II Università di Roma, Tor Vergata
Dipartimento d'Ingegneria Civile e Ingegneria Informatica
LM in Ingegneria dell'Informazione e dell'Automazione
Complementi di Probabilità e Statistica - Advanced Statistics
Instructors: Roberto Monte & Massimo Regoli
Final Test - 2020-02-12 - Probability

Problem 1 *The scrutiny of group of* $100,000$ *randomly chosen male people in the age* $40-79$ *in UK during* $2013-2015$ *reveals the following table of average lung cancer incidence*

	smoker	*not smoker*	*total*
lung cancer	$10,395$	$7,407$	$17,802$
not lung cancer	$50,078$	$32,120$	$82,198$
	$60,473$	$39,527$	$100,000$

Write Ω *for the sample space consisting of these* $100,000$ *people and write* S *[resp.* C*] for the subsets of* Ω *consisting of the smokers [the people affected by lung canger]. Let* $1_S : \Omega \to \{0,1\}$ *and* $1_C : \Omega \to \{0,1\}$ *the indicator functions of the events* S *and* C *respectively.*

1. *Determine the joint distribution of the random vector* $(1_S, 1_C)$ *and the marginal distributions of the random variables* 1_S *and* 1_C.
2. *Are the random variables* 1_S *and* 1_C *independent?*
3. *What is the probability that a randomly chosen person in* Ω *is affected by lung cancer, given that he is a smoker?*
4. *What is the probability that a randomly chosen person in* Ω *is a smoker, given that he is affected by lung cancer?*
5. *Check the validity of the total probability formula for* $\mathbf{P}(S)$ *and the Bayes Formula for* $\mathbf{P}(C \mid S)$.

Solution. .

Problem 2 *Let* $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ *be a probability space, let* $(\mathbb{R}, \mathcal{B}(\mathbb{R})) \equiv \mathbb{R}$ *be the Euclidean real line endowed with the Borel* σ*-algebra, and let* X *be a uniformly distributed random variable on* Ω *with states in the interval* $(-1,1)$*. In symbols* $X \sim Unif(-1,1)$*. Consider the function* $g : \mathbb{R} \to \mathbb{R}$ *given by*

$$g(x) \stackrel{def}{=} x^2, \quad \forall x \in \mathbb{R}.$$

1. *Can you state that the function* $Y : \Omega \to \mathbb{R}$ *given by*

$$Y(\omega) \stackrel{def}{=} g(X(\omega)), \quad \forall \omega \in \Omega.$$

   is a real random variable on Ω*?*
2. *Can you compute the distribution function* $F_Y : \mathbb{R} \to \mathbb{R}_+$ *of* $Y : \Omega \to \mathbb{R}$*?*
3. *Is* Y *absolutely continuous?*
4. *Are the first and second order moments of* Y *finite?*
5. *If the first and second order moments of* Y *are finite, can you compute* $\mathbf{E}[Y]$ *and* $\mathbf{D}^2[Y]$*?*

Solution. .

Problem 3 *Let U, V real random variables on a probability space Ω such that such that $U \sim V \sim N(0,1)$, the vector (U, V) is Gaussian, and $Corr(U, V) \equiv \rho < 1$. Consider the real random variables*

$$X \overset{def}{=} U - \rho V \quad and \quad Y \overset{def}{=} \sqrt{1-\rho} V.$$

1. *Can you prove that the vector (X, Y) Gaussian?*
2. *Are the random variables X and Y independent?*
3. *Compute the distributions of X and Y;*
4. *Compute $\mathbf{E}\left[X^2 Y^2\right]$, $\mathbf{E}\left[X Y^3\right]$, $\mathbf{E}\left[Y^4\right]$.*
5. *Compute $\mathbf{E}\left[U^2 V^2\right]$.*

Solution. .

Problem 4 *Let $(\Omega, \mathcal{E}, \mathbf{P}) \equiv \Omega$ a probability space and let B_1 and B_2 be standard Bernoulli random variables on Ω. In symbols, $B_k \sim Ber(1/2)$, for $k = 1, 2$. Assume that B_1 and B_2 are independent and set*

$$X \overset{def}{=} B_1 + B_2, \quad Y \overset{def}{=} B_1 \cdot B_2$$

1. *Compute $\mathbf{E}[B_k \mid X]$ and $\mathbf{E}[B_k \mid Y]$ for $k = 1, 2$.*
2. *Are the random variables $\mathbf{E}[B_1 \mid X]$ and $\mathbf{E}[B_2 \mid X]$ uncorrelated? Are they independent?*
3. *Are the random variables $\mathbf{E}[B_1 \mid Y]$ and $\mathbf{E}[B_2 \mid Y]$ uncorrelated? Are they independent?*
4. *Compute $\mathbf{E}[X \mid Y]$ and $\mathbf{E}[Y \mid X]$.*
5. *Are the random variables $\mathbf{E}[X \mid Y]$ and $\mathbf{E}[Y \mid X]$ uncorrelated? Are they independent?*
6. *Compute $\mathbf{E}\left[X^2 \mid Y\right]$ and $\mathbf{E}\left[Y^2 \mid X\right]$.*

Solution. .

Problem 5 *Let $\theta > 0$ and let X be a real random variable which is uniformly distributed in the interval $[0, \theta]$, in symbols $X \sim U(0, \theta)$. Let $X_1, \ldots, X_n$ a simple random sample of size $n \geq 1$ drawn from X and let $\left(\check{X}_n\right)_{n \geq 1}$ be the sequence of real random variables given by*

$$\check{X}_n \overset{def}{=} \max\{X_1, \ldots, X_n\}, \quad \forall n \geq 1.$$

1. *Prove that the sequence $\left(\check{X}_n\right)_{n \geq 1}$ converges in distribution to the Dirac random variable concentrated in θ.*
2. *Prove directly that the sequence $\left(\check{X}_n\right)_{n \geq 1}$ converges in probability to $Y \sim Dir(\theta)$.*
3. *Prove directly that the sequence $\left(\check{X}_n\right)_{n \geq 1}$ converges in mean to $Y \sim Dir(\theta)$.*
4. *Prove that the sequence $\left(\check{X}_n\right)_{n \geq 1}$ converges in square mean to $Y \sim Dir(\theta)$.*
5. *Does $\left(\check{X}_n\right)_{n \geq 1}$ converge almost surely to $Y \sim Dir(\theta)$?*

   Hint: Determine the distribution function of $\check{X}_n$ and exploit it.

Solution. .