Problem 1 (Series).

(a) Let $S_n = \sum_{i=0}^{n} \alpha^i$. Then $\alpha S_n = \sum_{i=0}^{n} \alpha^{i+1} = \sum_{i=1}^{n+1} \alpha^i$. Subtracting one equation from another, $S_n - \alpha S_n = \sum_{i=0}^{n} \alpha^i - \sum_{i=1}^{n+1} \alpha^i = 1 - \alpha^{n+1}$. Therefore, $S_n = (1-\alpha^{n+1})/(1-\alpha)$.

(b) $\sum_{i=0}^{\infty} \alpha^i = \lim_{n\to\infty} S_n$. Limit converges for $|\alpha| < 1$ to $1/(1-\alpha)$.

(c) We know that $\sum_{i=0}^{\infty} \alpha^i = 1/(1-\alpha)$. Differentiating with respect to α on both sides, we get $\sum_{i=1}^{\infty} i\alpha^{i-1} = 1/(1-\alpha)^2$. Multiplying by α on both sides, we have $\sum_{i=1}^{\infty} i\alpha^i = \alpha/(1-\alpha)^2$.

Problem 2 (Bayes' Theorem). The information can be placed into a joint probability distribution function:

Company	Defective	Good	Total
A	$0.05 * 0.50 = 0.025$	$0.50 - 0.025 = 0.475$	0.50
B	$0.07 * 0.30 = 0.021$	$0.30 - 0.021 = 0.279$	0.30
C	$0.10 * 0.20 = 0.020$	$0.20 - 0.020 = 0.180$	0.20
Total	0.066	0.934	1.00

(a) $\Pr\{\text{Defective}\} = 0.066$.

(b) $\Pr\{\text{Company B}|\text{Defective}\} = \Pr\{\text{Company B and Defective}\}/\Pr\{\text{Defective}\} = 0.021/0.066 \approx 0.318$.

(c) No. If they were, then $\Pr\{\text{Company B}|\text{Defective}\} = 0.318$ would have to equal $\Pr\{\text{Company B}\}$, but it does not.

Problem 3 (Probability Distributions).

(a) (Geometric Distribution)

(i) $E[X] = \sum_{t=0}^{\infty} tp(1-p)^t$. From Problem 1(c) we know that $\sum_{t=1}^{\infty} t(1-p)^t = (1-p)/(1-(1-p))^2 = (1-p)/p^2$. Therefore, $E[X] = (1-p)/p$.

$Var[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$. $E[X^2] = \sum_{t=0}^{\infty} t^2 p(1-p)^t$. We know that $\sum_{t=1}^{\infty} t(1-p)^t = (1-p)/p^2$. Differentiating both sides with respect to $(1-p)$, we get $\sum_{t=1}^{\infty} t^2(1-p)^{t-1} = (p^2 + 2p(1-p))/p^4 = (2-p)/p^3$. Multiplying both sides by $(1-p)$, we get $\sum_{t=1}^{\infty} t^2(1-p)^t = (2-p)(1-p)/p^3$. Therefore, $E[X^2] = (2-p)(1-p)/p^2$, and $Var[X] = (2-p)(1-p)/p^2 - (1-p)^2/p^2 = (1-p)/p^2$.

(ii) $\Pr\{X \le t\} = \sum_{i=0}^{t} \Pr\{X = i\} = \sum_{i=0}^{t} p(1-p)^i = p(1-(1-p)^{t+1})/(1-(1-p)) = 1-(1-p)^{t+1}$.

(iii) $\Pr\{X \ge t\} = 1 - \Pr\{X \le t-1\} = (1-p)^t$. Similarly, $\Pr\{X > s\} = (1-p)^{s+1}$ and $\Pr\{X > s+t\} = (1-p)^{s+t+1}$. Therefore, $\Pr\{X > s+t\} = \Pr\{X > s\}\Pr\{X \ge t\}$. But $\Pr\{X > s+t\} = \Pr\{X > s+t, X \ge t\} = \Pr\{X \ge t\}\Pr\{X > s+t|X \ge t\}$. Using the the two previous equations, we get $\Pr\{X > s+t|X \ge t\} = \Pr\{X > s\}$.

(iv) For a discrete memoryless random variable,

$$\begin{aligned}\Pr\{X > s+t\} &= \Pr\{X > s\}\Pr\{X \geq t\}, \\ \text{and } \Pr\{X > s+t-1\} &= \Pr\{X > s-1\}\Pr\{X \geq t\}.\end{aligned}$$

Subtracting one from another and using the fact that $\Pr\{X > s\} - \Pr\{X > s-1\} = \Pr\{X = s\}$, we get

$$\Pr\{X = s+t\} = \Pr\{X = s\}\Pr\{X \geq t\}.$$

Now,

$$\begin{aligned}E[X-t|X \geq t] &= \sum_{x=0}^{\infty}(x-t)\frac{\Pr\{X-t = x-t, X \geq t\}}{\Pr\{X \geq t\}} \\ &= \sum_{x=t}^{\infty}(x-t)\frac{\Pr\{X = x\}}{\Pr\{X \geq t\}} \\ &= \sum_{i=0}^{\infty} i\frac{\Pr\{X = i+t\}}{\Pr\{X \geq t\}} \\ &= \sum_{i=0}^{\infty} i\frac{\Pr\{X = i\}\Pr\{X \geq t\}}{\Pr\{X \geq t\}} \qquad \text{(memoryless property)} \\ &= \sum_{i=0}^{\infty} i\Pr\{X = i\} = E[X].\end{aligned}$$

This exercise can be similarly repeated for a continuous random variable X with the memoryless property by using integrals instead of sums.

(b) (Binomial Distribution) The expectation can be computed as follows:

$$\begin{aligned}E[X] &= \sum_{n=0}^{N} n\binom{N}{n}p^n(1-p)^{N-n} \\ &= \sum_{n=0}^{N} n\frac{N!}{(N-n)!n!}p^n(1-p)^{N-n} \\ &= \sum_{n=0}^{N} \frac{N!}{(N-n)!(n-1)!}p^n(1-p)^{N-n} \\ &= Np\sum_{n=0}^{N} \frac{(N-1)!}{(N-n)!n!}p^{n-1}(1-p)^{N-n} \\ &= Np\sum_{n=0}^{N} \binom{N-1}{n-1}p^{n-1}(1-p)^{N-n} \\ &= Np.\end{aligned}$$

It can also be obtained much simpler by noting that a binomially distributed random variable X can be written as a sum of N independent and identically distributed binary random variables $X_1, \cdots, X_N$ such that $\Pr\{X_i = 1\} = p$ and $\Pr\{X_i = 0\} = (1-p)$ for $i = 1, \cdots, N$. As $E[X_i] = p$, $E[X] = E[\sum_{i=1}^{N} X_i] = \sum_{i=1}^{N} E[X_i] = Np$. Similarly, using the fact that $Var[X_i] = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) = p(1-p)$ and $Var[X] = \sum_{i=1}^{N} Var[X_i]$ (because X_i are mutually independent), we get $Var[X] = Np(1-p)$.

(c) (Poisson Distribution)

(i) The expectation can be computed as follows:

$$\begin{aligned} E[X] &= \sum_{n=1}^{\infty} n e^{-\lambda} \frac{\lambda^n}{n!} \\ &= \lambda e^{-\lambda} \underbrace{\sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!}}_{=e^{\lambda}} \\ &= \lambda. \end{aligned}$$

The variance can be computed as follows:

$$\begin{aligned} Var[X] &= E[X^2] - E[X]^2 \\ &= \sum_{n=1}^{\infty} n^2 e^{-\lambda} \frac{\lambda^n}{n!} - \lambda^2 \\ &= \sum_{n=1}^{\infty} (n(n-1)+n) e^{-\lambda} \frac{\lambda^n}{n!} - \lambda^2 \\ &= \sum_{n=2}^{\infty} n(n-1) e^{-\lambda} \frac{\lambda^n}{n!} + \underbrace{\sum_{n=1}^{\infty} n e^{-\lambda} \frac{\lambda^n}{n!}}_{=E[X]=\lambda} - \lambda^2 \\ &= \lambda^2 e^{-\lambda} \underbrace{\sum_{n=2}^{\infty} \frac{\lambda^{n-2}}{(n-2)!}}_{=e^{\lambda}} + \lambda - \lambda^2 \\ &= \lambda. \end{aligned}$$

(ii)

$$\begin{aligned} \lim_{N\to\infty} \Pr\{X = n|N\} &= \lim_{N\to\infty} \binom{N}{n} p^n (1-p)^{N-n} \\ &= \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{n!} \left(\frac{\lambda}{N}\right)^n \left(1 - \frac{\lambda}{N}\right)^{N-n} \\ &= \frac{\lambda^n}{n!} \lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{N^n} \left(1 - \frac{\lambda}{N}\right)^{N-n} \\ &= \frac{\lambda^n}{n!} \underbrace{\lim_{N\to\infty} \frac{N(N-1)\cdots(N-n+1)}{N^n}}_{=1} \underbrace{\lim_{N\to\infty} \left(1 - \frac{\lambda}{N}\right)^{N-n}}_{=e^{-\lambda}} \\ &= e^{-\lambda} \frac{\lambda^n}{n!}. \end{aligned}$$

(d) (Exponential Distribution)

(i) The expectation can be computed as follows:

$$\begin{aligned} E[X] &= \int_{x=0}^{\infty} x\lambda e^{-\lambda x} dx \\ &= -xe^{-\lambda x}|_{x\to\infty} + xe^{-\lambda x}|_{x=0} + \int_{x=0}^{\infty} e^{-\lambda x} dx \\ &= -0 + 0 - \frac{1}{\lambda}\left(e^{-\lambda x}|_{x\to\infty} - e^{-\lambda x}|_{x=0}\right) \\ &= \frac{1}{\lambda}. \end{aligned}$$

The variance can be computed as follows:

$$\begin{aligned} Var[X] &= \int_{x=0}^{\infty} (x-1/\lambda)^2 \lambda e^{-\lambda x} dx \\ &= -(x-1/\lambda)^2 e^{-\lambda x}|_{x\to\infty} + (x-1/\lambda)^2 e^{-\lambda x}|_{x=0} + \int_{x=0}^{\infty} 2(x-1/\lambda)e^{-\lambda x} dx \\ &= \frac{1}{\lambda^2} - \frac{2(x-1/\lambda)}{\lambda} e^{-\lambda x}|_{x\to\infty} + \frac{2(x-1/\lambda)}{\lambda} e^{-\lambda x}|_{x=0} + \int_{x=0}^{\infty} 2e^{-\lambda x} dx \\ &= \frac{1}{\lambda^2} - \frac{2}{\lambda} - \frac{2}{\lambda}\left(e^{-\lambda x}|_{x\to\infty} - e^{-\lambda x}|_{x=0}\right) \\ &= \frac{1}{\lambda^2}. \end{aligned}$$

(ii) $\Pr\{X \geq t\} = \int_t^{\infty} \lambda e^{-\lambda x} dx = e^{-\lambda t}$. Similarly, $\Pr\{X > s\} = e^{-\lambda s}$ and $\Pr\{X > s+t\} = e^{-\lambda(s+t)}$. Therefore, $\Pr\{X > s+t\} = \Pr\{X > s\}\Pr\{X \geq t\}$. As $\Pr\{X > s+t, X \geq t\} = \Pr\{X > s+t\}$, we get $\Pr\{X > s+t | X \geq t\} = \Pr\{X > s\}$. In fact, this is the only continuous distribution that has this memoryless property.

(iii) This can be shown by seeing the complementary cumulative distribution function of $Z = X + Y$ which is given by $\Pr\{\min\{X, Y\} > z\} = \Pr\{X > z, Y > z\} = \Pr\{X > z\}\Pr\{Y > z\} = e^{-(\lambda_X + \lambda_Y)z}$.

(e) (Gaussian Distribution) The expectation can be computed as follows:

$$\begin{aligned} E[X] &= \int_{x=-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx \\ &= \int_{x=-\infty}^{\infty} (x-\mu) \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx + \mu \underbrace{\int_{x=-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx}_{=1} \\ &= -e^{-(x-\mu)^2/(2\sigma^2)} \frac{2\sigma^2}{\sigma\sqrt{2\pi}}\bigg|_{x\to-\infty} + e^{-(x-\mu)^2/(2\sigma^2)} \frac{2\sigma^2}{\sigma\sqrt{2\pi}}\bigg|_{x\to\infty} + \mu \\ &= \mu. \end{aligned}$$

The variance can be computed as follows:

$$
\begin{aligned}
Var[X] &= \int_{x=-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx \\
&= -\frac{(x-\mu)}{2} e^{-(x-\mu)^2/(2\sigma^2)} \frac{2\sigma^2}{\sigma\sqrt{2\pi}}\bigg|_{x\to-\infty} + \frac{(x-\mu)}{2} e^{-(x-\mu)^2/(2\sigma^2)} \frac{2\sigma^2}{\sigma\sqrt{2\pi}}\bigg|_{x\to\infty} \\
&\qquad + \sigma^2 \underbrace{\int_{x=-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} dx}_{=1} \\
&= \sigma^2.
\end{aligned}
$$

(f) (Multivariate Gaussian)

(i) (Matrix preliminaries) An $n \times n$ orthogonal matrix U is a matrix whose inverse is equal to its transpose, that is, $UU^T = U^TU = I_n$, where I_n is the $n \times n$ identity matrix. For any orthogonal matrix $(\det U)^2 = \det(UU^T) = 1$. Therefore, $|\det U| = 1$.

Any real symmetric matrix Σ can be written as $\Sigma = U\Lambda U^T$, where U is an orthogonal matrix and Λ is a diagonal matrix whose diagonal entries are the eigenvalues of Σ. This is known as the spectral theorem. Determinant of Σ is equal to the product of the eigenvalues of the matrix, which is equal to determinant of Λ, that is, $\det \Sigma = \det \Lambda$.

If Σ is also positive definite, all eigenvalues of Σ are strictly positive, and therefore it is non-singular (that is, its inverse exists). Inverse of a positive definite matrix $\Sigma = U\Lambda U^T$ is given by $\Sigma^{-1} = U\Lambda^{-1}U^T$. This can be verified by seeing that $\Sigma\Sigma^{-1} = U\Lambda U^T U\Lambda^{-1}U^T = U\Lambda I_n \Lambda^{-1}U^T = UU^T = I_n$.

Let $\mathbf{Y} = U^T(\mathbf{X} - \mu)$. Then $\mathbf{X} = U\mathbf{Y} + \mu$. Using transformation of random variables, we have

$$p_{\mathbf{Y}}(\mathbf{y}) = p_{\mathbf{X}}(U\mathbf{y} + \mu)|\det J|,$$

where J is the Jacobian, that is, $J_{ij} = \frac{\partial x_i}{\partial y_j}$. As $\frac{\partial x_i}{\partial y_j} = U_{ij}$, this implies that $J = U$. Since U is an orthogonal matrix, $|\det J| = |\det U| = 1$. Therefore, substituting for $p_{\mathbf{X}}(\cdot)$ we get

$$p_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(\det \Sigma)^{\frac{1}{2}}(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(U\mathbf{y})^T\Sigma^{-1}(U\mathbf{y})}.$$

Since Σ is positive definite, the exponent can be reduced as $(U\mathbf{y})^T\Sigma^{-1}(U\mathbf{y}) = (\mathbf{y}^TU^T)(U\Lambda^{-1}U^T)(U\mathbf{y}) = \mathbf{y}^T\Lambda^{-1}\mathbf{y}$. Using the fact that $\det\Sigma = \det\Lambda$, we get

$$p_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{(\det \Lambda)^{\frac{1}{2}}(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\mathbf{y}^T\Lambda^{-1}\mathbf{y}}.$$

As Λ is a diagonal matrix, Λ^{-1} is also a diagonal matrix with diagonal entries equal to inverse of the diagonal entries of Λ, that is, $\Lambda_{ii}^{-1} = 1/\Lambda_{ii}$. The exponent thus becomes $-\frac{1}{2}\mathbf{y}^T\Lambda^{-1}\mathbf{y} = -\frac{1}{2}\sum_{i=1}^n \frac{y_i^2}{\Lambda_{ii}}$. Also, $\det \Lambda = \prod_{i=1}^n \Lambda_{ii}$. Therefore,

$$\begin{aligned} p_{\mathbf{Y}}(\mathbf{y}) &= \frac{1}{(\prod_{i=1}^n \Lambda_{ii})^{\frac{1}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}\sum_{i=1}^n \frac{y_i^2}{\Lambda_{ii}}} \\ &= \prod_{i=1}^n \frac{1}{(2\pi\Lambda_{ii})^{\frac{1}{2}}} e^{-\frac{1}{2}\frac{y_i^2}{\Lambda_{ii}}} \\ &= \prod_{i=1}^n p_{Y_i}(y_i), \end{aligned}$$

where $Y_i \sim \mathcal{N}(0, \Lambda_{ii})$, $i = 1, \cdots, n$, are *independent* zero mean Gaussian random variables. They are independent because their joint distribution $p_{\mathbf{Y}}(\mathbf{y})$ can be expressed as a product of the distribution of the individual random variables $p_{Y_i}(y_i)$ as shown above. As a consequence, any multivariate Gaussian vector can be written as linear combination of independent Gaussian random variables.

Now, the expectation and covariance matrix of $\mathbf{X}$ can be computed as follows:

$$E[\mathbf{X}] = E[U\mathbf{Y} + \mu] = UE[\mathbf{Y}] + \mu = \mu$$

$$E[(\mathbf{X} - \mu)(\mathbf{X} - \mu)^T] = E[(U\mathbf{Y})(U\mathbf{Y})^T] = UE[\mathbf{Y}\mathbf{Y}^T]U^T = U\Lambda U^T = \Sigma.$$

(ii) Any subset $\mathbf{X}'$ of $\mathbf{X}$ is multivariate Gaussian because every linear combination of the elements of $\mathbf{X}'$ is a linear combination of the elements of $\mathbf{X}$ and thus by definition I, $\mathbf{X}'$ is multivariate Gaussian. This can also be proven by using only definition II. However the proof is tedious. Interested readers can find the proof here: http://fourier.eng.hmc.edu/e161/lectures/gaussianprocess/node7.html

(iii) By the previous result, the subset $\mathbf{X}' = (X_i, X_j) \sim \mathcal{N}(\mu', \Sigma')$. If X_i and X_j are uncorrelated, then Σ' is a diagonal matrix. As seen in part (i) of this problem, if the covariance matrix of a multivariate Gaussian is diagonal, its joint probability density function can be written as a product of the probability density functions of the individual random variables. Hence X_i and X_j are independent.