especialmente considerando que su política desde el inicio ha sido ofrecer solo productos de alta calidad.

Su amigo, experto en la materia, le explica que la calidad del cultivo depende de varios factores, pero que uno de los parámetros físico-químicos más relacionados con la calidad de la palta, y además estadísticamente observable, es la dureza de la fruta. Utilizando un penetrómetro digital, ha medido la dureza de varias paltas, categorizando su calidad según los estándares de la normativa alimentaria chilena. A partir de los datos recogidos, realizó un análisis estadístico de mezcla y concluyó que la dureza de cada palta, independiente del resto, sigue una mixtura de dos distribuciones Normales. Una mixtura de dos Normales es una distribución conformada por dos distribuciones Normales, tal que con una probabilidad entre 0 y 1 la VA distribuye según la primera Normal, y con probabilidad complementaria distribuye según la segunda Normal.

En este caso, la primera distribución corresponde a la dureza de las paltas de alta calidad. Dicha Normal tiene parámetros de media μ_1 y varianza σ_1^2 . Además, la probabilidad de que una palta sea de alta calidad es p , independiente del resto de las paltas. Por otro lado, la segunda distribución Normal representa la dureza de las paltas de baja calidad. Dicha distribución posee una media μ_2 y varianza σ_2^2 . Considere que $\mu_2 < \mu_1$.

Dado que la optimización de las condiciones de cultivo no es su especialidad, su misión será optimizar las utilidades de la empresa determinando el criterio óptimo de dureza a considerar. Para ello, denote a la variable aleatoria C como la variable dicotómica que categoriza la calidad de la palta, y D como la dureza de dicha palta. Note que C es una VA (no un evento), por ende debe tomar valores en los reales (por simplicidad, considere que toma valores 0 o 1).

Para los gráficos de los apartados numéricos de *a.i* y *a.ii*, y la parte *c.iv*; considere $p = 0.7$, $\mu_1 = 30[kgf]$, $\sigma_1 = 8$, $\mu_2 = 15.5[kgf]$, $\sigma_2 = 6$.

- a) i. (1 punto) Determine la distribución de D y grafique
Hint. Usted ya conoce la distribución de D dada la calidad de la palta. Por ende, solo le faltaría encontrar una expresión para su pdf utilizando adecuadamente la Ley de probabilidades totales.

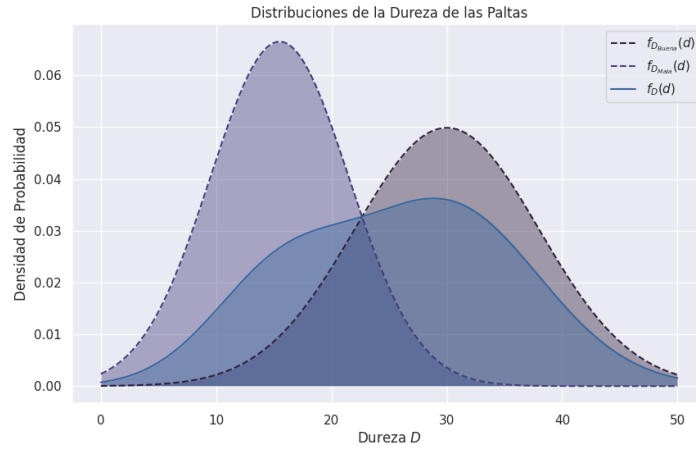
Solución (a.i)

Definimos primero la VA C que toma valor 1 si la palta está buena, y 0 en caso contrario.

Tal como dice la indicación encontramos la pdf de D aplicando la Ley de Probabilidades Totales:

$$\begin{aligned} f_D(d) &= f_{D|C}(d|0) \cdot p_C(0) + f_{D|C}(d|1) \cdot p_C(1) \\ &= f_{D|C}(d|0) \cdot (1-p) + f_{D|C}(d|1) \cdot p \\ &= \frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(d-\mu_2)^2}{2\sigma_2^2}\right) + \frac{p}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(d-\mu_1)^2}{2\sigma_1^2}\right) \end{aligned}$$

Seguido de esto el gráfico de la pdf se encuentra en la figura a continuación



- ii. (0.5 puntos) Utilizando adecuadamente Bayes determine la distribución de probabilidad de la VA C (es decir, la PMF o PDF, según corresponda), dado que la dureza de la palta (D) toma el valor d . ¿Podemos afirmar que la VA $C|D = d$ sigue alguna de las distribuciones vistas en clases? Si es que sí, ¿cuál? Justifique su respuesta claramente. Además, grafique la probabilidad de que una palta esté buena dado que la dureza es d , para diferentes valores de d . Hint: El Teorema de Bayes se puede usar si es que una VA es continua y la otra discreta. Por ejemplo, si X es VA continua, e Y es discreta, entonces se tiene que: (i) $f_{X|Y}(x|y) = \frac{p_{Y|X}(y|x) \cdot f_X(x)}{p_Y(y)}$ para todo $y \in R_Y$ y $p_{Y|X}(y|x) = \frac{f_{X|Y}(x|y) \cdot p_Y(y)}{f_X(x)}$ para todo $x \in R_X$.

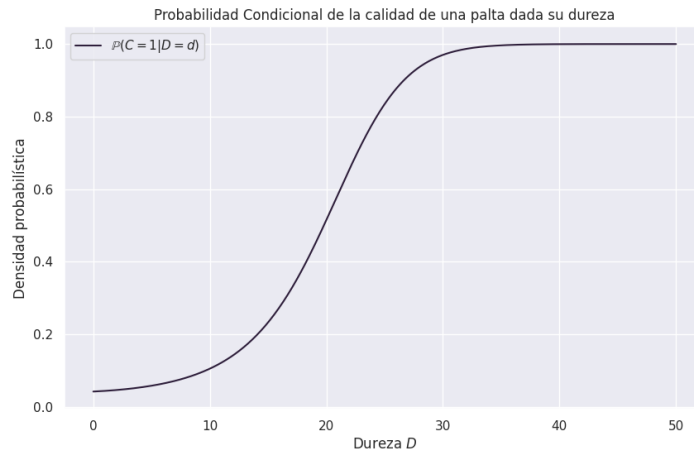
Solución (a.ii)

Tenemos que $p_{C|D}(c|d) = \frac{f_{D|C}(d|c) \cdot p_C(c)}{f_D(d)}$. Luego reemplazando se tiene

$$p_{C|D}(c|d) = \begin{cases} \frac{\frac{p}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(d-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(d-\mu_2)^2}{2\sigma_2^2}\right) + \frac{p}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(d-\mu_1)^2}{2\sigma_1^2}\right)}, & \text{si } c = 1, \\ \frac{\frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(d-\mu_2)^2}{2\sigma_2^2}\right)}{\frac{1-p}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(d-\mu_2)^2}{2\sigma_2^2}\right) + \frac{p}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(d-\mu_1)^2}{2\sigma_1^2}\right)}, & \text{si } c = 0. \end{cases}$$

Con lo que podemos afirmar que $C|D = d \sim \text{Bernoulli}(q)$, donde $q = \mathbb{P}(C = 1|D = d)$.

Finalmente el gráfico pedido es el visto a continuación



- iii. (0.5 punto) Calcule la probabilidad de que la dureza de una palta sea mayor que d . Deje expresado su resultado en términos de la función Φ (CDF de una Normal Estándar)
Hint. Le será útil dejar expresadas las CDFs de las VA $D|C = 0$ y $D|C = 1$, posteriormente encontrar la cdf de D y su complemento

Solución (a.iii)

Primero vemos que $\mathbb{P}(D \geq d) = 1 - F_D(d) = 1 - [(1-p)F_{D|C=0}(d) + pF_{D|C=1}(d)]$ donde dado que $D|C = 0$ y $D|C = 1$ son normales se tiene que sus CDFs pueden expresarse como $\Phi(\frac{d-\mu_2}{\sigma_2})$ y $\Phi(\frac{d-\mu_1}{\sigma_1})$ respectivamente. Finalmente

$$\mathbb{P}(D \geq d) = 1 - [(1-p)\Phi(\frac{d-\mu_2}{\sigma_2}) + p\Phi(\frac{d-\mu_1}{\sigma_1})]$$

Donde esta expresión indica la probabilidad de que una palta sacada al azar tenga una dureza mayor a un valor d

- b) Dado que no es posible conocer la calidad de una palta (a menos que se abra), la empresa considera el umbral de dureza δ , tal que una palta con dureza superior a dicho valor, es *considerada* como de alta calidad. Note que podrían haber casos en donde la calidad *considerada* de una palta no coincide con la calidad real de la palta (esta última dada por la VA C). Por ejemplo, una palta podría ser de alta calidad, pero ser *considerada* de baja calidad. Conteste las siguientes preguntas:
- i. (0.5 puntos) Calcule $\mathbb{P}(C = 1|D \leq \delta)$ e indique qué significa dicha probabilidad. Deje su respuesta en función de Φ .
Indicación. Use el Teorema de Bayes y los cálculos del ítem anterior.

Solución (b.i)

Usando la indicación se tiene que

$$\begin{aligned}\mathbb{P}(C = 1|D \leq \delta) &= \frac{\mathbb{P}(D \leq \delta|C = 1) \cdot p_C(1)}{F_D(\delta)} \\ &= \frac{F_{D|C=1}(\delta) \cdot p_C(1)}{F_D(\delta)} \\ &= \frac{p\Phi\left(\frac{d-\mu_1}{\sigma_1}\right)}{(1-p)\Phi\left(\frac{d-\mu_2}{\sigma_2}\right) + p\Phi\left(\frac{d-\mu_1}{\sigma_1}\right)}\end{aligned}$$

Esta representa la probabilidad de que una palta sea de buena calidad pero que termina siendo descartada dado que cuenta con una dureza inferior a un umbral δ

- ii. (0.5 puntos) Calcule $\mathbb{P}(C = 1|D \geq \delta)$
Hint. El cálculo es muy similar al ítem anterior.

Solución (b.ii)

$$\begin{aligned}\mathbb{P}(C = 1|D \geq \delta) &= \frac{\mathbb{P}(D \geq \delta|C = 1) \cdot p_C(1)}{\mathbb{P}(D \geq \delta)} \\ &= \frac{[1 - F_{D|C=1}(\delta)] \cdot p_C(1)}{1 - F_D(\delta)} \\ &= \frac{p[1 - \Phi\left(\frac{d-\mu_1}{\sigma_1}\right)]}{1 - [(1-p)\Phi\left(\frac{d-\mu_2}{\sigma_2}\right) + p\Phi\left(\frac{d-\mu_1}{\sigma_1}\right)]}\end{aligned}$$

Usted y su amigo venden su producto en cajas a un precio de r [\$/caja], cada una de las cuales contiene n paltas. Estas paltas son empaquetadas por una máquina que se asegura de que cada fruta ingresada tenga una dureza mínima de al menos a . Específicamente, el proceso para llenar una caja por la máquina empaquetadora consiste en tomar una palta, medir su dureza; si su dureza es mayor a a se ingresa a la caja, de lo contrario la palta es descartada para la venta (y por ende no ingresa a la caja). El proceso anterior continua hasta que se tienen n paltas al interior de la caja. Notar entonces que para producir una caja de n paltas es posible que se cosechan más de n paltas.

En su misión de maximizar los beneficios, enfrenta varios costos. Por un lado, incurre en un costo unitario de producción de c [\$] por palta. Además, se tiene que si es que una caja contiene k o más paltas de baja calidad (calidad real, no *considerada*) se debe pagar una multa de m [\$/caja].

- c) Considere las siguientes VAs. M : representa el número de paltas de baja calidad dentro de una caja de n paltas seleccionadas, y F : indica el número de paltas producidas hasta completar una caja de n paltas que cumplen con el umbral de dureza (a):
- i. (0.5 puntos) Determine la distribución de M . Si su distribución es conocida, indique su nombre y los valores de sus parámetros
Hint: Estamos hablando de falsos positivos, notar que para ingresar a la caja es condición necesaria tener una dureza a lo menos a , finalmente si hace uso apropiado de la Ley de Probabilidades Totales, no necesitará introducir ningún cálculo que no haya hecho en ítems previos como los vistos en b)



Solución (c.i)

Usando la definición de la v.a. se sigue que $M \sim \text{Binomial}(n, \mathbb{P}(C = 0|D \geq a))$ donde

$$\begin{aligned}\mathbb{P}(C = 0|D \geq a) &= 1 - \mathbb{P}(C = 1|D \geq a) \\ &= 1 - \frac{p[1 - \Phi(\frac{a-\mu_1}{\sigma_1})]}{1 - [(1-p)\Phi(\frac{a-\mu_2}{\sigma_2}) + p\Phi(\frac{a-\mu_1}{\sigma_1})]}\end{aligned}$$

- ii. (0.5 puntos) Determine la distribución de F . Si su distribución es conocida, indique su nombre y los valores de sus parámetros
Hint. No necesitará introducir ningún cálculo que no haya hecho en ítems previos como los vistos en a

Solución (c.ii)

Usando la definición de la v.a. se sigue que $F \sim \text{Binomial-Negativa}(n, \mathbb{P}(D \geq a))$

- iii. (1 punto) Sea U_a la utilidad de una caja. Esto es los ingresos menos los costos de producción más el potencial costo debido a multas. Entregue una expresión para la utilidad esperada de la venta de **una sola caja** en función de a , o sea $\mathbb{E}[U_a]$.
Indicación. No puede dejar ningún valor esperado en su expresión pero sí dejar en función de probabilidades calculadas en ítems previos tales como $\mathbb{P}(D \geq a)$ y $\mathbb{P}(C = 0|D \geq a)$

Solución (c.iii)

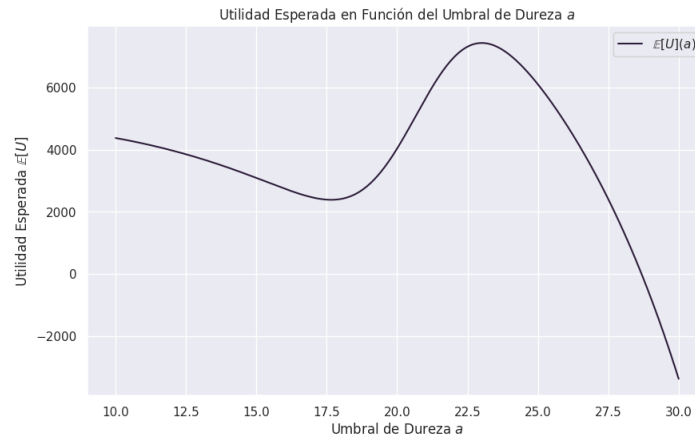
Primero vemos que $U = r - cF - m\mathbb{1}_{\{M \geq k\}}$, luego definiendo por conveniencia a $q := \mathbb{P}(C = 0|D \geq a)$ se tiene que

$$\begin{aligned}\mathbb{E}[U_a] &= r - c\mathbb{E}[F] - m\mathbb{P}(M \geq k) \\ &= r - \frac{cn}{\mathbb{P}(D \geq a)} - m \sum_{i=k}^n \binom{n}{i} q^i (1-q)^{n-i}\end{aligned}$$

- iv. (1 punto) Considere $r = 25000, c = 100, m = 10000, n = 100, k = 8$ y de manera numérica encuentre el valor óptimo de a medido en kilogramo-fuerza que maximiza las utilidades esperadas de una caja. Entregue un gráfico en donde se aprecie que el valor de a entregado es el óptimo. Además, adjunte el código utilizado.

Solución (c.iv)

Del siguiente gráfico se puede ver que el a óptimo se encuentra en el rango $[23, 24]$ cuya aproximación numérica es de $23.55[kgf]$ y ganancia de 6813clp/caja



INSTRUCCIONES

- Se debe trabajar en parejas o de forma individual. Pueden ser de distintas secciones.
- Está prohibido: cualquier tipo de colaboración con personas fuera del grupo, plagio, copia, compartir material usado para llegar a las respuestas, compartir respuestas, la divulgación de esta evaluación o partes de ésta con otras personas ya sea compañeros/as de clase o personas ajenas a la clase. El uso de **herramientas de inteligencia artificial tales como chatGPT o similares no está permitido**, excepto en las partes que requiera programar.
- Genere un reporte que contenga las respuestas a las preguntas de la tarea. Éste puede ser escrito a mano o en computador, pero debe ser entregado en formato PDF. El reporte de las respuestas debe tener una página de portada y como máximo 6 páginas de desarrollo. Además, puede incluir anexos adicionales si lo estima necesario.
- El encargado de esta tarea es el/la auxiliar **Adolfo Rojas**, por lo que las dudas deben ir dirigidas **exclusivamente a él**, de preferencia mediante el foro del curso para aumentar el conocimiento común o alternatively mediante correo.
- Es **responsabilidad de cada uno/a verificar** que los **archivos enviados** incluyan todo lo que se les solicita y en su versión final, recuerde descargar su tarea después de enviada y comprobar que envió lo correcto.
- La fecha de entrega es el día **08/11** a las **23:59** y es de carácter **inamovible**.
- Se aceptarán 2 días de atraso máximo, descontándose x puntos en su nota por día de atraso o fracción.
- Usted dispone de un plazo extra de 2 días a distribuir entre el total de tareas. Ejemplo: puedo usar 1 día extra en la tarea 2 y otro en la 3 (sin descuento).