1 Read me

These notes are extremely limited and sketchy. The reason is that I was familiar with probability theory before taking this class. Thus, I did not need to take many notes. My fellow student will do good in using this document only to corroborate exercises, problems, and final exams.

2 Preliminaries

Let Ω denote the sample space of an experiment; i.e. the set of all values which may result from an experiment. If $A \subseteq \Omega$ we say A is an event. If $\mathcal{A}$ is a σ-algebra over Ω we say $\mathcal{A}$ is a family of events.

> A σ-algebra on a set X is a non-empty collection of subsets of X that is closed under complement, countable unions and countable intersections. It is usual to take $\mathcal{A} = \mathcal{P}(\Omega)$.

As usual, if $\mathcal{A}$ a σ-algebra over Ω, for every $A \in \mathcal{A}$ we define $P(A)$ as the function that satisfies the following axioms:

- $P(A) \geq 0$
- $P(\Omega) = 1$
- If $A_1, A_2, \ldots \in \Omega$, $A_i \cap A_j = \emptyset$ for all $i \neq j$, then

$$P(A_1 \cup A_2 \cup \ldots) = \sum_{i=1}^{\infty} P(A_i)$$

A probabilistic model is a 3-uple $(\Omega, \mathcal{A}, P)$. We will assume from now on that Ω refers to a sample space, $\mathcal{A}$ to $\mathcal{P}(\Omega)$, and P to the probability function.

A random variable is a function $X : \Omega \mapsto \mathbb{R}$.

3 Elementary laws

3.1 Union, intersection, conditionality, etc.

This is a collection of notes. Their justification should be intuitively accessible if one stops and think of their formulas in terms of the subspaces of Ω involved.

Let $A, B \in \Omega$. The probability that A occurs given that B has occurred is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Observe that this gives a formula for $P(A \cap B)$. Furthermore,

$$P(A \cap B) = P(A)P(B \mid A) = P(B)P(A \mid B)$$

If A, B are independent, $P(A \cap B) = P(A)P(B)$.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

If A, B are mutually exclusive then $P(A \cap B) = \emptyset$ and $P(A \cup B) = P(A) + P(B)$.

It is useful to remember the following property too. Since $P(A \cup B) = P(A) + P(B)$ we have that $P((A \cap B) \cup (A - B)) = P(A \cap B) + P(A - B)$, which implies

$$P(A - B) = P(A) - P(A \cap B)$$

3.2 The law of total probability and Bayes' rule

Let $B_1, \ldots, B_k$, $k \in \mathbb{N}$, s.t.

$$\begin{aligned} \Omega &= B_1 \cup \ldots \cup B_k \\ \forall i, j \in [1, k] : i \neq j : B_i \cap B_j &= \emptyset \end{aligned}$$

Then $\{B_1, \ldots, B_k\}$ is a partition of Ω. If $A \subseteq \Omega$ then it can be decomposed using a partition $\{B_1, \ldots, B_k\}$ as $A = (A \cap B_1) \cup \ldots (A \cap B_k)$.

Theorem 1 *If $\{B_1, \ldots, B_k\}$, $k \in \mathbb{N}$, is a partition of Ω s.t. $P(B_i) > 0$ for all $1 \leq i \leq k$, then for any $A \subseteq \Omega$*

$$P(A) = \sum_{i=1}^{k} P(A \mid B_i)P(B_i)$$

Proof. Let $A \subseteq \Omega$. Because $B_1, \ldots, B_k$ partition Ω, $(A \cap B_i) \cap (A \cap B_j) = A \cap \emptyset = \emptyset$. Thus, the two events are mutually exclusive. Thus

$$\begin{aligned} P(A) &= P\left((A \cap B_1) \cup \ldots (A \cap B_k)\right) \\ &= P(A \cap B_1) + \ldots + P(A \cap B_k) \\ &= \sum_{i=1}^{k} P(A \mid B_i) P(B_i) \end{aligned}$$

Theorem 2 (Bayes' Rule) *Assume $\{B_1, \ldots, B_k\}$ is a partition of Ω and $P(B_i) > 0, i = 1, \ldots, k$. Then*

$$P(B_j \mid A) = \frac{P(A \mid B_j) P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i) P(B_i)}$$

The proof follows from the definition of conditional probability and the law of total probability.

4 Discrete random variables

A random variable $X : \mathcal{D}_X \subseteq \Omega \mapsto \mathbb{R}$ is discrete iff $\mathcal{D}_X$ is finite or countably infinite.

If Y is a random variable then expression $(Y = y) = \left\{\zeta \in \omega : X_\zeta = y\right\}$. In other words, $(Y = y)$ denotes the subset of Ω whose elements are assigned the value y by the random variable.

Example. In a coin toss, a random variable X may assign to the sample point "heads" the value 1 and the sample point "tails" the value -1. Then $(X = 1) = 1$, etc.

We define $P(Y = y) = \sum_{\zeta \in \Omega : Y_\zeta = y} P(\zeta)$. The probability distribution of Y is the general function

$$\begin{aligned} p : \mathbb{R} &\mapsto [0, 1] \\ y &\mapsto P(Y = y) \end{aligned}$$

Since the probability distribution p is defined as the probability of given sets of events, it follows that $0 \leq p(y) \leq 1$ for all y and $sum_y p(y) = 1$.

We asume the reader knows the definition of expected value. Let $g : \mathbb{R} \mapsto \mathbb{R}$. Then $g \circ Y$ (or simply $g(Y)$) has expected value

$$\mathbb{E}\left[g(Y)\right] = \sum_{y \in Im(Y)} g(y)p(y)$$

Proof. $P(g(Y) = g_i) = \sum_{y \in Im(Y), g(y)=g_i} p(y)$. Let this probability function for $g(Y)$ be called $p_g(y)$. Then

$$\begin{aligned}
\mathbb{E}[g(Y)] &= \sum_{y \in Im(g \circ Y)} y p_g(y) \\
&= \sum_{y \in Im(g \circ Y)} y \left[\sum_{x \in Im(Y), g(x)=y} p(x) \right] \\
&= \sum_{y \in Im(g \circ Y)} \left[\sum_{x \in Im(Y), g(x)=y} y p(x) \right] \\
&= \sum_{x \in Im(y)} g(x) p(x)
\end{aligned}$$

Definition 1 *Let $\mu = \mathbb{E}\left[Y\right]$. Then*

$$\mathbb{V}\left[Y\right] = \mathbb{E}\left[(Y - \mu)^2\right]$$

Theorem 3 *Let Y a random variable with p.m.f. p and $g_1, \ldots, g_k$ functions of Y. Then*

$$\mathbb{E}\left[g_1(Y) + \ldots + g_k(Y)\right] = \mathbb{E}\left[g_1(Y)\right] + \ldots + \mathbb{E}\left[g_k(Y)\right]$$

Theorem 4

$$\mathbb{V}[Y] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2$$

This is also easy to prove from the definition of $\mathbb{V}$.

5 Finales

5.1 Final 2003-12

Problem 1 *Prove a.* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$*, b.* $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$*, c.* $A \subset B \Rightarrow P(A) \leq P(B) \wedge P(B - a) = P(B) - P(A)$.

Let $(\Omega, \mathcal{A}, P)$ be an arbitrary probabilistic model and let $A, B, C \in \Omega$.

(1) Consider the set $A \cup B$ and let $A \cap B = I$. Observe that

$$A \cup B = (A \cap B) \cup (A - B) \cup (B - A)$$

Then $P(A \cup B) = P((A \cap B) \cup (A - B) \cup (B - A))$. Since the intersection of these events is empty, by the axioms of the probability function we have $P(A \cup B) = P(A \cap B) + P(A - B) + P(B - A)$. Using the fact that $P(X - Y) = P(X) - P(X \cap Y)$ we have

$$P(A \cap B) + P(A) - P(A \cap B) + P(B) - P(B \cap A) = P(A) + P(B) - P(A \cap B)$$

(2)

$$\begin{aligned} P(A \cup B \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P((A \cap C) \cup (B \cap C)) \\ &= P(A) + P(B) + P(C) - P(A \cap B) \\ &- [P(A \cap C) + P(B \cap C) - P(A \cap B \cap C)] \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \blacksquare \end{aligned}$$

Problem 2 *Define the variance of a random variable X. Show that $\mathbb{V}[cX] = c^2\mathbb{V}[X]$, $\mathbb{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y]$ if X, Y independent.*

(1) The variance of a random variable X is $\mathbb{V}[X] = \mathbb{E}\left[(X-\mu)^2\right]$ where $\mu = \mathbb{E}[X]$.

(2) Observe that

$$\begin{aligned}
\mathbb{V}[cX] &= \mathbb{E}\left[(cX)^2\right] - \mathbb{E}[cX]^2 \\
&= c^2\mathbb{E}\left[X^2\right] - \mathbb{E}[cX]\,\mathbb{E}[cX] \\
&= c^2\mathbb{E}\left[X^2\right] - c^2\mathbb{E}[X]^2 \\
& c^2\left(\mathbb{E}\left[X^2\right] - \mu^2\right) \\
& c^2\mathbb{V}[X]
\end{aligned}$$

(3)

$$\begin{aligned}
\mathbb{V}[X+Y] &= \mathbb{E}\left[(X+Y)^2\right] - \mathbb{E}[X+Y]^2 \\
&= \mathbb{E}\left[X^2 + 2XY + Y^2\right] - (\mu_X + \mu_Y)^2 \\
&= \mathbb{E}\left[X^2\right] + 2\mathbb{E}[XY] + \mathbb{E}\left[Y^2\right] - \mu_X^2 - 2\mu_x\mu_Y - \mu_Y^2 \\
&= \mathbb{E}\left[X^2\right] - \mu_X^2 + \mathbb{E}\left[Y^2\right] - \mu_Y^2 + 2\mathbb{E}[X]\,\mathbb{E}[Y] - 2\mu_x\mu_Y \quad \{\text{Independence}\} \\
&= \mathbb{V}[X] + \mathbb{V}[Y]
\end{aligned}$$

Problem 3 *Give a* 95% *confidence interval for the mean* μ *assuming the variance* σ^2 *is known. Then assuming the variance us unknown.*

(1) Given a sample $\mathbf{x} = X_1, X_2, \ldots, X_n$ with $X_i \sim \mathcal{N}(\mu, \sigma)$, we can use the fact that $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$ to construct the statistic

$$Z = \frac{\overline{X} - \mu}{\sigma}\sqrt{n}$$

With sufficiently large n, $Z \sim \mathcal{N}(0, 1)$. We want to choose a value of Z s.t. it occupies .975 of the area under the standard normal curve. Such value is $Z = 1.96$. The confidence interval is then

$$\left[\overline{X} - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right]$$

If σ is unknown we would simply use $\hat{\sigma} = \frac{1}{n-1}\sum(X_i - \overline{X})^2$ as an estimator and keep everything else the same.

(2) If the variance is unknown *and* the sample size is $n \leq 30$, then we must use $\hat{\sigma}$ as before, but use the t-Student distribution. Namely, our confidence interval will now be

$$\overline{X} \pm t_{0.025}\hat{\sigma}$$

The degrees of freedom of the t-Student distribution depends on n, of course.

Problem 4 *The number of kids that come to a vending machine during an hour is a discrete random variable Y with values in* $\{0, 8, 18, 30\}$.

(1) If $P(Y = 8) = \frac{1}{4}, P(Y = 18) = \frac{1}{3}, \mathbb{E}[Y] = 13$, what is the value of $P(Y = 30)$?

We know $\mathbb{E}[Y] = \sum_{y \in Im(Y)} yp(y) = 8 \cdot \frac{1}{4} + 18 \cdot \frac{1}{3} + 0p(0) + 30p(30) = 13$. Then

$$8 + 30p(30) = 13 \Rightarrow p(30) = \frac{5}{30} = \frac{1}{6}$$

(2) What is the value of $P(Y = 0)$?

We require that $\sum_{y \in Im(Y)} p(y) = 1$. We have

$$\begin{aligned} \sum_{y \in Im(Y)} p(y) &= \frac{1}{4} + \frac{1}{3} + \frac{1}{6} + p(0) \\ &= \frac{3}{4} + p(0) \end{aligned}$$

Then $\frac{3}{4} + p(0) = 1 \Rightarrow p(0) = \frac{1}{4}$

(3) Find $P(12 \le Y \le 20)$ and $P(Y \ne 30)$.

$P(12 \le Y \le 20) = P(18) = \frac{1}{3}$. $P(Y \ne 30) = \frac{1}{4} + \frac{1}{4} + \frac{1}{3} = \frac{5}{6}$ (Consistent with the fact that $1 - \frac{1}{6} = \frac{5}{6}$)

(4) If each sell makes 1.30 dollars and it costs 8 to maintain the machine for an hour, what is the expected value of the net profit in an hour?

The expected number of kids to approach the vending machine is 13. Each spends 1.30 dollars with an expected profit of 16.9. Minus the cost we have an expected net profit of 8.9.

Problem 5 *Let $X_1, X_2, \ldots, X_n$ random sample where each X_i has density*

$$f(x) = \begin{cases} \frac{1}{2}(1+\theta x) & -1 \leq x \leq 1 \\ 0 & otherwise \end{cases}$$

and where $\theta \in [-1, 1]$. Find $\mathbb{E}[X_i]$. What is the value of $\mathbb{E}\left[\overline{X}\right]$? If $\hat{\theta} = 3\overline{X}$, is it an unbiased estimator of θ?

(1) By definition,

$$\begin{aligned} \mathbb{E}[X_i] &= \frac{1}{2}\int_{\mathbb{R}} x + \theta x^2 \, dx \\ &= \frac{1}{2}\left(\int_{-1}^{1} x \, dx + \theta \int_{-1}^{1} x^2 \, dx\right) \\ &= \frac{1}{2}\left(\theta\left[\frac{1}{3} + \frac{1}{3}\right]\right) \\ &= \frac{\theta}{3} \end{aligned}$$

(2) Recall that $\overline{X} = \frac{1}{n}\sum X_i$. Then

$$\mathbb{E}\left[\overline{X}\right] = \frac{1}{n}\sum \mathbb{E}[X_i] = \frac{\theta}{3}$$

(3) Since $\mathbb{E}\left[\overline{X}\right] = \frac{\theta}{3}$ we have that $\mathbb{E}\left[3\overline{X}\right] = 3\mathbb{E}\left[\overline{X}\right] = \theta$. Thus, by definition, the estimator is unbiased.

5.2 Final

Problem 6 *En la producción de cierto artículo se pueden presentar sólo dos tipos de defectos A y B. Se sabe que A ocurre en un* 5% *de los artículos; B se presenta en un* 3% *de los artículos; y ambos ocurren juntos en un* 1% *de los artículos.*

(1) Dar la probabilidad de que un artículo tomado al azar presente a. solamente el defecto tipo A, b. al menos un defecto, c. ningún defecto.

(2) Sea Y la variable que cuenta el número de defectos encontrados en el artículo elegido al azar. Dé la PDF y la CDF de Y. Calcule el valor esperado de $X = 2 - Y^2$.

(1.a) Sea $\Omega = \{O, A, B, A \cap B\}$ y $\mathcal{A} = \mathcal{P}(\Omega)$. El problema nos da $P(A), P(B), P(A \cap B)$. Nos interesa ahora $P(A - B)$. Evidentemente $A - B \in \mathcal{A}$ y por lo tanto está bien definida la probabilidad. Veamos que

$$\begin{aligned} P(A - B) &= P(A) - P(A \cap B) \\ &= .05 - .01 \\ &= .04 \end{aligned}$$

(1.b) El conjunto deseado es $A \cup B$, pero sabemos que A, B no son disjuntos. Entonces usamos el hecho de que $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Esto da fácilmente $.05 + .03 - .01 = .07$.

(1.c) La probabilidad de que un elemento tenga algún error cualquiera es la probabilidad de que tenga solamente un error de tipo A, solamente un error de tipo B, o ambos. Esto es, $P(\overline{O}) = .04 + .03 + .01 = .08$. Luego $P(O) = .92$. (Otra forma de verlo es tomar directamente $P(A \cup B) = P(A) + P(B) - P(A \cap B) = .08$).

(2) Sabemos por el punto *(1)* que $P(Y = 2) = .01, P(Y = 1) = .07, P(Y = 0) = .98$. Es decir,

$$p(y)_Y = \begin{cases} .92 & y = 0 \\ .07 & y = 1 \\ .01 & y = 2 \end{cases}$$

Esto implica que

$$F(y)_Y = \begin{cases} 0 & = .92 \\ 1 & = .99 \\ 2 & = 1 \end{cases}$$

El valor esperado de $2 - Y^2$ es $2 - \mathbb{E}[Y^2]$. Es fácil observar que

$$\mathbb{E}\left[Y^2\right] = .07^2 + .01^2 \times 2 = .0053$$

Luego el valor esperado de $2 - Y^2$ es 1.9947.

Problem 7 *Una unidad de radar es usada para medir la velocidad de los automóviles en una vía durante la hora de mayor congestionamiento. La velocidad de los automóviles está normalmente distribuida con distribución* $\mathcal{N}(100, 8.5)$. *(1) Dé la probabilidad de que un auto elegido al azar viaje a una velocidad de a lo sumo* 85. *(2) Dé la probabilidad de que viaje a una velocidad entre* 58 *y* 110. *(3) Dé la probabilidad de que uno de diez automóviles elegidos al azar viaje a una velocidad mayor a* 88.

(1) Estandarizamos la variable y utilizamos la distribución normal estándar. Si $X \sim \mathcal{N}(100, 8.5)$ denota la variable de interés (velocidad de un vehículo en el contexto del problema),

$$P(X \leq 85) = \Phi\left(\frac{88 - 100}{8.5}\right) = \Phi(-1.411)$$

La tabla de la distribución estándar da $\Phi(-1.764) = .079$. Es decir, la probabilidad de que un vehículo viaje a a lo sumo 85 km/h es 7.9%.

(2) Según la misma lógica,

$$\begin{aligned} P(58 \leq X \leq 110) &= \Phi\left(\frac{110 - 100}{8.5}\right) - \Phi\left(\frac{58 - 100}{8.5}\right) \\ &= \Phi(1.176) - \Phi(-4.941) \\ &= .879 - 0 \\ &= .879 \end{aligned}$$

(3) Digamos que el evento de obtener un vehículo que viaje a más de 88 km/h es un éxito, y cualquier otro caso un fallo. Evidentemente la cantidad de vehículos que superan 88 km/h en una muestra de diez sigue sigue una distribución binomial $Y \sim \mathcal{B}(P(X \geq 88), 10)$. Se nos pide la probabilidad de que haya exactamente un éxito. Calculemos $p = P(X \geq 88)$. Evidentemente esto es $1 - P(X \leq 88) = 1 - .079 = .921$. Se sigue que

$$\begin{aligned} P(Y = 1) &= \binom{10}{1} \cdot .921 \cdot (1 - .921)^9 \\ &= 10 \cdot .921 \cdot 0 \\ &\approx 0 \end{aligned}$$

Problem 8 *Sea* $X_i \sim \mathcal{N}(\mu, \sigma)$. *Asumamos una muestra de* $X_1, \ldots, X_{18}$ *una muestra con media muestral* $\overline{X} = 99.45$ *y desviación estándar* $s_2 = 1.3$. *Dé estimaciones de máxima verosimilitud para la media, la varianza y el percentil* 5%. *Construya un intervalo de confianza del* 99% *para la media poblacional.*

Haremos solo el intervalo de confianza. La varianza poblacional es desconocida y la cantidad de datos es menor a 30. Usaremos la distribución t de Student. Recordemos que la media muestral sigue una distribución $\mathcal{N}(\mu, \frac{\sigma}{\sqrt{n}})$. Usaremos el estadístico

$$t = \frac{\overline{X} - \mu}{1.3}\sqrt{18}$$

Sabemos que t sigue una distribución t de Student con 17 grados de libertad. Queremos determinar el intervalo

$$\overline{X} \pm \frac{1.3}{\sqrt{18}} t_{.005}$$

(Vea que $\alpha = .01 \Rightarrow \frac{\alpha}{2} = .005$). Usando la table de la distribución t de Student, tenemos

$$\overline{X} \pm \frac{1.3}{\sqrt{18}} 2.567 = \mu \pm 0.30 \times 2.567$$

Esto resulta en

$$99.45 \pm .7701 = [98.6799, 100.2201]$$

Problem 9 *En el diseño de mascarillas de bomberos se prueba un conjunto de* 120 *mascarillas.* 48 *fallaron la prueba. Dé un intervalo de conianza del* 90% *para p. Determine el tamaño de muestra necesario para que un intervalo de confianza del* 90% *tenga una longitud de a lo sumo la mitad de la obtenida en el item anterior, independientemente del valor de* $\hat{p}$*.*

Si se quiere determinar si hay suficiente evidencia para decir que p es menor a 0.5*, plantee las hipótesis, establezca la región de rechazo con nivel de significación del* 5%*, calcule el p-valor y tome una decisión dado* $\alpha = 0.01$.

(1) Tenemos $\hat{p} = \frac{48}{120} = 0.4$. Para muestras suficientemente grandes, el estimador sigue una distribución normal. Sabemos que la desviación estándar de este estimador es

$$\hat{s} = \sqrt{\frac{0.4 \cdot 0.6}{120}} = 0.044$$

Entonces, usamos el estadístico

$$Z = \frac{0.4 - p}{0.044}$$

que sigue una distribución normal estándar y calculamos el intervalo

$$\begin{aligned} 0.4 \pm 0.044 z_{.05} &= 0.4 \pm 0.044 \times 1.645 \\ &= 0.4 \pm 0.072 \\ &= [.328, .472] \end{aligned}$$

(2) La longitud del intervalo obtenido es .144. El tamaño de muestra rnecesario para que el intervalo tenga una longitud de $.\frac{144}{2} = .072$, independientemente del valor de $\overline{p}$, es dada por la ecuación

$$\begin{aligned} \bar{p} + \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}1.645 - \left(\bar{p} - \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}1.645\right) &\leq .072 \\ 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &\leq .072 \\ \hat{p}(1-\hat{p}) &\leq .001n \\ \frac{\hat{p}(1-\hat{p})}{.001} &\leq n \end{aligned}$$

Si hacemos $u = \hat{p}(1 - \hat{p})$ y observamos que $\frac{1}{.001} = \frac{1}{\frac{1}{1000}} = 1000$, tenemos que $n \geq 1000u$ es el tamaño de muestra necesario para que el intervalo tenga la longitud deseada o menos. En el caso particular de nuestra $\hat{p} = 0.4$, deberíamos tener 240 observaciones. Observe que esto es el doble de las observaciones que tenemos ($n = 120$). Esto tiene sentido, pues se nos pidió reducir la longitud del intervalo a la mitad.

(3) Hagamos la prueba de hipótesis. Sea $H_0 : p = 0.5$. La hipótesis alternativa será $H_1 : p < 0.5$.

Asuma que la hipótesis nula es verdadera. ¿Cuál es la probabilidad de haber encontrado $\hat{p} = 0.4$ en este caso? Si la hipótesis nula fuera verdadera, la desviación estándar debería ser $\sqrt{\frac{0.5^2}{120}} = .045$. El valor observado $\hat{p}$ estaría entonces a

$$z = \frac{0.4 - 0.5}{.045} = -2.222$$

desviaciones estándar de la media. El p-valor será el área de la distribución normal estándar a la izquierda de -2.222—es decir, la probabilidad de observar un valor tan o más extremo que -2.222—. Tomamos el área a la izquierda porque la hipótesis alternativa es que p es *menor* a es 0.5. Entonces, la tabla de la distribución normal nos dice que p-valor = .0132. Como esto es superior a $\alpha = 0.01$, no rechazamos la hipótesis nula.

Problem 10 *Sea $X_1, \dots, X_n$ una muestra con $n \geq 3$ y $X_i \sim Poisson(\lambda)$. (1) Encuentre el estimador de λ usando el método de los momentos. (2) Encuentre el estimador de λ usando máxima verosimilitud. (3) Considere los estimadores*

$$\overline{\lambda_1} = X_1, \overline{\lambda_2} = \frac{X_1 + X_n}{2}, \overline{\lambda_3} = \frac{X_1 + 2X_2 + X_3}{3}, \overline{\lambda_4} = \overline{X}$$

¿Cuál es insesgado? ¿Cuál tiene menor varianza?

Recuerde que si $X \sim \text{Poisson}(\lambda)$ entonces $f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$ para $x \geq 0$.

(1) El primer momento muestral es $\overline{X}$. El primer momento (o la esperanza) de una Poisson es su parámetro λ. Igualando ambos obtenemos $\overline{X} = \lambda$ y vemos que la media muestral es un estimador por el método de los momentos de λ.

(2) Usando máxima verosimilitud, observemos que

$$\mathcal{L}(\lambda \mid X_1, \dots, X_n) = \prod_{i=1}^{n} f(x_i \mid \lambda)$$

Maximizar la expresión de arriba equivale a maximizar su logaritmo. En consecuencia, observamos que

$$\begin{aligned}
\ln\left[\prod_{i=1}^{n} e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right] &= \sum_{i=1}^{n} \ln\left(e^{-\lambda}\frac{\lambda^{x_i}}{x_i!}\right) \\
&= \sum_{i=1}^{n} \left[\ln(e^{-\lambda}) + \ln\left(\lambda^{x_i}\right) - \ln\left(x_i!\right)\right] \\
&= \sum_{i=1}^{n} \left[-\lambda + \ln(\lambda^{x_i}) - \ln\left(x_i!\right)\right] \\
&= -\lambda n + \sum_{i=1}^{n} \ln(\lambda^{x_i}) - \ln(x_i!)
\end{aligned}$$

Sea Λ la expresión arriba. Entonces

$$
\begin{aligned}
\frac{\partial \Lambda}{\partial \lambda} &= -n + \sum_{i=1}^{n} \frac{\partial \ln(u)}{\partial u}\frac{\partial u}{\partial \lambda} && \{u = \lambda^{x_i}\} \\
&= -n + \sum_{i=1}^{n} \frac{1}{\lambda^{x_i}} x_i \lambda^{x_i - 1} \\
&= -n + \sum_{i=1}^{n} \frac{x_i}{\lambda} \\
&= -n + \frac{1}{\lambda}\sum_{i=1}^{n} x_i
\end{aligned}
$$

Encontramos los puntos críticos respecto a λ tomando

$$
-n + \frac{1}{\lambda}\sum x_i = 0 \Rightarrow \sum x_i = \lambda n
$$

o equivalentemente

$$
\lambda = \frac{1}{n}\sum x_i = \overline{X}
$$

Que este punto es un máximo se sigue de que la distribución de Poisson es cóncava y carece de mínimo.

(3) Observe que $\mathbb{E}[X_1] = \lambda$, $\frac{1}{2}\mathbb{E}\left[X_1 + X_2\right] = \frac{1}{2}2\lambda = \lambda$, y $\frac{1}{3}\mathbb{E}\left[X_1 + 2X_2 + X_3\right] = \frac{1}{3}\left(\lambda + 2\lambda + \lambda\right) = \frac{4\lambda}{3}$. Se sigue que el primer y segundo estimador son insesgados y el tercero es sesgado. Ya hemos establecido que el primer momento muestral $\overline{X}$ es insesgado en el punto *(1)*.

Observe que

$$
\begin{aligned}
\mathbb{V}\left[X_1\right] &= \lambda \\
\frac{1}{4}\mathbb{V}\left[X_1 + X_2\right] &= \frac{1}{4}\lambda \\
\frac{1}{9}\left[\lambda + \frac{1}{4}\lambda + \lambda\right] = \left[\frac{1}{9}\frac{9\lambda}{4}\right] &= \frac{\lambda}{4}
\end{aligned}
$$

El mejor estimador es el segundo, porque de los insesgados es el que tiene menor varianza.

5.3 Final 2021-07-26

Problem 11 *In making a certain article, two types of defect exist: Type I and Type II. Type I occurs* 5% *of the times; type II occurs* 10% *of the times. We can assume the occurrence of one defect is independent of the occurrence of the other. A random article is selected. (1) What is the probability that it is flawed? (2) Assuming it is flawed, what is the probability that it only contains a Type I defect?*

(1) The probability that it is flawed is simply $P(\text{Type I}\cup\text{Type II}) = P(\text{Type I}) + P(\text{Type II}) - P(\text{Type I}\cap\text{Type 2})$. Since the defects are independent, this gives $.05 + .1 - .05 \cdot .1 = .145$.

(2) For brevity, let A denote the event of a Type I defect, B the event of a Type II defect. Then we want to find

$$\begin{aligned} P(A \cap \neg B \mid A \cup B) &= \frac{P\left((A \cap \neg B) \cap (A \cup B)\right)}{P(A \cup B)} \\ &= \frac{P\left((A \cap \neg B \cap A) \cup (A \cap \neg B \cap B)\right)}{.145} \\ &= \frac{P\left((A \cap \neg B) \cup \emptyset\right)}{.145} \\ &= \frac{P\left(A \cap \neg B\right)}{.145} \\ &= \frac{.05 \cdot (1 - .1)}{.145} \\ &= .310 \end{aligned}$$

In other words, assuming that an item is flawed, the probability that its flaw is of Type I is 31%.

Problem 12 *Let X a random var with CDF*

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{(x+2)^2}{8} & -2 \leq x < 0 \\ 1 - \frac{(-x+2)^2}{8} & 0 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

(1) Is X continuous or discrete? Justify. (2) Find the PDF or PMF of X. (3) Find the 25 *percentile of X. (4) Find the expected value and standard deviation of X.*

(1) X is said to be discrete if there is some finite or countably infinite set A s.t. $P(X \in A) = 1$. Evidently $P(X \in A) = 1$ if and only if $A = [c, 2]$ with $c \in \mathbb{R}$ and $c \leq -2$. Any set of this form is infinite. Then X is continuous.

(2) The PDF will be the case-to-case derivative of the CDF:

$$f(x) = \begin{cases} 0 & x < -2 \\ \frac{2(x+2)}{8} & -2 \leq x < 0 \\ 1 + \frac{2(-x+2)}{8} & 0 \leq x < 2 \\ 0 & x \geq 2 \end{cases}$$

To make sure this is correct, you can cerify that $\int_{\mathbb{R}} f(x)\, dx = 1$ (I skip this).

(3) We first need to determine which "part" of the function f contains the 25 percentile. Lets integrate the first "part":

$$\frac{1}{4} \int_{-2}^{0} (x+2)\, dx = \frac{1}{2}$$

Since 50% of the probability lies within the region $(-2, 0]$, clearly the 25 percentile is within this region. Now observe that

$$\begin{aligned} \frac{1}{4} \int_{-2}^{t} (x+2)\, dx &= \frac{1}{4} \left[\left(\frac{t^2}{2} - 2 \right) + (2t + 4) \right] \\ &= \frac{t^2}{8} - \frac{1}{2} + \frac{t}{2} + 1 \end{aligned}$$

We want to find the t that contains 25% of the distribution. Solving the equation

$$\begin{aligned} \frac{t^2}{8} - \frac{1}{2} + \frac{t}{2} + 1 &= .25 \\ \frac{t^2 + 4t}{8} &= -.25 \\ t^2 + 4t &= -2 \\ t^2 + 4t + 2 &= 0 \end{aligned}$$

The roots are $-2 \pm \sqrt{2}$. Obviously, $-2 - \sqrt{2}$ falls out of the range we are interested in. Then $-2 + \sqrt{2}$ is the 25 percentile.

(4) We skip the calculations but give the formula. The expected value is

$$\mathbb{E}[X] = \frac{1}{4} \int_{-2}^{0} (x + 2)x \, dx + \int_{0}^{2} x \, dx + \frac{1}{4} \int_{0}^{2} (-x + 2)x \, dx = 2$$

Then

$$\begin{aligned} \mathbb{V}[X] &= \mathbb{E}\left[(X - 2)^2\right] \\ &= \frac{1}{4} \int_{-2}^{0} (x + 2)(x - 2)^2 \, dx + \int_{0}^{2} (x - 2)^2 \, dx + \frac{1}{4} \int_{0}^{2} (-x + 2)(x - 2)^2 \, dx \\ &= \frac{22}{3} \end{aligned}$$

Problem 13 *Let* $X \sim (40, 8), Y \sim (30, 6)$. *(1) Give the probability that* $Y \in [17.52, 33.84]$. *(2) Give the probability that* $Y \leq X$. *(3) Ten draws* $Y_1, \ldots, Y_{10}$ *are taken; what is the probability that only three of the ten draws exceeds* 33.84*? And what the probability that* $\overline{Y}$ *is inferior to* 33.84*?*

(1) Such probability is $\Phi(\frac{33.84-30}{6}) - \Phi(\frac{17.52-30}{6}) = \Phi(.64) - \Phi(-2.08)$. Using the z-score table we observe that this gives $.738 - .018 = .72$. The probability is 72%.

(2) Observe that $Y \leq X \iff Y - X \leq 0$. Using the properties of normal distributions, we have that $Z = Y - X \sim \mathcal{N}(-10, 14)$. Then we require to compute only $\Phi(\frac{10}{14}) = \Phi(0.714) = .761$.

(3) The experiment is binomial with $p = P(X > 33.84) = 1 - P(X \leq 33.84) = 1 - .72 = .28$ and $n = 10$. Then the desired event has probability

$$\binom{10}{3}(.28)^3(.72)^7 = .264$$

We know $\overline{Y} \sim \mathcal{N}(30, \frac{6}{\sqrt{10}})$. Then $P(\overline{Y} \leq 33.84)$ is given by

$$\Phi(\frac{3.84 \times \sqrt{10}}{6}) = \Phi(2.023) = .978$$

Problem 14 *An article says only one out of three people get a job after college. A study found* 85 *out of* 200 *people got jobs. (1) Build a* 98% *confidence interval for the true proportion. (2) Can you conclude with a significance level* $\alpha = .02$ *that the proportion is greater than the one published in the article?*

We have $\hat{p} = .425$. The standard deviation of this estimator is

$$s_{\hat{p}} = \sqrt{\frac{(.425)(.575)}{200}} = .035$$

Since the estimator approximates a normal distribution with sufficiently large samples, we build the confidence interval $\hat{p} \pm s_{\hat{p}} \cdot z_{.01} = .425 \pm .035 \cdot 2.32$. This gives the interval $[0.3438, 0.5062]$.

(2) Let $H_0 : p = 0.33, H_a : p > 0.33$. Let us assume H_0 holds. We ask: What is the probability of having a value as extreme or more than $\overline{p} = .425$ under this assumption? In other words, what is the area under the curve of the distribution of p, under hypothesis H_0, to the right of .425? Since

$$\Phi(\frac{.425 - .33}{.035}) = \Phi(2.71) = .996$$

is the area to the left of .425, $1 - .996 = .004$ is the area to the right. Incidentally, this is the p-value. Since the p-value is less than .02 we reject the null hypothesis and accept the alternative hypothesis.

Problem 15 () 25 *measures of the amount of a substance were made with a mean* 7975 *and* $s_n = 74$. *Assume the random variable follows* $\mathcal{N}(\mu, \sigma)$. *(1) Give a MLE estimator of* $\sqrt{\mu}, \sigma^2$ *and* $P(X \leq 7990)$. *Justify. (2) Give a confidence interval of* 95% *for* μ. *(3) Is there evidence to conclude that* $\mu > 7950$? *(*$\alpha = .05$*) (4) Assume* $\sigma = 73$, *is there evidence to conclude that* $\mu > 7950$? *(*$\alpha = .05$*)*

(1) It is a property of MLE estimation that it satisfies *functional invariance*. This means that if $\hat{\theta}$ is the MLE of θ, then $g(\hat{\theta})$ is the MLE of $g(\theta)$. We know $\overline{X}$ is the MLE of μ. Then $\sqrt{\overline{X}}$ is the MLE of $\sqrt{\mu}$. The same principle gives that s_n^2 is the MLE estimator of σ^2. Once more, the same priniciples give that the MLE estimator of $P(X \leq 7990)$ is $\Phi\left(\frac{7990-\overline{X}}{s_n}\right)$.

(2) We must use the t-Student distribution with 24 degrees of freedom because $n = 25 \leq 30$ and σ is unknown. We then have the interval $\overline{X} \pm s_n \cdot t_{.025}$. This gives $[7975 - 74 \cdot 2.064, 7975 + 74 \cdot 2.064] = [7822.264, 8117.736]$.

(3) Let $H_0 : \mu = 7950, H_a : \mu > 7950$. Assuming H_0, the probability of observing a value as extreme or more (to the right) than 7975 is

$$1 - \Phi(\frac{7975 - 7950}{74}) = 1 - \Phi(0.338) = 1 - .62930 = 0.3707$$

This is the p-value. There isn't enough evidence to reject the null hypothesis.

(4) If we assume $\sigma = 73$, then

$$1 - \Phi(\frac{7975 - 7950}{73}) = 1 - \Phi(0.342)$$

and we still do not reject.

Problem 16 *Let $X1, \dots, X_n$ a random sample with uniform distribution $\mathcal{U}\left[\theta; \theta+1\right]$, with $\theta > 0$. (1) Consider $\hat{\theta} = \max X_i$ an estimator of θ whose PDF is*

$$f_{\hat{\theta}}(x) = \begin{cases} n(x-\theta)^{n-1} & x \in (\theta, \theta+1) \\ 0 & otherwise \end{cases}$$

Find the expected value of $\hat{\theta}$. (2) Find the method of moments estimator of θ. Is it unbiased? (3) Let $\hat{\theta}_2 = \hat{\theta} - \frac{n}{n+1}$. Es it an unbiased estimator of θ?

(1) By definition,

$$\mathbb{E}\left[\hat{\theta}\right] = n \int_{\theta}^{\theta+1} x(x-\theta)^{n-1}\, dx$$

Let $u = x - \theta$ s.t. $x = u + \theta$ and $du = dx$.

$$\begin{aligned} \mathbb{E}\left[\hat{\theta}\right] &= n \int_{u(\theta)}^{u(\theta+1)} (u+\theta)u^{n-1}\, du \\ &= n \int_{0}^{1} u^n + \theta u^{n-1}\, du \\ &= \frac{n}{n+1} + \theta \end{aligned}$$

(2) The first sample moment is $\overline{X}$. The first moment of the uniform distribution is $\frac{a+b}{2}$, or in our case $\frac{2\theta+1}{2}$. Then

$$\overline{X} = \theta + \frac{1}{2} \Rightarrow \theta = \overline{X} - \frac{1}{2}$$

is the method of moments estimator of θ. Observe that

$$\begin{aligned} \mathbb{E}\left[\overline{X} - \frac{1}{2}\right] &= \mathbb{E}\left[\overline{X}\right] - \frac{1}{2} \\ &= \frac{1}{n}\sum \mathbb{E}\left[X_i\right] - \frac{1}{2} \\ &= \theta + \frac{1}{2} - \frac{1}{2} \\ &= \theta \end{aligned}$$

The estimator is by definition unbiased.

(3) Observe that

$$\begin{aligned}\mathbb{E}\left[\hat{\theta} - \frac{n}{n+1}\right] &= \mathbb{E}\left[\hat{\theta}\right] - \frac{n}{n+1} \\ &= \frac{n}{n+1} + \theta - \frac{n}{n+1} \\ &= \theta\end{aligned}$$

It is unbiased.