Q1:

Suppose I gather some n independent measurements of a quantity. Let us treat the measurements as independent random variables with mean μ and standard deviation σ. If I want to be 99% certain that the average of these measurements is accurate to within $\pm\frac{\sigma}{4}$ units, how many measurements must I take, i.e. what is the value of n? [10 points]

Solution:

We know that $\bar{X}$, the average of the measurements, will have mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ and has an approximately Gaussian distribution. Hence $\sqrt{n}\frac{\bar{X}-\mu}{\sigma}$ is a standard normal random variable. We want $\mu - \frac{\sigma}{4} \leq \bar{X} \leq \mu + \frac{\sigma}{4}$, i.e. $-\frac{\sigma}{4} \leq \bar{X} - \mu \leq \frac{\sigma}{4}$, i.e. $-\frac{\sqrt{n}}{4} \leq \sqrt{n}(\frac{\bar{X}-\mu}{\sigma}) \leq \frac{\sqrt{n}}{4}$. Now $P(-\frac{\sqrt{n}}{4} \leq \sqrt{n}(\frac{\bar{X}-\mu}{\sigma}) \leq \frac{\sqrt{n}}{4}) \geq 0.99$. Hence from the given table, we must have $\sqrt{n}/4 \geq 2.7$, i.e. $\sqrt{n} \geq 10.8$, i.e. $n \geq 117$.

Q2:

Show that the sum of two independent Poisson random variables with mean λ_1 and λ_2 respectively is another Poisson random variable. What is its mean? Show all steps clearly. [10 points]

Solution:

$$
\begin{aligned}
P(Z=k) &= \sum_{l=0}^{k} P(X=l)P(Y=k-l) && (1)\\
&= \sum_{l=0}^{k} \frac{\lambda_1^l e^{-\lambda_1}}{l!}\frac{\lambda_2^{k-l} e^{-\lambda_2}}{(k-l)!} && (2)\\
&= e^{-\lambda_1-\lambda_2}\sum_{l=0}^{k} \frac{\lambda_1^l}{l!}\frac{\lambda_2^{k-l}}{(k-l)!} && (3)\\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{l=0}^{k} k!\frac{\lambda_1^l}{l!}\frac{\lambda_2^{k-l}}{(k-l)!} && (4)\\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{l=0}^{k} C(k,l)\lambda_1^l\lambda_2^{k-l} && (5)\\
&= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}(\lambda_1+\lambda_2)^k && (6)\\
& && (7)
\end{aligned}
$$

which is a Poisson pmf with parameter $\lambda_1 + \lambda_2$.

Q3:

Let X and Y be independent random variables on a probability space $(\Omega, \Sigma, \Pr)$.

Let g and h be real-valued functions defined on the codomains of X and Y respectively.

Then $g(X)$ and $h(Y)$ are independent random variables.

Solution:
https://proofwiki.org/wiki/Functions_of_Independent_Random_Variables_are_Independent

Q4:

Let X be a continuous random variable with a strictly increasing distribution function $F_X(x)$. Consider random variables $Y_1 = F_X(X)$ and $Y_2 = F_X^{-1}(U)$ where U is a random variable from a $(0,1)$ uniform distribution and F_X^{-1} is the inverse of the function F_X. Derive the CDF of Y_1 and the CDF of Y_2. Hence write down a simple procedure to draw a sample from a Gaussian distribution with mean 0 and standard deviation 1 assuming you have access to a ready-made algorithm to draw a sample from Uniform$(0,1)$. [10 points]

Solution: see next page

$$P(X \le x) = P(F_X(X) \le F_X(x))$$
$$= P(Y \le F_X(x)) = F_X(x)$$

$\therefore$ Y is uniformly distributed since for any $[0,1]$ uniform random variable Z

$$P(Z \le z) = z.$$

$$P(F_X^{-1}(U) \le x) = P(U \le F_X(x)) = F_X(x).$$

$$\therefore P(F_X^{-1}(U) \le x) = F_X(x)$$

$\therefore$ $F_X^{-1}(U) = Y_2$ is a random variable with distribution F_X.

This gives rise to an elegant procedure to draw a sample from any distribution F_X.

1. Generate a random sample $u \sim \text{Uniform}(0,1)$
2. Compute $F_X^{-1}(u) = \tilde{x}$
3. $\tilde{x}$ is distributed as per F_X.

For a Gaussian distribution, tables for $F_X^{-1}(u)$ are readily available

Q5:

Consider that $\mathbf{X} = \{X_1, X_2, ..., X_n\}$ is a set of n independent random variables from an underlying distribution $F_X(x)$. Define $X_{(k)}$ to be the k^{th} smallest element in $\mathbf{X}$. Derive the CDF of $X_{(k)}$ in terms of $F_X(x)$. Define $N_x = \sum_{i=1} \mathbf{1}(X_i \le x)$ where $\mathbf{1}(y) = 1$ for $y =$ true and 0 otherwise. Derive the CDF of N_x. How does it relate to the CDF of $X_{(k)}$? [10 points]

See solution on next page:

Q3 $N_x = \sum_{i=1}^{n} \underbrace{\mathbb{1}(X_i \le x)}$

↳ Bernoulli random variables (independent) with parameter $p = F_X(x)$

$\therefore N_x$ is a binomial random variable.

$$\therefore P(N_x \ge k) = \sum_{j=k}^{n} C(n,j)\left(F_X(x)\right)^j \left(1 - F_X(x)\right)^{n-j}$$

$$P(X_{(k)} \le x) = P(N_x \ge k) = 1 - P(N_x < k)$$

because $X_{(k)} \le x$ implies $N_x \ge k$ and conversely

$$\therefore F_{X(k)}(x) = 1 - F_{N_x}(k)$$

$$f_{X(k)}(x) = \sum_{j=k}^{n} C(n,j)\left[j\left(F_X(x)\right)^{j-1} f_X(x)\left(1 - F_X(x)\right)^{n-j} + \left(F_X(x)\right)^j (n-j)\left(1 - F_X(x)\right)^{n-j-1} (-1) f_X(x)\right]$$

$$= \sum_{j=k}^{n} C(n,j)\, F_X(x)^{j-1} \left(1 - F_X(x)\right)^{n-j-1} f_X(x) \left[j\left(1 - F_X(x)\right) + F_X(x)(n-j)(-1)\right]$$

$$= \sum_{j=k}^{n} C(n,j)\, f_X(x)\, F_X(x)^{j-1} \left(1 - F_X(x)\right)^{n-j-1} \left(j - \cancel{j F_X(x)} - n F_X(x) + \cancel{j F_X(x)}\right)$$

From the desk of DATE

$$= \sum_{j=k}^{n} j\, C(n,j)\, f_X(x)\, F_X(x)^{j-1} \left(1 - F_X(x)\right)^{n-j-1}$$

$$- \sum_{j=k}^{n} n\, C(n,j)\, f_X(x)\, F_X(x)^{j} \left(1 - F_X(x)\right)^{n-j-1}$$