Math 394B - Summer 2010

Solutions to the Problems for the Final Exam

1 Quizzes

1.1

Suppose $A_1 \cap A_2 = \emptyset$, and $P[A_1 \cup A_2] = 1$. Then, for any set B

○ $P[B] = P[B\,|A_1] + P[B\,|A_2]$ We need to weigh the conditional probabilities with their "likelihood", i.e. with the corresponding $P[A_k]$

⊗ $P[B] = P[B\,|A_1]\,P[A_1] + P[B\,|A_2]\,P[A_2]$ This is the "total probabilities" formula.

○ $P[B] = P[B \cap A_1]\,P[A_1] + P[B \cap A_2]\,P[A_2]$ This is the reverse mistake, compared to the first expression: it would be true if we had *not* "weighted" each summand with the $P[A_k]$.

○ $P[B] = \frac{P[B \cap A_1]}{P[A_1]} + \frac{P[B \cap A_2]}{P[A_2]}$ This is the same as the first expression

1.2

If we denote by $\mathbb{N}$ the set of natural numbers, $\mathbb{N} = \{1, 2, \ldots\}$, which of the following is a probability mass function on $\mathbb{N}$?

○ $P[k] = 1; k = 1, 2, \ldots$ No way: infinitely many ones add up to "infinite"

○ $P[k] = \frac{2}{k}; k = 1, 2, \ldots$ Doesn't work either: $\sum_k \frac{1}{k}$ is the so-called "harmonic series", and it is divergent (it diverges, approximately, as $\log n$, when we sum up to n)

○ $P[k] = \frac{k}{2}; k = 1, 2, \ldots$ This is more divergent than ever: the sum up to n is $\frac{n(n+1)}{4}$

⊗ $P[k] = 2^{-k}; k = 1, 2, \ldots$ This series is the geometric series, with "ratio" $\frac{1}{2}$, hence it sums to $\frac{1}{2} \cdot \frac{1}{1-\frac{1}{2}} = 1$.

○ $P[k] = 2^k; k = 1, 2, \ldots$ This series is a divergent geometric series! Up to n, it sums up to $2 \cdot \frac{2^n - 1}{2-1} = 2\,(2^n - 1)$

1.3

Suppose $P[X \leq 0] = 1$. Then it is certainly true that (F_X is the cdf of the distribution of X; $\lim_{x \to 0^-}$ means limit from the left)

○ $F_X(0) = 0$ $F_X(0) = P[X \leq 0]$, and we just said it was 1

⊗ $F_X(0) = 1$ As we just said...

○ $F_X(0) = \frac{1}{2}$ This is out of line: X takes no values larger than 0...

○ $\lim_{x \to 0^-} F_X(x) = 0$ See the comment at the first expression

○ $\lim_{x \to 0^-} F_X(x) = 1$ This is mean... This is true if F_X is continuous at 0, but we didn't say that - X could very well be a discrete RV, or, in any case, have a positive probability of taking the value 0. Since cdf's are *right* continuous (due to the definition in terms of $\leq$, rather than $<$), it may not be true that the *left* limit is 1.

2 Problems

2.1

Two syndromes have very similar symptoms, but are very different in their seriousness. Syndrome A is generally benign, and is the cause of the symptoms in 90% of the cases. Syndrome B is extremely serious, hence a test is devised to identify the cause. The test is 99% effective in identifying an A case (if a patient has syndrome A, it is recognized 99% of the time, while 1% of the time it is erroneously reported as B), and 95% effective in recognizing syndrome B, while the remaining 5% are erroneously reported as A.

1. A patient exhibiting symptoms is tested, and the test result indicate A. What chance is there that the patient actually has the more serious B syndrome?

2. A patient exhibiting symptoms is tested, and the test result indicate B. What is the probability that this is indeed the cause of the symptoms?

3. Due to the cost of the test, a cost-benefit analysis is performed. If the cost of an unrecognized B-affected patient is \$100,000 (due to expensive advanced condition treatment), and the cost of recognized B-affected patient is \$10,000, while each test costs \$1,000 to administer, what would the expected cost of testing and consequently treating patients be, as opposed to not testing anybody, so that every B-affected patient would go unrecognized?

Solution: Apart form the third question, this is a very familiar problem on Bayes' Formula...

1. We have (denote the syndrome of a patient by A or B, and the result of the test by a or b, depending on the outcome)

$$P[a|A] = .99; P[b|B] = .95; P[A] = .9$$

Hence,

$$P[B|a] = \frac{P[a|B]P[B]}{P[a]} = \frac{P[a|B]P[B]}{P[a|A]P[A] + P[a|B]P[B]} =$$

$$\frac{.05 \cdot .1}{.99 \cdot .9 + .05 \cdot .1} = .005804$$

(almost 1/50th as the "a priori" probability of being a B)

2. This should be extremely familiar:

$$P[B|b] = \frac{P[b|B]P[B]}{P[b]} = \frac{P[b|B]P[B]}{P[b|A]P[A] + P[b|B]P[B]} =$$

$$\frac{.95 \cdot .1}{.01 \cdot .9 + .95 \cdot .1} = .91346$$

which is not as "surprising" as previous examples, but we should keep in mind that the probability of a B is already 10% from the start...

3. If nobody got screened, the average cost per person would be $.1 \cdot 100000$, i.e. \$10,000 (10% of the symptomatics would have to be treated at the costliest rate). On the other hand, from the computations above, we would miss .58% of the B-affected, but $1 - .91827 = .081731$ (or about 8%) would be false positives. Now, we also know that

$$P[a] = .896; P[b] = .104$$

Hence, the cost per patient would be
$>$ \$1000 for the test
$> .104 \cdot 10^4 = \$1040$ for the positives (false or real)
$> .0058 \cdot 10^5 = \$580$ for the missed positives
for a total of 1000+1040+580=\$2,620, i.e. 26% of the cost of *not* testing people.

2.2

A component has to be fed an electric current in a circuit, and, given its specs, it needs to "see" a voltage of 1.5V within 0.2V (a higher or lower voltage will either burn the component, or make the system non operative). The generating apparatus is producing a voltage W that is normally distributed, with $\mu = 1.4$, and $\sigma^2 = 2.5 \cdot 10^{-3}$.

1. What is the probability that the component will burn?
2. What is the probability that the voltage will fall within the specs?
3. What should the value of σ^2 be to make sure that the probability of the voltage being too low would be no greater than 10^{-3}?

Solution: The acceptable range for W is $[1.3, 1.7]$.

1. The component will burn if the voltage is too high, i.e. $W > 1.7$.

$$P[W > 1.7] = P\left[\frac{W - 1.4}{.05} > \frac{1.7 - 1.4}{.05}\right] = 1 - \Phi(6) = 9.8659 \cdot 10^{-10}$$

2. Now the question is

$$P[1.3 < W < 1.7] = P\left[\frac{1.3 - 1.4}{.05} < \frac{W - 1.4}{.05} < \frac{1.7 - 1.4}{.05}\right] = \Phi(6) - \Phi(-2) =$$

$$= .97725$$

3. We would like σ to be such that

$$P\left[\frac{W - 1.4}{\sigma} < -\frac{.1}{\sigma}\right] \leq 10^{-3}$$

Now, from tables or software, we can get that

$$P[Z < -3.0902] = 10^{-3} \tag{2}$$

(one often writes $z_{.001} = 3.0902$, meaning that the probability that Z, a standard normal, is greater than 3.0902 is .001 - by symmetry, we get (2)). Hence, we need

$$-\frac{.1}{\sigma} \leq -3.0902$$

or

$$\sigma \leq \frac{.1}{3.0902} = .032360$$

2.3

2.3.1

We look for cars going through a checkpoint on I-5. Assume that n cars are approaching, and that the arrival time of each car is distributed according to an exponential distribution with parameter $\lambda = 5$, all being independent of each other.

1. What is the the probability that one car or more will go through over a time span t?
2. * What is the probability that *all* n cars will have gone through by time t?

Solutions

1. As usual, it is easier to compute $P[N_t \geq 1] = 1 - P[N_t = 0]$. Given n cars, that will be $1 - e^{-5nt}$
2. That asks for the maximum of the arrival times to be less than t: $\left(1 - e^{-5t}\right)^n$

2.3.2

Assume now that the flow of cars at our checkpoint follows a Poisson process whose rate depends on the time of the day, as follows: over a 10 hour period, starting at $t = 0$, taken as Midnight, time measured in hours)

$$\lambda = \begin{cases} \lambda_1 = 10 & 0 \leq t < 5 \\ \lambda_2 = 120 & 5 \leq t < 10 \end{cases}$$

1. What is the probability of no car passing the checkpoint between Midnight and 5 a.m.?
2. What is the *average* (*expected*) number of cars passing in the first 5 hours? What is the average number of cars passing in the 10 hours under consideration?
3. * What is the probability of no cars passing between 4 a.m. and 6 a.m.?
4. * What is the probability of exactly one car passing between 4 a.m. and 6 a.m.?
5. * Let T be the time the first car passes the checkpoint after Midnight. Set $T = \infty$ (a symbolic value), if no car appears in the 10 hours. What is the distribution of T?

Hint: For question 5, consider that we should consider that T could be less than 5, or more than 5 and less than 10, or more than 10. These three cases occur, depending on whether or not no car appears in the first 5 hours, and if none did, whether or not no car appeared in the following 5 hours. If you go for $F_T(t)$, or equivalent, you will want to consider separately the cases when $t \leq 5$, $5 < t \leq 10$, $t > 10$ which would be equivalent, in our notation, to $t = \infty$.

Solutions:

1. $e^{-5\lambda_1} = e^{-5\cdot 10} = e^{-50} \approx 1.9287 \times 10^{-22}$
2. It's $5 \cdot 10 = 50$ for the first, and $5 \cdot 10 + 5 \cdot 120 = 650$ for the second
3. No cars have to come in the first hour, and in the second hour: $e^{-\lambda_1} e^{-\lambda_2} = e^{-10} \cdot e^{-120} = e^{-130} \approx 3.4811 \times 10^{-57}$
4. There are only two disjoint possibilities: $\{N_1 = 1, N_2 = 0\}$ and $\{N_1 = 0, N_2 = 1\}$. The probability is thus

$$10e^{-10} \cdot e^{-120} + e^{-10} \cdot 120e^{-120} = 130 \cdot e^{-130} \approx 4.5254 \cdot 10^{-55}$$

5. Consider various possible values for t, in $P[T \leq t]$, or $P[T > t]$:
   (a) $0 \leq t < 5$: $P[T > t] = e^{-\lambda_1 t}$
   (b) $5 \leq t < 10$: $P[T > t] = P[T > 5, T - 5 > t - 5] = P[T > 5]P[T - 5 > t - 5] = e^{-5\lambda_1} e^{-\lambda_2 (t-5)} = e^{-50+600-120t} = e^{550-120t}$
   (c) $t > 10$ actually is not an acceptable value, since after 10, we "jump" to ∞: $P[T = \infty] = e^{-5\lambda_1 - 120\lambda_2} = e^{-130}$