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 ally $i \neq j$, 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

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 $P(A \cup B) = P(A) + P(B)$.

3.2 The law of total probability and Bayes' rule

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

$$\Omega = B_1 \cup \ldots \cup B_k$$
$$\forall i, j \in [1, k] : i \neq j : B_i \cap B_j = \emptyset$$

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 $\mu = \mathbb{E}[Y]$. Then*

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

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

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

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 $P(A \cup B) = P\left((A \cap B) \cup (A - B) \cup (B - A)\right)$. 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)

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

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 $P(12 \leq Y \leq 20)$ and $P(Y \neq 30)$.

$P(12 \leq Y \leq 20) = P(18) = \frac{1}{3}$. $P(Y \neq 30) = \frac{1}{4} + \frac{1}{4} + \frac{1}{3} = \frac{5}{6}$ (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*

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

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* $X = 2 - Y$.

(1) Sea $\mathcal{A} \subseteq \Omega$ el subconjunto del espacio muestral $\Omega = \{O, A, B, AB\}$ asociado a todos los artículos que tienen el error tipo A. Definimos de manera equivalente $\mathcal{B}$.

Evidentemente, el espacio $\mathcal{A}'$ que nos interesa es $\mathcal{A}' = \mathcal{A} \cap \overline{\mathcal{B}}$. Pero estos eventos no son necesariamente independientes. Sin embargo, podemos observar que

$$\mathcal{A} \cap \overline{\mathcal{B}} = \mathcal{A} - (\mathcal{A} \cap \mathcal{B})$$

Sabemos que $P(\mathcal{A} \cap \mathcal{B}) = 0.01$. Luego

$$P(\mathcal{A} \cap \overline{\mathcal{B}}) = P(\mathcal{A} - \mathcal{A} \cap \overline{B})$$