Homework 1, Math 305, Spring 2018

Last name, first name, ID number : Score:

1. Assuming the blood type distribution in a population to be

   - A: 41%,
   - B: 10%,
   - AB: 3%,
   - O: 46%,

   what is the probability that the blood of a randomly selected person

   (a) will contain the A antigene?
   Solution: $P(A \cup AB) = P(A) + P(AB) = 46\%$.

   (b) will contain the B antigen?
   Solution: $P(B \cup AB) = P(B) + P(AB) = 13\%$.

   (c) that it will contain neither the A nor the B antigene?
   Solution: $P(O) = 46\%$.

2. Consider 2 mutually exclusive events A and B such that $P(A) > 0$ and $P(B) > 0$. Are A and B independent? Justify your answer.
   Solution: Assume A and B are mutually exclusive, $P(A) > 0$ and $P(B) > 0$. We have $P(A \cap B) = P(\emptyset) = 0 \neq P(A)P(B)$. Therefore A and B are not independent.

3. Assume that the the number of cars entering a tunnel per-2 minute period is a Poisson random variable and equals one in average. Find the probability that the number of cars entering the tunnel during a 2 minute period exceeds three. Solution: Let X be the number of cars entering the tunnel in a 2-minute period. By assumption, X is a discrete random that follows a Poisson distribution with parameter $\lambda = 1$ (the mean value of X). Therefore

$$
\begin{aligned}
P(X > 3) = 1_P(X \leq 2) &= 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) = 1 \\
&= 1 - (\frac{1^0}{0!} - \frac{1^1}{1!} - \frac{1^2}{2!} - \frac{1^3}{3!})e^{-1} \\
&= 1 - \frac{8}{3e} \\
&\approx 0.019
\end{aligned}
$$

4. Let $x_1, \ldots, x_n$ be a random sample of data points taken from a Poisson distribution with parameter λ. Find the maximum likelyhood estimator $\hat{\lambda}$.
   Solution: The likelihood function $L(\lambda)$ is given by

$$L(\lambda) = \Big(\frac{\lambda^{x_1} e^{-\lambda}}{x_1!}\Big) \ldots \Big(\frac{\lambda^{x_n} e^{-\lambda}}{x_n!}\Big) = \frac{e^{-n\lambda}\lambda^{x_1+\cdots+x_n}}{x_1! \ldots x_n!}.$$

   Since maximizing $L(\lambda)$ is equivalent to maximizing $\ln(L(\lambda))$, the maximum likelihood estimator is the solution of the maximization problem

$$\max_{\lambda} G(\lambda)$$

   where $G(\lambda) = -n\lambda + (x_1 + \cdots + x_n)\ln(\lambda)$. Differentiating we find $G'(\lambda) = -n + (x_1 + \cdots + x_n)\frac{1}{\lambda}$ thus $G'(\lambda) = 0$ if and only if $\lambda = \frac{x_1+\cdots+x_n}{n} = \bar{x}$ that is the arithmetic mean of the data points. Differentiating again, we find $G''(\lambda) = -\frac{x_1+\cdots+x_n}{\lambda^2} < 0$ so $\bar{x}$ is a maximum of G. We conclude that the maximum likelihood estimator of λ is $\hat{\lambda} = \bar{x}$.

5. Assume that Y is a continuous random variable whose probability density function is

$$f(y) = \begin{cases} \frac{3}{8}y^2 \text{ for } 0 \leq y \leq 2, \\ 0 \text{ otherwise.} \end{cases}$$

Find $E(Y)$ and $Var(Y)$.
Solution: We find

$$E(Y) = \int_{-\infty}^{+\infty} y\frac{3}{8}y^2 dy = \frac{3}{8}.\frac{1}{4}y^4\Big|_0^2 = \frac{3}{2}$$

and

$$Var(Y) = E(Y^2) - E(Y)^2 = \int_{-\infty}^{+\infty} y^2\frac{3}{8}y^2 dy - \frac{9}{4} = \frac{3}{8}.\frac{1}{5}y^5\Big|_0^2 - \frac{9}{4} = \frac{12}{5} - \frac{3}{2} = \frac{3}{20}.$$

6. A study shows that the death rate of a person who smokes is twice that of a nonsmoker. In terms of hazard rates (section 5.3.3), it means that $\lambda_s(t) = 2\lambda_n(t)$ (where s stands for smoker and n for nonsmoker). Conclude that, of two people the same age, one smoker and one not, the probability that the smoker survives to any given age is the square of the corresponding probability for the nonsmoker.
Solution: Denote T_s (resp. T_n) the death time of a smoker (resp. of a nonsmoker) and F_s (resp. F_n) the c.d.f of T_s (resp. T_n). By definition of a hazard rate, we have

$$\int_0^t \frac{F_s'(y)}{1 - F_s(y)} dy = 2\int_0^t \frac{F_n'(y)}{1 - F_n(y)} dy.$$

Integrating both sides, by using the substitutions $u = 1 - Fs(y)$ on the left-hand side and $v = 1 - Fn(y)$ on the right-hand side before substituting back, and then exponentiating, we find

$$1 - F_s(t) = (1 - F_n(t))^2$$

that is

$$P(T_s > t) = (P(T_n > t))^2$$

7. Achievement test scores of high school seniors in a state have mean 60 and variance 64. Consider a random sample of $n = 100$ students from some high school. What is the probability that the sample mean is at most 58?
Solution: Denote $\bar{Y}$ the sample mean of the student's scores and Z a standard normal random variable. Applying to the central limit theorem, we find

$$P(\bar{Y} \leq 58) = P(\frac{\bar{Y} - 60}{0.8} \leq \frac{58 - 60}{0.8}) \approx P(Z \leq -2.5) = 0.0062.$$

8. Let X and Y be independent exponential random variables with parameter λ. Let $Z = X + Y$. Show that Z is gamma-distributed. (Hint: calculate $P(Z \leq z)$.)