Quiz 1: CS 215

Name: ____ **Roll Number:** ________________________

Attempt all five questions, each carrying 10 points. Clearly mark out rough work.

Useful Information

1. Binomial theorem: $(x+y)^n = \sum_{k=0}^{n} C(n,k)x^k y^{n-k}$
2. The empirical mean of n independent and identically distributed random variables is approximately Gaussian distributed. The approximation accuracy is better when n is larger.
3. Defining $\Phi(x) = \int_{-\infty}^{x} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$, we have the following table:

n	$\Phi(n) - \Phi(-n)$
1	68.2%
2	95.4%
2.6	99%
2.8	99.49%
3	99.73%

4. For a non-negative random variable X, we have $P(X \geq a) \leq E(X)/a$ where $a > 0$.
5. For a random variable X with mean μ and variance σ^2, we have $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$.
6. Integration by parts: $\int u dv = uv - \int v du$.
7. Gaussian pdf: $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$
8. Poisson pmf: $P(X=i) = \frac{e^{-\lambda}\lambda^i}{i!}$

Additional space

1. An exponential random variable X has a pdf which is given as $f_X(x) = \lambda e^{-\lambda x}$ where $x \in [0, \infty)$ and $\lambda > 0$. Derive the pmf of floor(X) and ceil(X). Recall that ceil(X) is the smallest integer greater than or equal to X and floor(X) is the largest integer less than or equal to X. [5+5=10 points]
Solution: $P(\text{floor}(X) = n) = P(n \leq X < n+1) = F_X(n+1) - F_X(n) = (1 - e^{-\lambda(n+1)}) - (1 - e^{-\lambda n}) = e^{-\lambda n}(1 - e^{-\lambda})$.
$P(\text{ceil}(X) = n) = P(n-1 \leq X < n) = F_X(n) - F_X(n-1) = (1 - e^{-\lambda(n)}) - (1 - e^{-\lambda(n-1)}) = e^{-\lambda(n-1)}(1 - e^{-\lambda})$.

2. Verify whether true or false with justification (no credit without it): (a) The minimum of n iid Bernoulli random variables is also a Bernoulli random variable. (If your answer is in the affirmative, what is the parameter of the Bernoulli random variable?). (b) The minimum of n iid geometric random variables is also a geometric random variable. (If your answer is in the affirmative, what is the parameter of the geometric random variable?). Recall that the pmf of a geometric random variable has the form $P(X = i) = (1-p)^{i-1}p$. [5+5=10 points]
Solution: Part a:
Let $Y = min\{X_i\}_{i=1}^n$ where $X_i \sim Bernoulli(p)$. Then $P(Y = 1) = \Pi_{i=1}^n P(X_i = 1) = p^n$. Also $P(Y = 0) = 1 - P(Y = 1) = 1 - p^n$. So Y is a Bernoulli random variable with parameter p^n.
Part b:
Let $Y = min\{X_i\}_{i=1}^n$ where $X_i \sim Geometric(p)$. Then $P(Y \geq y) = \Pi_{i=1}^n P(X_i \geq y) = ((1-p)^y)^n = (1-p)^{ny}$. Hence $P(Y < y) = 1 - [(1-p)^{ny}]$. This is the CDF of a geometric random variable with the parameter $1 - (1-p)^n$.

3. If $X \sim \text{Poisson}(\lambda)$, $Y \sim \text{Uniform}(-a, a)$ where $a > 0$ and $Z = X + Y$, derive an expression for $E[(Z - E(Z))^3]$. Assume X and Y are independent. [10 points]
Solution: We have $E(Z) = E(X) + E(Y) = \lambda$. We also have $E(X^2) = \lambda^2 + \lambda$.
Now LHS $= E[(X + Y - \lambda)^3] = E[(X+Y)^3 - \lambda^3 + 3(X+Y)\lambda^2 - 3\lambda(X+Y)^2]$.
Now $E[(X+Y)^3] = E[X^3 + Y^3 + 3X^2Y + 3XY^2]$. Now $E[Y] = E[Y^3] = 0, E[Y^2] = a^2/3$, so we have $E[(X+Y)^3] = E[X^3 + 3XY^2] = E[X^3] + 3\lambda a^2/3 = E[X^3] + a^2\lambda$.
We have $E[X^3] = \sum_{k=0}^{\infty} k^3 \dfrac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k^2 \dfrac{e^{-\lambda}\lambda^k}{(k-1)!}$ which upon replacing k by $l+1$ further yields $\sum_{l=0}^{\infty}(l+1)^2 \dfrac{e^{-\lambda}\lambda^{l+1}}{l!} = \lambda E((X+1)^2) = \lambda(E[X^2 + 2X + 1]) = \lambda^3 + 3\lambda^2 + \lambda$.
Hence $E[(X+Y)^3] = \lambda^3 + 3\lambda^2 + \lambda + \lambda a^2$.
Also $E[3(X+Y)\lambda^2 - 3\lambda(X+Y)^2] = 3\lambda^3 + 3\lambda^2 E[Y] - 3\lambda E[X^2 + Y^2 + 2XY] = 3\lambda^3 - 3\lambda(\lambda^2 + \lambda + a^2/3 + 0) = -3\lambda^2 - \lambda a^2$.
Combining all these results together, we get $E[(X + Y - \lambda)^3] = \lambda^3 + 3\lambda^2 + \lambda + \lambda a^2 - \lambda^3 - 3\lambda^2 - \lambda a^2 = \lambda$.

4. Let $X_1, X_2, ..., X_n$ be independent random variables with the PDF $f_X(x; a, b) = 1/(b-a)$ if $a \leq x \leq b$ and 0 otherwise. Here $a < b$. Derive the maximum likelihood estimate for a for the special case when $b = a+1$. What would be the maximum likelihood estimate of b if the PDF for each of the random variables $X_1, X_2, ..., X_n$ was $f_X(x; 0, b) = 1/b$ if $0 \leq x < b$ and 0 otherwise? Notice the strict inequality in the second sub-question. [7+3=10 points]
Solution: The value of the likelihood function given all the samples is 1 if $\forall i, a \leq x_i \leq a+1$. Hence $a \leq x_{min}$ and $a + 1 \geq x_{max}$, i.e. $a \geq x_{max} - 1$. As the likelihood is constant in the domain $\forall i, a \leq x_i \leq a+1$, we know that any value lying inside the interval $[x_{max} - 1, x_{min}]$ is a maximum likelihood estimate for a. This is an example where the MLE is not unique!
For the second part, we have the constraint that for a non-zero PDF, we must have $x \geq 0$ and $x < b$ (strict inequality). In such a case, we cannot have $b = x_{max}$ unlike the earlier example. Hence we would want to consider an MLE of $x_{max} + \delta$ for an infinitesimally small δ. However for any δ you choose, one can reduce its value (i.e. the value of δ) further and increase the likelihood from $1/(x_{max} + \delta)^n$ to $1/(x_{max} + \delta')^n$ where $\delta' < \delta$. This is a case where the MLE of b is therefore not defined!

5. A storage device contains the monthly expenses of a group $\mathcal{G}$ of n individuals in a country. A computer program has read through these records, and has computed and stored in memory the value $S = \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2$ where x_i is the monthly expenses of the i^{th} individual. Some analysis you wish to perform requires the sample standard deviation of the monthly expenses of the individuals in $\mathcal{G}$. However,

the storage device is incredibly slow and you do not have the option of reading any of the data again. How will you compute the standard deviation given S and n? Derive all required formulae if necessary. [10 points]
Solution: The value of S is actually proportional to the sample variance $\sigma^2 = \sum_i (x_i - m)^2/(n-1)$. To see this, consider that $S = \sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2 = \sum_{i,j} (x_i - m + m - x_j)^2 = \sum_{i,} (x_i - m)^2 + (x_j - m)^2 + 2(x_i - m)(x_j - m) = n \sum_i (x_i - m)^2 + n \sum_j (x_j - m)^2 + 0$. Here m is the sample mean. Note that the cross-term in the summation above is 0 because $\sum_i (x_i - m) = 0$ by definition of m. Hence we have $S = 2n \sum_i (x_i - m)^2 = 2n(n-1)\sigma^2$. Hence $\sigma = \sqrt{S/(2n(n-1))}$.