National Chiao Tung University
Department of Computer Science

DCP1206, Fall 2019: Probability – Final Exam

2020/01/08, 10:10AM-12PM

- This exam contains 6 problems. Total of points is 110.
- Write your solutions in the booklet, and only the solutions in the booklet will be graded.
- You are allowed to bring a two-sided, A4 formula sheet.
- When giving a formula for an MGF/CDF/PMF/PDF, make sure to specify the range over which the formula holds.
- For Problem 1, please mark each subproblem as 'O' or 'X'.
- Problem 1 (true/false questions) will receive no partial credit. Partial credit for Problems 2-6 will be awarded.
- Good luck! May the force be with you.

1. (28 points) Please mark whether each of the following arguments is true (O) or false (X). If you mark a subproblem as false (X), please briefly explain the reason or provide a counterexample. If you mark a subproblem as true (O), then no justification is required.

   (a) (4 points) Let X and Y be two random variables. Moreover, let $g(\cdot)$ and $h(\cdot)$ be two real-valued functions of X and Y, respectively. If X and Y are independent, then $g(X)$ and $h(Y)$ are also independent.

   (b) (4 points) Let $X_1 \sim \mathcal{N}(\mu = 1, \sigma^2 = 2)$ and $X_2 \sim \mathcal{N}(\mu = -1, \sigma^2 = 2)$ be two normal random variables. Then, $X_1 + X_2$ is also normal with mean 0 and variance 4.

   (c) (4 points) Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with the following PDF:

   $$f(x) = \begin{cases} 1 - |x|, & \text{if } x \in (-1, 1) \\ 0, & \text{else} \end{cases}$$

   For every $n \in \mathbb{N}$, define $Y_n = \max(X_1, \cdots, X_n)$. Then, Y_n converges to 1, almost surely.

   (d) (4 points) Let Z_1 and Z_2 be two arbitrary normal random variables. Then, Z_1 and Z_2 can always be written as a bivariate normal random variable.

   (e) (4 points) Consider a stationary Markov chain with state space $S = \{0, 1, 2, 3\}$ and the transition matrix P given by

   $$P = \begin{bmatrix} 0.3 & 0.5 & 0.2 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0 & 0.1 & 0.3 & 0.6 \\ 0.7 & 0 & 0 & 0.3 \end{bmatrix}$$

   Then, this Markov chain is aperiodic and irreducible.

   (f) (4 points) Let $X_1, X_2, \cdots$ be a sequence of random variables. If X_n converges to 0 in probability, then we also have $\lim_{n\to\infty} E[X_n^4] = 0$.

   (g) (4 points) Suppose that X is a random variable with $E[X] = \text{Var}[X] = \mu$. By Chebyshev's inequality, we know $P(X > (\pi + 1)\mu) \leq \frac{1}{\pi^2 \mu}$.

2. (10 points) Let $\bar{X}$ denote the mean of 28 i.i.d. samples from an exponential distribution with mean = 1. Approximate $P(0.95 < \bar{X} < 1.05)$ by using CLT. (Hint: What's the variance?)

3. (20 points) Let X be a *double exponentially distributed* random variable with PDF given by

   $$f(x) = \frac{1}{2} \cdot \exp(-|x|), \; -\infty < x < \infty.$$

   (a) (10 points) Find the MGF of X. Please specify the range in which the MGF exists.

   (b) (10 points) In Homework 3, we have shown that $E[X^{2n}] = (2n)!$ and $E[X^{2n+1}] = 0$, for all $n \in \mathbb{N}$ through integration by parts. Here, please reproduce the same results by using the MGF obtained in (a). (Hint: It might be easier to consider the Taylor expansion of the MGF before taking differentiation)

4. (16 points) Let $X_1, \cdots, X_N$ be non-negative independent random variables with continuous distributions (but $X_1, \cdots, X_N$ are not necessarily identically distributed). Assume that the PDFs of X_i's are uniformly bounded by some constant $C > 0$.
   (a) (6 points) Show that for every i, $E[\exp(-tX_i)] \le \frac{C}{t}$, for all $t > 0$.
   (b) (10 points) By using (a), show that for any $\varepsilon > 0$, we have
   $$P\Big(\sum_{i=1}^{N} X_i \le \varepsilon N\Big) \le (Ce\varepsilon)^N.$$
   Please carefully justify every step of your proof. (Hint: For any $t > 0$, $P(\sum_{i=1}^{N} X_i \le \varepsilon N) = P(e^{t\sum_{i=1}^{N} X_i} \le e^{t\varepsilon N}) = P\big(e^{-t\sum_{i=1}^{N} X_i} \ge e^{-t\varepsilon N}\big))$

5. (18 points)
   (a) (8 points) Consider two sequences of random variables $X_1, X_2, \cdots$ and $Y_1, Y_2, \cdots$ defined on the same sample space. Suppose that X_n converges to a and Y_n converges to b, almost surely. Moreover, suppose the random variables Y_n cannot be equal to zero. Show that X_n/Y_n converges to a/b, almost surely.
   (b) (10 points) Let $U_1, U_2, \cdots$ and $V_1, V_2, \cdots$ be two i.i.d. sequences of random variables. We assume that the U_i's and V_i's have finite mean (denoted by $E[U]$ and $E[V]$, respectively), and that $V_1 + V_2 + \cdots + V_n$ cannot be equal to zero, for every $n \ge 1$. For each $n \ge 1$, define
   $$Z_n = \frac{U_1 + \cdots + U_n}{V_1 + \cdots + V_n}. \tag{1}$$
   Does the sequence Z_n converge to some constant c, almost surely? If so, what is the value of c? Please carefully justify your answer. (Hint: Use SLLN and the result of (a))

6. (18 points) To celebrate the New Year, every Pokemon Go player is able to catch special Pikachus with party hats during the first week of January 2020. The combat power (CP) of each special Pikachu is drawn independently from a normal distribution with an unknown mean μ and a known standard error $\sigma = 100$ (suppose the CP can be any real number). We would like to estimate μ based on the information collected from the caught Pikachus:
   (a) (10 points) Given that 5 Pikachus with CP$= 200, 150, 250, 350, 400$ are caught, what is the likelihood function of μ? Then, what is the maximum likelihood estimator (MLE) of μ? (Hint: For MLE, you could either take differentiation by yourself or simply use the results discussed in the lecture slides)
   (b) (8 points) Next, we use the Bayesian approach to estimate μ with the help of a conjugate prior. Specifically, we choose the prior distribution on μ to be normal with mean $\mu_0 = 300$ and standard error $\sigma_0 = 100$. Given that 5 Pikachus with CP$= 200, 150, 250, 350, 400$ are caught, what is the estimated value of μ under the maximum a posteriori (MAP) criterion? (Hint: Note that posterior $\propto$ prior $\times$ likelihood. To find the parameter that maximizes the posterior, it would be easier to consider the logarithm of the posterior)