Prof. Dr. P. Nolin ETH Zürich Winter 2014

Possible Solutions

Probability Theory and Statistics (BSc D-ITET)

1. a) The following probabilities are given in the exercise

$$\begin{aligned} P(<\text{The train departs on time}>) &= P(DoT) = 0.7 \\ P(<\text{The train arrives on time}>) &= P(AoT) = 0.75 \\ P(<\text{The train departs on time AND arrives on time}>) &= P(DoT \cap AoT) = 0.6. \end{aligned}$$

Therefore, we obtain

$$\begin{aligned} P(<\text{The train departs with delay}>) &= P(DwD) = 1 - P(DoT) = 0.3 \\ P(<\text{The train arrives with delay}>) &= P(AwD) = 1 - P(AoT) = 0.25. \end{aligned}$$

We need to compute $P(AoT \mid DwD)$. From the definition of conditional probabilities, one has

$$P(AoT \mid DwD) = \frac{P(AoT \cap DwD)}{P(DwD)}.$$

Since we already found that $P(DwD) = 0.3$, we now focus on the numerator. Since a train can either depart on time or with delay it holds

$$P(AoT) = P(AoT \cap DwD) + P(AoT \cap DoT).$$

Therefore, one has

$$P(AoT \cap DwD) = P(AoT) - P(AoT \cap DoT) = 0.75 - 0.6 = 0.15.$$

Hence, we calculate

$$P(AoT \mid DwD) = \frac{0.15}{0.3} = 0.5.$$

b) Let T be the event <The member has trained>, and let F be the event <The member finishes the tour>. It is given in the exercise $P(T^c) = 0.2$. Moreover, we know from the exercise that

$$P(F \mid T) = 0.85, \quad P(F \mid T^c) = 0.6.$$

Please Turn!

i) We need to compute $P(F)$. Applying the formula of total probabilities, we get

$$P(F) = P(F \mid T)P(T) + P(F \mid T^c)P(T^c) = 0.85 \cdot 0.8 + 0.6 \cdot 0.2 = 0.68 + 0.12 = 0.8.$$

ii) We need to compute $P(T \mid F)$. Applying Bayes' formula, we obtain

$$P(T \mid F) = \frac{P(F \mid T)P(T)}{P(F)} = \frac{0.85 \cdot 0.8}{0.8} = 0.85.$$

c) The unconditioned law of N is given by

$$\begin{aligned} P(N = n) &= \int_0^\infty e^{-\lambda} \frac{\lambda^n}{n!} \frac{1}{2} e^{-\lambda/2} d\lambda = \int_0^\infty e^{-\frac{3\lambda}{2}} \lambda^n \frac{d\lambda}{2n!} \\ &= \int_0^\infty e^{-y} \left(\frac{2}{3} y\right)^n \frac{2}{3} \frac{1}{2n!} dy = \frac{1}{2n!} \left(\frac{2}{3}\right)^{n+1} \underbrace{\int_0^\infty e^{-y} y^n dy}_{=n!} \\ &= \frac{1}{3}\left(\frac{2}{3}\right)^n. \end{aligned}$$

d) Observe that the distribution obtained in **c)** differs from the one in the hint only by the value of the parameter p. We solve here the question for a general $p \in (0, 1)$.

$$\begin{aligned} E[N] &= \sum_{k \geq 0} kp(1-p)^k = p \sum_{k \geq 0} k(1-p)^k \\ &= p(1-p) \sum_{k \geq 1} k(1-p)^{(k-1)} = p(1-p) \left(-\sum_{k \geq 0} (1-p)^k \right)' \\ &= p(1-p) \left(-\frac{1}{1-(1-p)} \right)' = p(1-p) \frac{1}{p^2} = \frac{1-p}{p}. \end{aligned}$$

Hence for $p = \frac{1}{3}$ as in **c)**, we get $E[X] = 2$; otherwise, using the hint, $E[X] = 3$.

Remark: N has a geometric distribution shifted by -1, therefore it is consistent to have $\frac{1}{p} - 1$ as expected value for N.

2. **a)** Y is a discrete random variable which only takes two values, 0 and 1. We can compute the probability of each of these by integrating the joint density function. We obtain the following:

$$P[Y = 0] = \int_0^1 12(x^2 - x^3)(1 - x) \, \mathrm{d}x = \frac{2}{5},$$

and

$$P[Y = 1] = 1 - P[Y = 0] = \frac{3}{5}.$$

Therefore, $Y \sim \mathrm{Ber}(p = 3/5)$.

See next page!

b) We have $E[X+Y] = E[X] + E[Y] = \int_0^1 12(x^3 - x^4)\,\mathrm{d}x + \frac{3}{5} = \frac{6}{5}$.

c) Since $\mathrm{Cov}(X,Y) = E[XY] - E[X]E[Y]$, we only need to calculate $E[XY]$ since we have the other two from the previous exercise. Using the joint density function, we find that

$$E[XY] = \sum_{y=0}^{1} \int_0^1 xy f_{X,Y}(x,y)\,\mathrm{d}x = \int_0^1 12(x^2 - x^3)x^2\,\mathrm{d}x = \frac{2}{5}.$$

Therefore, $\mathrm{Cov}(X,Y) = \frac{2}{5} - \frac{9}{25} = \frac{1}{25}$, Hence, X, Y are not independent.

d) Note that $\sin(2\pi Y) \equiv 0$, therefore it is independent of any other random variable. In particular, it is independent of X^2.

3. **a)** We recognize immediately that $X \sim \mathrm{Bin}(N = 100, q = (1-p) = 0.99)$. We then calculate $\mathrm{Var}(X) = Nq(1-q) = 0.99$, and we get

$$P(X \le \mathrm{Var}(X)) = P(X = 0) = \binom{100}{0} q^0 (1-q)^{100} = (0.01)^{100}.$$

b) Since X has a Binomial distribution, we have $E[X] = N(1-p) = 100(1-p)$. Using the natural approximation

$$\frac{\sum_{i=1}^n X_i}{n} \simeq E[X] = 100(1-p),$$

we obtain the natural estimator $\tilde{p} = 1 - \frac{\sum_{i=1}^n X_i}{100n}$.

c) The likelihood function is defined by $L(p; x_1, \ldots, x_n) =$ the probability to observe the sequence $x_1, \ldots, x_n$. Since the X_i, with $1 \le i \le n$, are independent, we get

$$L(p; x_1, \ldots, x_n) = \prod_{i=1}^n \binom{100}{x_i} q^{x_i}(1-q)^{100-x_i} = \prod_{i=1}^n \binom{100}{x_i}(1-p)^{x_i} p^{100-x_i}$$

which yields a log-likelihood function given by

$$\log L(p; x_1, \ldots, x_n) = \sum_{i=1}^n \binom{100}{x_i} + x_i \log(1-p) + (100 - x_i)\log(p)$$

Taking its derivative with respect to p and setting it equal to 0, gives then the maximum likelihood estimator $\widehat{p}$:

$$\frac{\partial \log L(p; x_1, \ldots, x_n)}{\partial p} = \sum_{i=1}^n \left(-\frac{x_i}{1-p} + \frac{100 - x_i}{p} \right) = 0$$

$$\iff (100n - s_n)(1-p) = s_n p \iff 100n - s_n = 100np$$

$$\iff p = 1 - \frac{s_n}{100n},$$

where $s_n = \sum_{i=1}^n x_i$. Consequently, we have $\widehat{p} = 1 - \frac{\sum_{i=1}^n X_i}{100n}$.

REMARK: This yields the same estimator as in part **b)**.

Please Turn!

d) We clearly have $n = 1+64+20+15 = 100$, and $s_n = 100n - 1\cdot 0 - 64\cdot 1 - 20\cdot 2 - 15\cdot 3 = 10'000 - 149 = 9'851$. Therefore, we get

$$p^* = \frac{100n - s_n}{100n} = \frac{149}{10000} = 0.0149.$$

e) Let us calculate

$$E[\widehat{p}] = \frac{100n - \sum_{i=1}^n E[X_i]}{100n} \overset{E[X_i]=100(1-p)}{=} \frac{100n - 100n(1-p)}{100n} = p,$$
$$\mathrm{Var}(\widehat{p}) = \frac{\mathrm{Var}(\sum_{i=1}^n X_i)}{(100n)^2} \overset{\text{indep.}}{=} \frac{\sum_{i=1}^n \mathrm{Var}(X_i)}{(100n)^2} \overset{\textbf{a)}}{=} \frac{100n(1-p)p}{(100n)^2} = \frac{(1-p)p}{100n}.$$

Hence, we get

$$\frac{p^* - E[\widehat{p}]}{\sqrt{\mathrm{Var}(\widehat{p})}} = \frac{\sqrt{100n}(p^* - p)}{\sqrt{(1-p)p}} = \frac{100(0.0049)}{\sqrt{0.99}\sqrt{0.01}} \sim \frac{0.49}{0.1} = 4.9.$$

Clearly, under the assumption $p = 0.01$, the likelihood of observing p^* is very small, since it is roughly 4.9 standard deviations away from its expected value. Therefore, it doesn't support the vendor's claim and indicates that the true value of p is larger than 0.01.