Quiz 1: CS 215

Name: ______________________________ **Roll Number:** ____________________

Attempt all five questions, each carrying 10 points. Clearly mark out rough work. You may directly use results/theorems that we derived in class or in homework - you do not need to prove them afresh.

Useful Information

1. Binomial theorem: $(x+y)^n = \sum_{k=0}^{n} C(n,k) x^k y^{n-k}$
2. 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%

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

1. Let $X_1, X_2, ..., X_n$ be independent random variables from $\mathcal{N}(\mu, \sigma^2)$. Show that the random variable $\bar{X} = \frac{X_1 + X_2 + ... + X_n}{n}$ is also Gaussian distributed. (Note: you cannot invoke the central limit theorem here as it would yield only approximate Gaussianity.) What is its mean and variance? [7+3=10 points]
Solution: The MGF of $\bar{X}$ is $\phi_{\bar{X}}(t) = \prod_{i=1}^{n} \phi_{X_i}(t/n) = \prod_{i=1}^{n} e^{\mu t/n + \sigma^2 t^2/(2n^2)} = e^{\mu t + \sigma^2 t^2/(2n)}$. The latter is clearly the MGF of a Gaussian random variable with mean μ and variance σ^2/n. By uniqueness of MGF, the assertion is proved.

2. If $X \sim$ Poisson(λ), $Y \sim \mathcal{N}(0, \sigma^2)$ 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 $E[Y] = E[Y^3] = 0, E[Y^2] = \sigma^2$, so we have $E[(X+Y)^3] = E[X^3 + 3XY^2] = E[X^3] + 3\lambda\sigma^2$. We have $E[X^3] = \sum_{k=0}^{\infty} k^3 \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{(k-1)!}$ which upon replacing k by $l+1$ further yields $\sum_{l=0}^{\infty} (l+1)^2 \frac{e^{-\lambda}\lambda^{l+1}}{l!} = \lambda E((X+1)^2) = \lambda(E[X^2 + 2X + 1]) =$

$\lambda^3 + 3\lambda^2 + \lambda$. Hence $E[(X+Y)^3] = \lambda^3 + 3\lambda^2 + \lambda + 3\lambda\sigma^2$.
Also $E[3(X+Y)\lambda^2 - 3\lambda(X+Y)^2] = 3\lambda^3 + 3\lambda^2 E[Y] - 3\lambda E[X^2 + Y^2 + 2XY] = 3\lambda^3 - 3\lambda(\lambda^2 + \lambda + \sigma^2 + 0) = -3\lambda^2 - 3\lambda\sigma^2$.
Combining all these results together, we get $E[(X+Y-\lambda)^3] = \lambda^3 + 3\lambda^2 + \lambda + 3\lambda\sigma^2 - \lambda^3 - 3\lambda^2 - 3\lambda\sigma^2 = \lambda$.

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

4. Consider sample values $x'_1, x'_2, ..., x'_n$ respectively from $n > 0$ iid random variables $X_1, X_2, ..., X_n$, each having the CDF $F_X(x)$. Then the so-called empirical CDF is defined as $F_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}(x'_i \leq x)$ where $\mathbf{1}(q)$ is the indicator function which produces 1 if the predicate q is true, and 0 otherwise. Prove that $F_n(x)$ is an unbiased estimate of $F_X(x)$ and derive its variance. Hence prove that $\lim_{n\to\infty} E([F_n(x) - F_X(x)]^2) = 0$. [3+4+3=10 points]
Solution: We have $E[F_n(x)] = E[\frac{1}{n}\sum_{i=1}^n \mathbf{1}(X_i \leq x)] = \frac{1}{n}\sum_{i=1}^n P(X_i \leq x) = \frac{1}{n}\sum_{i=1}^n F_X(x) = F_X(x)$. Hence, it is an unbiased estimate. Note that for all i from 1 to n, the random variables $\mathbf{1}(X_i \leq x)$ are Bernoulli distributed with success parameter $F_X(x)$.
Now $Var(F_n(x)) = Var(\frac{1}{n}\sum_{i=1}^n \mathbf{1}(X_i \leq x)) = \frac{1}{n^2} \times nP(X \leq x)(1 - P(X \leq x) = F_X(x)(1 - F_X(x))/n$. This step made use of the independence of $\mathbf{1}(X_i \leq x)$.
Alos $\lim_{n\to\infty} E[(F_n(x) - F_X(x))^2] = \lim_{n\to\infty} E[(F_n(x) - E(F_n(x)))^2] = \lim_{n\to\infty} Var(F_n(x)) = 0$ from the earlier expression for variance.

5. Consider a sequence of independent Bernoulli trials $X_1, X_2, ...$ each with success probability p. Let N be a random variable that denotes the trial number of the first success. Derive an expression for $P(N > n)$ and $E(N)$. [7+3=10 points]
Solution: We have $P(N = n) = p(1-p)^{n-1}$ since the first $n-1$ trials were failures and the n^{th} trial was a success. $N > n$ implies the first n trials are all failures. Hence $P(N > n) = (1-p)^n$. Hence $P(N \leq n) = 1 - (1-p)^n$. Also we have $E(N) = \sum_{n=1}^\infty np(1-p)^{n-1} = p\sum_{n=1}^\infty n(1-p)^{n-1} = p\sum_{n=1}^\infty -\frac{d}{dp}(1-p)^n = -p\frac{d}{dp}\sum_{n=0}^\infty (1-p)^n = -p\frac{d}{dp}(1/p) = \frac{1}{p}$.