$X \sim Binomial(4, \frac{1}{3})$. The probability that all the rolls result in successes is given by $\mathbb{P}(X = 4)$ – which can now be computed from the Binomial distribution. If we now consider the probability that at least two rolls result in a number at least 5, then that probability is given by $\mathbb{P}(X \geq 2)$.

**Example 1.353** (Negative Binomial RV)**.** Consider a random experiment with two outcomes 'Success' and 'Failure' with probability of success $p \in (0, 1)$. We consider repeating the experiment until we have $r$ successes, with $r$ being a positive integer. Let $X$ denote the number of failures observed till the $r$-th success. Then $X$ is a discrete RV with the support of $X$ being $S_X = \{0, 1, \cdots\}$. Note that for $k \in S_X$, using independence of the trials we have

$$
\begin{aligned}
&\mathbb{P}(X = k) \\
&= \mathbb{P}(\text{there are } k \text{ failures before the } r\text{-th success}) \\
&= \mathbb{P}(\text{first } k + r - 1 \text{ trials result in } r - 1 \text{ successes and the } k + r\text{-th trial results in a success}) \\
&= \mathbb{P}(\text{first } k + r - 1 \text{ trials result in } r - 1 \text{ successes}) \times \mathbb{P}(\text{the } k + r\text{-th trial results in a success}) \\
&= \binom{k + r - 1}{r - 1} p^{r-1} (1 - p)^k \times p \\
&= \binom{k + r - 1}{k} p^r (1 - p)^k.
\end{aligned}
$$

Therefore the p.m.f. of $X$ is given by

$$
f_X(x) = \begin{cases} \binom{x+r-1}{x} p^r (1 - p)^x, & \text{if } x \in S_X, \\ 0, & \text{otherwise.} \end{cases}
$$

In this case, we say $X$ follows the negative Binomial$(r, p)$ distribution or equivalently, $X$ is a negative Binomial $(r, p)$ RV. Here, $r$ denotes the number of successes at which the trials are terminated and $p$ being the probability of success. The MGF can now be computed as follows.

$$
\begin{aligned}
M_X(t) &= \mathbb{E}e^{tX} \\
&= \sum_{k=0}^{\infty} e^{tk} \binom{k + r - 1}{k} p^r (1 - p)^k
\end{aligned}
$$

$$= \sum_{k=0}^{\infty} \binom{k+r-1}{k} p^r \left[(1-p)e^t\right]^k$$
$$= p^r[1-(1-p)e^t]^{-r}, \forall t < -\ln(1-p).$$

Using the MGF, we can compute the mean and variance of $X$ as $\mathbb{E}X = \frac{rq}{p}, Var(X) = \frac{rq}{p^2}$, with $q = 1-p$.

*Remark* 1.354 (Connection between negative Binomial distribution and the Geometric distribution)*.* A negative Binomial$(1,p)$ RV $X$ has the p.m.f.

$$f_X(x) = \begin{cases} p(1-p)^x, \text{ if } x \in \{0,1,\cdots\}, \\ 0, \text{ otherwise.} \end{cases}$$

which is exactly the same as the p.m.f. for the Geometric$(p)$ distribution. Since the p.m.f. of a discrete RV determines the distribution, we conclude that a *Geometric*$(p)$ RV can be identified as the number of failures observed till the first success in independent trials of a random experiment with two outcomes 'Success' and 'Failure' with probability of success $p \in (0,1)$.

**Note 1.355** (No memory property for Geometric Distribution)**.** Let $X \sim$ *Geometric*$(p)$ for some $p \in (0,1)$. For any non-negative integer $n$, we have

$$\mathbb{P}(X \geq n) = \sum_{k=n}^{\infty} p(1-p)^k = p(1-p)^n \sum_{k=0}^{\infty} (1-p)^k = (1-p)^n.$$

Then, for any non-negative integers $m, n$, we have

$$\mathbb{P}(X \geq m+n \mid X \geq m) = \frac{\mathbb{P}(X \geq m+n \text{ and } X \geq m)}{\mathbb{P}(X \geq m)} = \frac{\mathbb{P}(X \geq m+n)}{\mathbb{P}(X \geq m)} = (1-p)^n = \mathbb{P}(X \geq n).$$

Here, the probability of obtaining at least $n$ additional failures (till the first success) beyond the first $m$ or more failures remain the same as in the the probability of obtaining at least $n$ failures till the first success. In the situation where we stress test a device under repeated shocks, if we consider the survival or continued operation of the device under shocks as 'Failures' in our trial and if the number of shocks till the device breaks down follows *Geometric*$(p)$ distribution, then we can interpret that the age of the device (measured in number of shocks observed) has no effect on the remaining lifetime of the device. This property is usually referred to as the 'No memory' property of the Geometric distribution.

**Note 1.356.** See problem set 8 for a similar property for the Exponential distribution.

**Example 1.357.** Let us consider the random experiment of rolling a standard six-sided fair die till we observe an outcome of at least 5. As mentioned in Example 1.352, the probability of success is $\frac{1}{3}$. Since the last roll results in a success, the number $Y$ of rolls required is exactly one more than the number $X$ of failures observed. Here $X \sim Geometric(\frac{1}{3})$. Then, the probability that an outcome of 5 or 6 is observed in the 10-th roll for the first time is given by

$$\mathbb{P}(Y = 10) = \mathbb{P}(X = 9) = \frac{1}{3}\left(\frac{2}{3}\right)^9.$$

If we want to look at $Z$ which is the number of failures observed till 5 or 6 is rolled twice, then $Z$ follows negative Binomial$(2, \frac{1}{3})$. Now, the number of rolls required is $Z + 2$. The probability that 10 rolls are required is given by

$$\mathbb{P}(Z + 2 = 10) = \mathbb{P}(Z = 8) = \binom{9}{8}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^8.$$

**Note 1.358.** Suppose that a box contains $N$ items, out of which $M$ items have been marked/labelled. In our experiment, we consider all labelled items to be identical and the same for all the unlabelled items. If we draw items from the box with replacement, then the probability of drawing a marked/labelled item is $\frac{M}{N}$ does not change between the draws. If we draw $n$ items at random with replacement, then the number $X$ of marked/labelled items follow $Binomial(n, \frac{M}{N})$ distribution. The case where the draws are conducted without replacement is of interest.

**Example 1.359** (Hypergeometric RV)**.** In the setup of Note 1.358, consider drawing $n$ items at random without replacement. Here, the probability of drawing a marked/labelled item may change between the draws and the number $X$ of marked/labelled items in the $n$ drawn items need not follow $Binomial(n, \frac{M}{N})$ distribution. Here, the number of labelled items among the items drawn satisfies the relation

$$0 \leq X \leq \min\{n, M\} \leq N$$

and the number of unlabelled items among the items drawn satisfies the relation

$$0 \leq n - X \leq N - M$$

and hence $X$ is a discrete RV with support $S_X = \{\max\{0, n-(N-M)\}, \max\{0, n-(N-M)\} + 1, \cdots, \min\{n, M\}\}$. The p.m.f. of $X$ is given by

$$f_X(x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, \text{ if } x \in S_X, \\ 0, \text{ otherwise.} \end{cases}$$

In this case, we say $X$ follows the Hypergeometric distribution or equivalently, $X$ is a Hypergeometric RV. This distribution has the three parameters $N, M$ and $n$. Using properties of binomial coefficients, we can compute the factorial moments of $X$ (left as exercise in problem set 9) and using these values we have,

$$\mathbb{E}X = \frac{nM}{N}, \quad Var(X) = \frac{nM}{N^2(N-1)}(N-M)(N-n) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}.$$

**Note 1.360.** In the setup of a Hypergeometric RV, if we consider $p = \frac{M}{N}$ as the probability of success and $n$ as the number of trials, then $\mathbb{E}X$ matches with that of a $Binomial(n, \frac{M}{N})$ RV and $Var(X)$ is close to that of a $Binomial(n, \frac{M}{N})$ RV for small sample sizes $n$.

**Example 1.361.** Suppose that there are multiple boxes each containing 100 electric bulbs and we draw 5 bulbs from each box for testing. If a box contains 10 defective bulbs, then the number $X$ of defective bulbs in the drawn bulbs follows Hypergeometric distribution with parameters $N = 100, M = 10, n = 5$. Here,

$$\mathbb{P}(X = 2) = \frac{\binom{10}{2}\binom{100-10}{5-2}}{\binom{100}{5}}.$$

**Note 1.362.** We continue with the setting of Note 1.358, where a box contains $N$ items, out of which $M$ items have been marked/labelled or are defective. In our experiment, we consider all labelled items to be identical and the same for all the unlabelled items. If we draw items from the box with replacement until the $r$-th defective item is drawn, then the number of draws required can be described in terms of negative Binomial$(r, \frac{M}{N})$ distribution, where the last draw yields the $r$-th defective item (see Example 1.357). The case where the draws are conducted without replacement is of interest.

**Example 1.363** (Negative Hypergeometric RV)**.** In the setting of Note 1.362, consider drawing the items without replacement till the $r$-th defective item is obtained. We then have $1 \leq r \leq M$. Let $X$ be the number of draws required. Then $X$ is a discrete RV with support $S_X = \{r, r+1, \cdots, N\}$. For $k \in S_X$, using independence of the draws we have

$$\begin{aligned}
&\mathbb{P}(X = k) \\
&= \mathbb{P}(\text{first } k-1 \text{ trials result in } r-1 \text{ defective items and the } k\text{-th trial results in a defective item}) \\
&= \mathbb{P}(\text{first } k-1 \text{ trials result in } r-1 \text{ defective items}) \times \mathbb{P}(\text{the } k\text{-th trial results in a defective item}) \\
&= \frac{\binom{M}{r-1}\binom{N-M}{k-r}}{\binom{N}{k-1}} \times \frac{M-(r-1)}{N-(k-1)}.
\end{aligned}$$

Therefore the p.m.f. of $X$ is given by

$$f_X(x) = \begin{cases} \frac{M-(r-1)}{N-(x-1)} \frac{\binom{M}{r-1}\binom{N-M}{x-r}}{\binom{N}{x-1}}, & \text{if } x \in \{r, r+1, \cdots, N\}, \\ 0, & \text{otherwise.} \end{cases}$$

In this case, we say $X$ follows the negative Hypergeometric distribution or equivalently, $X$ is a negative Hypergeometric RV.

We now discuss an example of a discrete random vector.

*Remark* 1.364. While considering a Bernoulli or a Binomial RV, we looked at random experiments with exactly two outcomes. We now consider random experiments with two or more than two outcomes. Suppose a random experiment terminates in one of $k$ possible outcomes $A_1, A_2, \cdots, A_k$ for $k \geq 2$. More generally, we may also consider random experiments which terminate in one of $k$ mutually exclusive and exhaustive events $A_1, A_2, \cdots, A_k$ with $k \geq 2$. Write $p_j = \mathbb{P}(A_j), j = 1, 2, \cdots, k$, which does not change from trial to trial. Then, $p_1 + p_2 + \cdots + p_k = 1$. Suppose $n$ trials are conducted independently and let $X_j, j = 1, 2, \cdots, k$ denote the number of times event $A_j$ has occured, respectively. Then the RVs $X_1, X_2, \cdots, X_k$ satisfy the relation $X_1 + X_2 + \cdots + X_k = n$ and we have

$$\mathbb{P}(X_1 = x_1, \cdots, X_k = x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$$

for $x_1, \cdots, x_k \in \{0, 1, \cdots, n\}$ with $x_1 + \cdots + x_k = n$. The probability is zero otherwise. Removing the redundancy we have the joint p.m.f. of $(X_1, X_2, \cdots, X_{k-1})$ given by

$$\begin{aligned}
&f_{X_1, \cdots, X_{k-1}}(x_1, \cdots, x_{k-1}) \\
&= \frac{n!}{x_1! \cdots x_{k-1}!(n - x_1 - \cdots - x_{k-1})!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}} (1 - p_1 - \cdots - p_{k-1})^{n - x_1 - \cdots - x_{k-1}}
\end{aligned}$$

for $x_1, \cdots, x_k \in \{0, 1, \cdots, n\}$ with $x_1 + \cdots + x_{k-1} \leq n$ and zero otherwise.

**Example 1.365** (Multinomial Distribution)**.** A random vector $X = (X_1, \cdots, X_{k-1})$ is said to follow the Multinomial distribution with parameters $n$ and $p_1, p_2, \cdots, p_k$ if the joint p.m.f. is as in Remark 1.364 above. We now list some properties of the multinomial distribution.

(a) We first compute the joint MGF. For $t_1, t_2, \cdots, t_{k-1} \in \mathbb{R}$,

$$\begin{aligned}
&M_X(t_1, t_2, \cdots, t_{k-1}) \\
&= \mathbb{E} \exp(t_1 X_1 + t_2 X_2 + \cdots + t_{k-1} X_{k-1}) \\
&= \sum_{\substack{x_1, \cdots, x_k \in \{0, 1, \cdots, n\} \\ x_1 + \cdots + x_{k-1} \leq n}} \frac{n! \exp(t_1 x_1 + t_2 x_2 + \cdots + t_{k-1} x_{k-1})}{x_1! \cdots x_{k-1}!(n - x_1 - \cdots - x_{k-1})!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}} p_k^{n - x_1 - \cdots - x_{k-1}} \\
&= \sum_{\substack{x_1, \cdots, x_k \in \{0, 1, \cdots, n\} \\ x_1 + \cdots + x_{k-1} \leq n}} \frac{n!}{x_1! \cdots x_{k-1}!(n - x_1 - \cdots - x_{k-1})!} \left(p_1 e^{t_1}\right)^{x_1} \cdots \left(p_{k-1} e^{t_{k-1}}\right)^{x_{k-1}} p_k^{n - x_1 - \cdots - x_{k-1}} \\
&= \left(p_1 e^{t_1} + p_2 e^{t_2} + \cdots + p_{k-1} e^{t_{k-1}} + p_k\right)^n
\end{aligned}$$

(b) If $t = (t_1, 0, \cdots, 0) \in \mathbb{R}^{k-1}$, then $M_X(t) = \mathbb{E} \exp(t_1 X_1) = M_{X_1}(t_1)$. But, using the above expression for the joint MGF, we have $M_{X_1}(t_1) = M_X(t) = (p_1 e^{t_1} + p_2 + \cdots + p_{k-1} + p_k)^n = (p_1 e^{t_1} + 1 - p_1)^n$. Therefore, $X_1 \sim Binomial(n, p_1)$. Similarly, $X_i \sim Binomial(n, p_i), \forall i = 2, \cdots, k-1$. In particular, $\mathbb{E} X_i = np_i, Var(X_i) = np_i(1 - p_i)$.

(c) For distinct indices $i, j \in \{1, 2, \cdots, k-1\}$,

$$M_{X_i,X_j}(t_i, t_j) = M_X(0, \cdots, 0, t_i, 0, \cdots, 0, t_j, 0, \cdots, 0) = \left(p_i e^{t_i} + p_j e^{t_j} + 1 - p_i - p_j\right)^n, \forall (t_i, t_j) \in \mathbb{R}^2.$$

Therefore $(X_i, X_j)$ follows the trinomial distribution with the parameters $p_i, p_j, 1-p_i-p_j$, i.e. multinomial distribution with the parameters $n = 3$ and $p_i, p_j, 1 - p_i - p_j$.

(d) For distinct indices $i, j \in \{1, 2, \cdots, k-1\}$, consider $t_i = t_j = t \in \mathbb{R}$. Then,

$$M_{X_i+X_j}(t) = M_{X_i,X_j}(t, t) = \left[(p_i + p_j)\, e^t + 1 - (p_i + p_j)\right]^n,$$

which shows $X_i + X_j \sim Binomial(n, p_i + p_j)$. Then $Var(X_i + X_j) = n(p_i + p_j)(1 - p_i - p_j)$. Using the relation

$$Var(X_i + X_j) = Var(X_i) + Var(X_j) + 2Cov(X_i, X_j),$$

we have $Cov(X_i, X_j) = -np_ip_j$. Consequently, the correlation between $X_i$ and $X_j$ is

$$\rho(X_i, X_j) = \frac{Cov(X_i, X_j)}{\sqrt{Var(X_i)\, Var(X_j)}} = -\left(\frac{p_ip_j}{(1 - p_i)(1 - p_j)}\right)^{\frac{1}{2}}.$$

**Note 1.366.** We now look at distributions that arise in practice from random samples. Such distributions are usually referred to as sampling distributions. More specifically, if $X_1, X_2, \cdots, X_n$ is a random sample from $N(\mu, \sigma^2)$ distribution, we shall look at various statistics involving the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$.

**Note 1.367** (Distribution of square of a standard Normal RV)**.** Let $X \sim N(0, 1)$. Recall that the p.d.f. of $X$ is

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \forall x \in \mathbb{R}$$

We consider the distribution of $Y = X^2$ by first computing the MGF. We have,

$$M_Y(t) = \mathbb{E}e^{tX^2} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx^2} e^{\left(-\frac{x^2}{2}\right)}\, dx = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{\infty} e^{\left(t-\frac{1}{2}\right)x^2}\, dx = (1 - 2t)^{-\frac{1}{2}}, \forall t < \frac{1}{2}$$

Comparing with the MGF of the $Gamma(\alpha, \beta)$ distribution, we conclude that $X^2 \sim Gamma(\frac{1}{2}, 2)$.

**Note 1.368.** If $X \sim N(\mu, \sigma^2)$, then $\frac{X-\mu}{\sigma} \sim N(0, 1)$ and hence $\left(\frac{X-\mu}{\sigma}\right)^2 \sim Gamma(\frac{1}{2}, 2)$.

*Remark* 1.369 (Distribution of the sample mean for a random sample from the Normal distribution)*.* If $X_1, X_2, \cdots, X_n$ is a random sample from $N(\mu, \sigma^2)$ distribution, then for $Y = X_1 + X_2 + \cdots + X_n$, using independence of $X_i$'s we have

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t) = \exp(n\mu t + \frac{1}{2} n\sigma^2 t^2)$$

and hence $X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)$. Now, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and consequently, $\sqrt{n}\left(\frac{\bar{X}-\mu}{\sigma}\right) \sim N(0,1)$ and $n\left(\frac{\bar{X}-\mu}{\sigma}\right)^2 \sim Gamma(\frac{1}{2}, 2)$.

**Note 1.370.** Let $X_1, X_2, \cdots, X_n$ be independent RVs with $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \cdots, n$. Then $\left(\frac{X_i - \mu_i}{\sigma_i}\right)^2, i = 1, 2, \cdots, n$ are i.i.d. with the common distribution $Gamma(\frac{1}{2}, 2)$. The independence follows from Remark 1.330(h). Using problem set 8, we have

$$\sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2 \sim Gamma\left(\frac{n}{2}, 2\right).$$

**Definition 1.371** (Chi-Squared distribution with $n$ degrees of freedom)**.** Let $n$ be a positive integer. We refer to the $Gamma\left(\frac{n}{2}, 2\right)$ distribution as the Chi-Squared distribution with $n$ degrees of freedom. If an RV $X$ follows the Chi-Squared distribution with $n$ degrees of freedom, we write $X \sim \chi_n^2$.

**Note 1.372.** Using Note 1.370, we conclude that $\sum_{i=1}^{n}\left(\frac{X_i-\mu_i}{\sigma_i}\right)^2 \sim \chi_n^2$, where $X_1, X_2, \cdots, X_n$ are independent RVs with $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \cdots, n$.

**Note 1.373.** As argued in Note 1.370, using Remark 1.330(h) we conclude that $X + Y \sim \chi_{m+n}^2$, where $X, Y$ are independent RVs with $X \sim \chi_m^2$ and $Y \sim \chi_n^2$.

**Note 1.374.** If $X \sim \chi_n^2$, then using properties of the $Gamma\left(\frac{n}{2}, 2\right)$ distribution, we have $\mathbb{E}X = n, Var(X) = 2n$ and $M_X(t) = (1-2t)^{-\frac{n}{2}}, \forall t < \frac{1}{2}$.

*Remark* 1.375 (Distribution of the sample variance for a random sample from the Normal distribution)*.* Let $X_1, X_2, \cdots, X_n$ be a random sample from $N(\mu, \sigma^2)$ distribution. By looking at the joint MGF of $X_1 - \bar{X}, \cdots, X_n - \bar{X}$ and $\bar{X}$, it can be shown that $\sum_{i=1}^{n}(X_i - \bar{X})^2$ and $\bar{X}$ are independent.

Now,
$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X}+\bar{X}-\mu)^2=\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2+\frac{n(\bar{X}-\mu)^2}{\sigma^2}$$
where $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2 \sim \chi_n^2$ and $\frac{n(\bar{X}-\mu)^2}{\sigma^2} \sim \chi_1^2$. Since, $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2$ and $\frac{n(\bar{X}-\mu)^2}{\sigma^2}$ are independent, we conclude $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \chi_{n-1}^2$. Taking the sample variance as $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$, we conclude that $\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$.

**Note 1.376.** Given a random sample $X_1, X_2, \cdots, X_n$ from $N(\mu,\sigma^2)$ distribution, the sample mean $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \sim N\left(\mu,\frac{\sigma^2}{n}\right)$ and sample variance $S_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ has the property that $\frac{(n-1)S_n^2}{\sigma^2}\sim\chi_{n-1}^2$. The distribution of $\frac{\bar{X}-\mu}{S_n}$ is of interest.

**Definition 1.377** (Student's $t$-distribution with $n$ degrees of freedom)**.** Let $n$ be a positive integer. Let $X \sim N(0,1)$ and $Y \sim \chi_n^2$ be independent RVs. Then,
$$T=\frac{X}{\sqrt{\frac{Y}{n}}}$$
is said to follow the $t$-distribution with $n$ degrees of freedom. In this case, we write $T \sim t_n$. The p.d.f. is given by
$$f_T(t)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\Gamma\left(\frac{1}{2}\right)\sqrt{n}}\left(1+\frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \forall t\in\mathbb{R}.$$
Here, $\mathbb{E}T^k$ exists if $k<n$. Since, the distribution is symmetric about 0 and hence $\mathbb{E}T^k=0$ for all $k$ odd with $k<n$. If $k$ is even and $k<n$, then
$$\mathbb{E}T^k=n^{\frac{k}{2}}\frac{\Gamma\left(\frac{k+1}{2}\right)\Gamma\left(\frac{n-k}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)}.$$
In particular, if $n>2$, then $\mathbb{E}T=0$ and $Var(T)=\frac{n}{n-2}$. The $t$-distribution appears in the tests for statistical significance.

**Note 1.378.** If $X_1, X_2, \cdots, X_n$ is a random sample from $N(\mu,\sigma^2)$ distribution, then $\sqrt{n}\frac{\bar{X}-\mu}{S_n} \sim t_{n-1}$.

**Note 1.379.** Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distribution, respectively. Consider the sample variances $S_1^2 := \frac{1}{m-1}\sum_{i=1}^m (X_i - \bar{X})^2$ and $S_2^2 := \frac{1}{n-1}\sum_{j=1}^n (Y_j - \bar{Y})^2$. The distribution of $\frac{S_1^2}{S_2^2}$ is of interest. Note that $\frac{(m-1)S_1^2}{\sigma_1^2} \sim \chi^2_{m-1}$ and $\frac{(n-1)S_2^2}{\sigma_2^2} \sim \chi^2_{n-1}$.

**Definition 1.380** ($F$-distribution with degrees of freedom $m$ and $n$)**.** Let $m$ and $n$ be positive integers. Let $X \sim \chi^2_m$ and $Y \sim \chi^2_n$ be independent RVs. Then,

$$F = \frac{\frac{X}{m}}{\frac{Y}{n}}$$

is said to follow the $F$-distribution with degrees of freedom $m$ and $n$. In this case, we write $F \sim F_{m,n}$. The p.d.f. is given by

$$f_F(x) = \begin{cases} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \frac{m}{n}\left(\frac{m}{n}x\right)^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}, \text{ if } x > 0, \\ 0, \text{ otherwise.} \end{cases}$$
$$= \begin{cases} \frac{1}{B\left(\frac{m}{2}, \frac{n}{2}\right)} \frac{m}{n}\left(\frac{m}{n}x\right)^{\frac{m}{2}-1}\left(1+\frac{m}{n}x\right)^{-\frac{m+n}{2}}, \text{ if } x > 0, \\ 0, \text{ otherwise.} \end{cases}$$

**Note 1.381.** If $F \sim F_{m,n}$, then $\frac{1}{F} \sim F_{n,m}$.

**Note 1.382.** Let $X_1, X_2, \cdots, X_m$ and $Y_1, Y_2, \cdots, Y_n$ be independent random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$ distribution, respectively. Consider the sample variances $S_1^2 = \frac{1}{m-1}\sum_{i=1}^m (X_i - \bar{X})^2$ and $S_2^2 = \frac{1}{n-1}\sum_{j=1}^n (Y_j - \bar{Y})^2$. The distribution of $\frac{\sigma_2^2 S_1^2}{\sigma_1^2 S_2^2} \sim F_{m-1,n-1}$.

We now discuss an example of a bivariate continuous random vector. Appropriate generalizations to higher dimensions are possible, but we do not discuss that here.

**Definition 1.383** (Bivariate Normal distribution)**.** A bivariate random vector $X = (X_1, X_2)$ is said to follow bivariate Normal distribution $N_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ for $\mu_1 \in \mathbb{R}, \mu_2 \in \mathbb{R}, \sigma_1 > 0, \sigma_2 > 0, \rho \in (-1, 1)$, if the joint p.d.f. is given by

$$f_{X_1,X_2}(x_1, x_2)$$