Makeup 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 with finite variance 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. Consider a random variable Z that takes on the values of another random variable X with probability p and the values of random variable Y with probability $1-p$. Let $f_X(.)$ and $f_Y(.)$ be the PDFs of X and Y respectively. Now, students ABC and PQR are having a debate about a procedure to draw a sample value (call it z) of Z.
ABC proposes the following: Draw a sample of X from $f_X(.)$. Call it x_1. Draw a sample of Y from $f_Y(.)$. Call it y_1. Draw a number q from Uniform$(0,1)$. If $q < p$, then $z = x_1$ otherwise $z = y_1$.
PQR instead proposes the following: Draw a sample of X from $f_X(.)$. Call it x_1. Draw a sample of Y from $f_Y(.)$. Call it y_1. Then $z = px_1 + (1-p)y_1$.
Which of these procedures is/are correct? Justify the correctness or lack thereof. [5+5=10 points]
Solution: PQR's procedure is incorrect, because the z values thus drawn would belong to a PDF g which is equal to the convolution of pf_X and $(1-p)f_Y$, i.e. $g(z) = \int_{-\infty}^{+\infty} pf_X(z-t)(1-p)f_Y(t)dt$. On the other hand, the PDF of Z is given by $f_Z(z) = pf_X(z) + (1-p)f_Y(z)$. ABC's procedure is correct because with probability p, the values would be drawn from f_X and with probability $(1-p)$, they would be drawn from f_Y. This is the definition of a mixture of distributions.

2. The entropy of a discrete random variable X is defined as $H(X) = -\sum_{i=1}^{K} p_i \log p_i$ where $p_i = P(X=i)$ and $\sum_{i=1}^{K} p_i = 1; \forall i, 0 \leq p_i \leq 1$. In this definition, $0 \log 0$ is considered to be 0. For which PMF (i.e. for what values of $\{p_i\}_{i=1}^{K}$) will the entropy be maximum? What is this maximum value? Derive your answer by setting the first derivatives of the entropy to 0. Obtain the sign of the second derivatives, i.e. sign of $\frac{\partial^2 H}{\partial p_i^2}$. (Note: Given what you have learned so far, you will be able to find only a local maximum. But it turns out that the unexpectedness measure is a concave function, due to which a local maximum is also the global maximum. You are not expected to prove that it is a concave function.) For what PMF, will the entropy be the least? Give an intuitive answer (it is not so easy to prove your answer for the minimum, in a quiz/exam). What is this minimum value? [5+1+2+1+1=10 points]
Solution: We have $H(X) = -\sum_{i=1}^{K-1} p_i \log p_i - p_K \log p_K = -\sum_{i=1}^{K-1} p_i \log p_i + (1 - \sum_{i=1}^{K-1} p_i \log p_i) \log(1 - \sum_{i=1}^{K-1} p_i \log p_i)$. Hence, we have for $j \in \{1, ..., K-1\}$ that $\dfrac{\partial H(X)}{\partial p_j} = -(1+\log p_j) - (-1 - \log(1 - \sum_{i=1}^{K-1} p_i)) = -\log p_j + \log(1 - \sum_{i=1}^{K-1} p_j)$. Setting this derivative to zero, we have $p_j = 1 - \sum_{i=1}^{K-1} p_j = p_K$. Thus, all the p_j values are equal to $1/K$, i.e. we have a discrete uniform PMD. The second derivative is given as $\dfrac{\partial^2 H(X)}{\partial p_j^2} = -1/p_j - 1/(1 - \sum_{j=1}^{K-1} p_j)$ which is clearly less than 0, and thus the second derivative test is passed. Thus, a discrete uniform PMF maximizes the entropy. The least possible entropy value is 0, which occurs if $p_j = 1$ for some $j \in \{1, ..., K\}$ with the other values being all 0. This is called a Kronecker delta function. The entropy is always non-negative and because $0 \leq p_i \leq 1$ and hence $-\log p_i$ is always positive.

3. Let Y be a Gaussian random variable with mean μ and variance σ^2. Derive the CDF and PDF of the random variable $X = |Y|$. Also derive $E(X^2)$. [4+3+3=10 points]
Solution: We have $F_X(x) = P(X \leq x) = P(|Y| \leq x) = P(-x \leq Y \leq x) = \int_{-x}^{+x} f_Y(y)dy = F_Y(x) - F_Y(-x) = \Phi((x-\mu)/\sigma) - \Phi((-x-\mu)/\sigma)$ where Φ stands for the CDF of a zero-mean Gaussian random variable with unit variance. The PDF of X is given by $f_X(x) = \frac{1}{\sigma}[\phi((x-\mu)/\sigma) + \phi((-x-\mu)/\sigma)] = \frac{1}{\sigma}[\phi((x-\mu)/\sigma) + \phi((x+\mu)/\sigma)]$ where ϕ stands for the PDF of a zero-mean Gaussian random variable with unit variance, and where we use the symmetry of ϕ. This yields $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}[\exp\left(-(x-\mu)^2/2\sigma^2\right) + \exp\left(-(x+\mu)^2/2\sigma^2\right)]$.

4. Let $\{X_i\}_{i=1}^{n}$ be n i.i.d. random variables from a Gaussian distribution with mean μ and variance μ^2. Determine the maximum likelihood estimate of μ. [3.5+3.5+3=10 points]
Solution: The negative log likelihood is given by $NLL(\mu) = \sum_{i=1}^{n} \frac{(x_i-\mu)^2}{2\mu^2} + \log \mu +$ constants. Taking the derivative and setting it to zero, we get $\sum_{i=1}^{n}(\dfrac{-x_i^2}{\mu^3} + \dfrac{x_i}{\mu^2} + \dfrac{1}{\mu} = 0$. Multiplying throughout by μ^2, we obtain $\mu^2 + \mu \sum_{i=1}^{N} x_i/N - \sum_{i=1}^{N} x_i^2/N = 0$. This yields $\mu = \dfrac{-S \pm \sqrt{S^2 + 4S_2}}{2}$ where $S = \sum_{i=1}^{N} x_i/N$ and $S_2 = \sum_{i=1}^{N} x_i^2/N$.

5. 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: $P(\text{floor}(X) = n) = P(n \leq X < n+1) = F_X(n+1) - F_X(n) = (1 - e^{-\lambda(n+1)}) - (1 - e^{-\lambda n}) = e^{-\lambda n}(1 - e^{-\lambda})$.
$P(\text{ceil}(X) = n) = P(n-1 \leq X < n) = F_X(n) - F_X(n-1) = (1 - e^{-\lambda(n)}) - (1 - e^{-\lambda(n-1)}) = e^{-\lambda(n-1)}(1 - e^{-\lambda})$.