Math 621: Probability (Graduate)

Darshan Patel

Fall 2017

A discrete random variable (rv) X has a probability mass function (PMF)

$$p(x) := \mathbb{P}(X = x)$$

and cumulative distribution function (CDF)

$$F(x) = \mathbb{P}(X \leq x)$$

. The random variable X has "support"

$$\text{Supp}[X] := \{x : p(x) > 0, x \in \mathbb{R}\}$$

Since X is discrete, $|\text{Supp}(X)| \leq |\mathbb{N}|$.
Support and pmf are related as follows:

$$\sum_{x \in \text{Supp}(X)} p(x) = 1$$

The most fundamental discrete random variable is the Bernoulli:

$$X \sim \text{Bern}(p) := \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

What is p? p is a parameter. Parameters have parameter spaces. For example, $p \in (0, 1)$, thus $p \neq 0$ and $p \neq 1$.

$X \sim \text{Deg}(c) = \{c$ with probability 1
This means that $\text{Deg}(c) = \mathbb{1}_{x=c}$, where $\mathbb{1}_{x=c}$ is an indicator function.

$$\mathbb{1}_A = \begin{cases} 1 & \text{if } A \\ 0 & \text{if } A^c \end{cases}$$

The random variables X_1, X_2 are independent if joint mass function $\mathbb{P}(X_1, X_2) = \mathbb{P}_{X_1}(X_1)\mathbb{P}_{X_2}(X_2)$ for all x_1, x_2 in their supports.

Let $X_1 \stackrel{d}{=} X_2$. The random variables X_1 and X_2 are equal in distribution if $\mathbb{P}_{X_1}(X) = \mathbb{P}_{X_2}(X)$.

Let $X_1, X_2 \overset{iid}{\sim}$. The random variables X_1, X_2 are independent and identically distributed if $X_1, X_2 \overset{iid}{\sim}$ and $X_1 \overset{d}{=} X_2$.

Let $T_2 = X_1 + X_2$ where $X_1, X_2 \overset{iid}{\sim} \text{Bern}(p)$. Then

$$\text{Supp}[T_2] = \{0, 1, 2\} = \text{Supp}[X_1] + \text{Supp}[X_2]$$

In fact,

$$\begin{aligned}
\mathbb{P}_{T_2}(()2) &= p^2 \\
\mathbb{P}_{T_2}(()0) &= (1-p)^2 \\
\mathbb{P}_{T_2}(()1) &= 2p(1-p)
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}_{T_2}(t) &= \sum_{x \in \text{Supp}[X]} \mathbb{P}_{X_1}(x)\mathbb{P}_{X_2}(x) \\
&= \sum_{x \in \{0,1\}} \left[(p^x(1-p)^{1-x})(p^{t-x}(1-p)^{1-t+x})\right] \\
&= p^t \sum_{x \in \{0,1\}} (1-p)^{2-t} \\
&= p^t(1-p)^{2-t} \sum_{x \in \{0,1\}} 1 \\
&= 2p^t(1-p)^{2-t}
\end{aligned}$$

But this is wrong because $\mathbb{P}_{T_2}(2) = 2p^2 \neq p^2$.

Let

$$\begin{aligned}
p(t) = \mathbb{P}(T_2 = t) &= \sum_{x \in \text{Supp}[X]} \mathbb{P}_{X_1}(x)\mathbb{P}_{X_2}(t-x) \\
&= \sum_{x \in \{0,1\}} p^x(1-p)^{1-x}\mathbb{1}_{x \in \{0,1\}} p^{t-x}(1-p)^{1-t+x}\mathbb{1}_{t-x \in \{0,1\}} \\
&= p^t(1-p)^{2-t} \sum_{x \in \{0,1} \mathbb{1}_{x \in \{0,1\}}\mathbb{1}_{t-x \in \{0,1\}} \\
&= p^t(1-p)^{2-t}\left(\underbrace{\mathbb{1}_{0 \in \{0,1}}_{1}\mathbb{1}_{t- \in \{0,1\}} + \underbrace{\mathbb{1}_{1 \in \{0,1\}}}_{1}\mathbb{1}_{t-1 \in \{0,1\}}\right) \\
&= p^t(1-p)^{2-t}\left(\mathbb{1}_{t \in \{0,1\}} + \mathbb{1}_{t-1 \in \{0,1}\right) \\
&= \binom{2}{t}p^t(1-p)^{2-t}
\end{aligned}$$

This equation does satisfy $p(0), p(1), p(2)$.

Let $X \sim \text{Bern}(p) = \text{Binom}(1, p) = \binom{1}{x}p^x(1-p)^{1-x}$. Now $\binom{n}{k}$ is only valid with $k \leq n$;

otherwise, it's 0. Now back to $\mathbb{P}_{T_2}(t)$.

$$\begin{aligned}
\mathbb{P}(T_2 = t) &= \sum_{x \in \text{Supp}[X]} \mathbb{P}_{X_1}(x)\mathbb{P}_{X_2}(t-x) \\
&= \sum_{x \in \{0,1\}} \binom{1}{x} p^x (1-p)^{1-x} \binom{1}{t-x} p^{t-x}(1-p)^{1-t+x} \\
&= pt(1-p)^{2-t} \sum_{x \in \{0,1\}} \binom{1}{x}\binom{1}{t-x} \\
&= \binom{2}{t} p^t (1-p)^{2-t} \text{ by } \binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}
\end{aligned}$$

Convolution of Two Independent PMFs:

$$p(t) = \mathbb{P}(T_2 = t) = \mathbb{P}_{X_1}(x) \cdot \mathbb{P}_{X_2}(x) := \sum_{x \in \text{Supp}[X]} \mathbb{P}_{X_1}(x)\mathbb{P}_{X_2}(t-x)$$

Let $X_1, X_2, X_2 \overset{iid}{\sim} \text{Bern}(p)$. Let

$$\begin{aligned}
T_3 &= X_1 + X_2 + X_3 = X_3 + T_2 \\
&= \mathbb{P}_{X_3}(x) \cdot \mathbb{P}_{T_2}(x) \\
&= \sum_{x \in \text{Supp}[X]} \mathbb{P}_{X_3}(x)\mathbb{P}_{T_2}(()t-x) \\
&= \sum_{x \in \{0,1\}} \binom{1}{x} p^x (1-p)^{1-x} \binom{2}{t-x} p^{t-x}(1-p)^{2-t+x} \\
&= p^t(1-p)^{3-t} \sum_{x \in \{0,1\}} \binom{1}{x}\binom{2}{t-x} \\
&= \binom{3}{t} p^t (1-p)^{3-t}
\end{aligned}$$

Let $X_1, X_2 \overset{iid}{\sim} \text{Bern}(\frac{1}{2})$ and $T_2 = X_1 + X_2$. For Bern(p):

$$\begin{aligned}
\mathbb{P}_{T_2}(x) &= \mathbb{P}_{X_1}(x) \cdot \mathbb{P}_{X_2}(x) \\
&= \sum_{x \in \text{Supp}[X]} \mathbb{P}_{X_1}(x)\mathbb{P}_{X_2}(t-x) \\
&= \sum_{x \in \{0,1\}} p^x(1-p)^{1-x} p^{t-x}(1-p)^{1-t+x}
\end{aligned}$$

$$= \sum_{x \in \{0,1\}} p^t (1-p)^{2-t}$$

$$\begin{aligned}
&= p^t(1-p)^{2-t} \underbrace{\sum 1}_{2} \\
&= 2p^t(1-p)^{2-t}
\end{aligned}$$

This was wrong.

$$p(2) = p^0(1-p)^{1-0} \qquad \underbrace{p^{2-0}(1-p)^{t-2}}_{\text{turned off using indicator function}} \qquad +p^1(1-p)^{t-1}p^{2-1}(1-p)^{1-2+1}$$

Let $X_1, X_2 \overset{iid}{\sim} \text{Binom}(n,p)$. Let $Y = X_1 + X_2$. Then

$$\begin{aligned}
\mathbb{P}_Y(x) &= \mathbb{P}_{X_1}(x) \cdot \mathbb{P}_{X_2}(x) \\
&= \sum_{x \in \text{Supp}[X]} \mathbb{P}_{X_1}(x)\mathbb{P}_{X_2}(y-x) \\
&= \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} \underbrace{\mathbb{1}_{x \in \{0,1,\dots,n\}}}_{\text{not needed}} \binom{n}{y-x} p^{y-x}(1-p)^{1-y+x} \underbrace{\mathbb{1}_{y-x \in \{0,1,\dots,n\}}}_{\text{not needed}} \\
&= \sum_{x \in \{0,1,\dots,n\}} \binom{n}{x} p^x (1-p)^{n-x} \binom{n}{y-x} p^{y-x}(1-p)^{n-y+x} \\
&= p^y(1-p)^{2n-y}\binom{2n}{y} \text{ by Vandermonde's Identity} \\
&= \text{Binom}(2n, p)
\end{aligned}$$

Consider $B_1, B_2, \dots \overset{iid}{\sim} \text{Bern}(p)$. Let $X = \overset{\min}{t}\ \{B_t = 1\} - 1$. This is called a geometric random variable. So $X \sim \text{Geom}(p)$. $\text{Supp}[X] = \{0, 1, \dots\} = \mathbb{N}$. Parameter Space: $0 < p < 1$. In fact

$$\begin{aligned}
\mathbb{P}(X = 0) &= p \\
\mathbb{P}(X = 1) &= (1-p)p \\
\mathbb{P}(X = 2) &= (1-p)^2 p \\
\mathbb{P}(X = x) &= (1-p)^x p
\end{aligned}$$

Now, for the convolution of $\text{Geom}(p)$. Let $T_2 = X_1 + X_2$.

$$\begin{aligned}
p(t) &= \mathbb{P}_{X_1}(x) \cdot \mathbb{P}_{X_2}(x) \\
&= \sum_{x \in \text{Supp}[X]} \mathbb{P}_{X_1}(x)\mathbb{P}_{X_2}(t-x) \\
&= \sum_{x \in \mathbb{N}_0} (1-p)^x p (1-p)^{t-x} p \mathbb{1}_{t-x \in \mathbb{N}_0} \\
&= (1-p)^t p^2 (t+1)
\end{aligned}$$

Now $\text{Supp}[T_2] = \{0, 1, \dots\}$. Let $T_3 = X_1 + X_2 + X_3 = X_3 + T_2$.

$$\begin{aligned}
p(t) &= \mathbb{P}_{X_3}(x) \cdot \mathbb{P}_{T_2}(x) \\
&= \sum_{x \in \text{Supp}[X_3]} \mathbb{P}_{X_3}(x) \mathbb{P}_{T_2}(t-x) \\
&= \sum_{x \in \mathbb{N}_0} (1-p)^x p (t-x+1)(1-p)^{t-x} p^2 \mathbb{1}_{t-x \in \text{Supp}[T_2] = \mathbb{N}_0} \\
&= p^3 (1-p)^t \sum_{x \in \mathbb{N}_0} (t-x+1) \mathbb{1}_{x \leq t} \\
&= (1-p)^t p^3 \Big((t+1) \sum_{x \in \mathbb{N}_0} \mathbb{1}_{x \leq t} - \sum_{x \in \mathbb{N}_0} x \mathbb{1}_{x \leq t} \Big) \\
&= (1-p)^t p^3 \Big((t+1) \underbrace{\sum_{x=0}^{t} 1}_{t+1} - \underbrace{\sum_{x=0}^{t} x}_{\frac{t(t+1)}{2}} \Big) \\
&= (1-p)^t p^3 \Big(\frac{t^2 + 3t + 2}{2} \Big)
\end{aligned}$$

In fact, $T_3 =$ number of failures until 3 successes.

$$\mathbb{P}(T_3 = t) = \binom{t+2}{2} (1-p)^t p^3$$

Note that

$$\binom{t+2}{2} = \frac{(t+2)!}{2!t!} = \frac{(2+t)(1+t)}{2} = \frac{t^2 + 3t + 2}{2}$$

These have a name. $T_2 \sim \text{NegBinom}(2, p)$. $T_3 \sim \text{NegBinom}(3, p)$.

Let $X \sim \text{Binom}(n, p)$ where $\text{Supp}[X] = \{0, \dots, n\}$. What if n is really big? What if p is really small? Let n and p be related such that $\lambda = np$ or $p = \frac{\lambda}{n}$. What is the pmf if $n \to \infty$?

$$\begin{aligned}
\lim_{n \to \infty} p(x) &= \lim_{n \to \infty} \binom{n}{x} p^x (1-p)^{n-x} \\
&= \lim_{n \to \infty} \binom{n}{x} \Big(\frac{\lambda}{n} \Big)^x \Big(1 - \frac{\lambda}{n} \Big)^{n-x} \\
&= \lim_{n \to \infty} \frac{n!}{x!(n-x)!} \frac{\lambda^x}{n^x} \Big(1 - \frac{\lambda}{n} \Big)^n \Big(1 - \frac{\lambda}{n} \Big)^{-x} \\
&= \frac{\lambda^x}{x!} \lim_{n \to \infty} \frac{n!}{(n-x)! n^x} \underbrace{\lim_{n \to \infty} \Big(1 - \frac{\lambda}{n} \Big)^n}_{e^{-\lambda}} \underbrace{\lim_{n \to \infty} \Big(1 - \frac{\lambda}{n} \Big)^{-x}}_{1} \\
&= \frac{\lambda^x e^{-\lambda}}{x!} = \text{Poisson}(\lambda)
\end{aligned}$$

Let $X \sim \text{Poisson}(\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$. $\text{Supp}[X] = \{0, 1, \dots\} = \mathbb{N}_0$. Parameter Space: $\lambda \in (0, \infty)$.

Convolution of Poisson: Let $X_1, X_2 \overset{iid}{\sim}$ Poisson(λ). Let $T = X_1 + X_2$.

$$\begin{aligned}
p(t) &= \sum_{x \in \text{Supp}[X]} \mathbb{P}_{X_1}(x)\mathbb{P}_{X_2}(t-x) \\
&= \sum_{x \in \mathbb{N}_0} \frac{\lambda^x e^{-\lambda}}{x!} \frac{\lambda^{t-x} e^{-\lambda}}{(t-x)!} \mathbb{1}_{x \leq t} \\
&= \lambda^t e^{-2\lambda} \sum_{x \in \mathbb{N}_0} \frac{1}{x!(t-x)!} \mathbb{1}_{x \leq t} \frac{t!}{t!} \\
&= \frac{\lambda^t e^{-2\lambda}}{t!} \sum_{x \in \mathbb{N}_0} \binom{t}{x} \mathbb{1}_{x \leq t} \\
&= \frac{\lambda^t e^{-2\lambda}}{t!} \sum_{x=0}^{t} \binom{t}{x} \\
&= \frac{\lambda^t e^{-2\lambda}}{t!} \cdot 2^t \\
&= \frac{(2\lambda)^t e^{-2\lambda}}{t!} \\
&= \text{Poisson}(2\lambda)
\end{aligned}$$

Let $X_1, X_2 \overset{iid}{\sim}$ Bern(p) and $T = X_1 + X_2$. Then

$$p(t) = \mathbb{P}_{X_1}(x) \cdot \mathbb{P}_{X_2}(x) = \sum_{x \in \text{Supp}[X]} \mathbb{P}_{X_1}(x) \cdot \mathbb{P}_{X_2}(t-x) \overset{?}{=} 2p^t(1-p)^{2-t}$$

$$\begin{aligned}
p(2) &\overset{?}{=} \mathbb{P}_{X_1}(0)\mathbb{P}_{X_2}(2-0) + \mathbb{P}_{X_1}(1)\mathbb{P}_{X_2}(2-1) \\
&= \mathbb{P}_{X_1}(0)\mathbb{P}_{X_2}(2) + \mathbb{P}_{X_1}(1)\mathbb{P}_{X_2}(2-1) \\
&= p^{-}(1-p)^2 p^2(1-p)^0 + p^1(1-p)^1 \cdot p^1(1-p)^1 \\
&= 2p^2(1-p)^2
\end{aligned}$$

Let $A = \Big\{w_1, w_2, \dots, w_n\Big\}$ where $|A| = n$. Let

$$\begin{aligned}
2^A &= \Big\{B : B \subseteq A\Big\} \\
&= \Big\{B : B \subseteq A \text{ and } |A| = 0\Big\} \bigcup \\
&\Big\{B : B \subseteq A \text{ and } |A| = 1\Big\} \bigcup \\
&\Big\{B : B \subseteq A \text{ and } |A| = 2\Big\} \bigcup \\
&\dots \\
&\bigcup \Big\{B : B \subseteq A \text{ and } |A| = n\Big\} \\
2^n &= |2^A| \\
&= \sum_{i=1}^{n} |\Big\{B : B \subseteq A \text{ and } |A| = i\Big\}| \\
&= \sum_{i=0}^{n} \binom{n}{i}
\end{aligned}$$

This proves that

$$2^n = \sum_{i=0}^{n} \binom{n}{i}$$

Recall $\text{E}[X] = \sum_{x \in \text{Supp}[X]} xp(x)$ for discrete random variables. Consider a function of a random variable g. Then $\text{E}[g(x)] = \sum_{x \in \text{Supp}[X]} g(x)p(x)$
Let $z = \mathbb{1}_A$. Then $z \sim \text{Bern}(P(A))$. Hence $\text{E}[z] = P(A)$.
If $z = g(x, y)$, a function of two random variables,

$$\text{E}[z] = \text{E}[g(x, y)] = \sum_{x \in \text{Supp}[X]} \sum y \in \text{Supp}[Y] g(x, y) \mathbb{P}_{X,Y}(x, y)$$

where $\mathbb{P}_{X,Y}(x, y)$ is a jmf.
Let $X, Y \overset{iid}{\sim} \text{Geom}(p) = (1 - p)^x p$. Then

$$\text{E}[X] = \mathbb{P}(X \leq x) = 1 - \mathbb{P}(X > x) = 1 - (1 - p)^{x+1}$$

What is $\mathbb{P}(X > Y)$? Let $z = \mathbb{1}_{x>y} = g(x,y)$. Then

$$\begin{aligned}
\mathbb{P}(X > Y) &= \mathrm{E}[z] \\
&= \sum_{y\in\mathbb{N}_0}\sum_{x\in\mathbb{N}_0} \mathbb{1}_{x>y}\mathbb{P}_{X,Y}(x,y) \\
&= p^2 \sum_{y\in\mathbb{N}_0}(1-p)^y \sum_{x\in\mathbb{N}_0}(1-p)^x \mathbb{1}_{x>y} \\
&\text{since } X, Y \overset{iid}{\sim},\ \mathbb{P}_{X,Y}(x,y) = \mathbb{P}_X(x)\mathbb{P}_Y(y) = p(1-p)^x p(1-p)^y \\
&= p^2 \sum_{y\in\mathbb{N}_0}(1-p)^y \sum_{x=y+1}^{\infty}(1-p)^x \\
&\text{Let } x' = x-(y+1) = x-y-1 \to x = x'+y+1 \\
&= p^2 \sum_{y\in\mathbb{N}_0}(1-p)^y \sum x' \in \mathbb{N}_0 (1-p)^{x'+y+1} \\
&= p^2 \sum_{x\in\mathbb{N}_0}(1-p)^{2y+1} \sum_{x'\in\mathbb{N}_0}(1-p)^{x'} \\
&= p^2(1-p) \underbrace{\sum_{y\in\mathbb{N}_0}\left((1-p)^2\right)^y}_{\underbrace{\frac{1}{1-(1-p)^2}}_{\frac{1}{p(2-p)}}} \underbrace{\sum_{x'\in\mathbb{N}_0}(1-p)^{x'}}_{\underbrace{\frac{1}{1-(1-p)}}_{\frac{1}{p}}} \\
&= \frac{1-p}{2-p}
\end{aligned}$$

In fact,

$$\lim_{p\to 0}\mathbb{P}(X > Y) = \frac{1}{2}$$

What is $\mathbb{P}(X = Y)$? Let $z = \mathbb{1}_{x=y}$. Then

$$\begin{aligned}
\mathbb{P}(X = Y) &= \mathrm{E}[z] \\
&= \sum\sum \mathbb{1}_{x=y}\mathbb{P}_{X,Y}(x,y) \\
&= \sum_{y\in\mathbb{N}_0} p(1-p)^y \underbrace{\sum_{x=y}^{y} p(1-p)^x}_{\text{one element}} \\
&= p^2 \sum_{y\in\mathbb{N}_0}(1-p)^{2y} \\
&= p^2 \frac{1}{p(2-p)} \\
&= \frac{p}{2-p}
\end{aligned}$$

Let $X, Y \overset{iid}{\sim} \text{Binom}(n,p)$.Then

$$\mathbb{P}(X > Y) = \sum_{y \in \mathbb{N}_0} \mathbb{P}(Y = y)(1 - F_X(y))$$

But $F_X(y)$ has no closed form.

A basket has apples and bananas. Let p_1 = probability of getting apples and p_2 = probability of getting bananas. It is true that $p_2 = 1 - p_1$. Furthermore, $p_1 \in (0,1)$. Represent apples as x_1. Then bananas can be represented as $x_2 = n - x_1$ where n is the total number of fruits in the basket. A vector can be created that represents this:

$$\vec{X} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Let's add cantaloupes to the basket. p_3 = probability of getting cantaloupes. Now, the parameter space is such that $p_1 + p_2 + p_3 = 1$ and $\vec{X} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$.

What's $\mathbb{P}(\vec{X} = \vec{x})$?

$$\mathbb{P}_{\vec{X}}(x_1, x_2, x_3) = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \mathbb{1}_{x_1+x_2+x_3=n}$$

where the factorials term can be simplified to $\binom{n}{x_1,x_2,x_3}$.

In general,

$$\vec{X} \sim \text{Multinom}(n, \vec{p}) := \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} p_2^{x_2} \cdot \dots \cdot p_k^{x_k} \mathbb{1}_{\sum x_i = n}$$

such that $\binom{n}{x_1,x_2,\dots,x_k} = \frac{n!}{x_1!x_2!\dots x_k!}$. Note that $\vec{X}$ is a multidimensional random variable of dim K and $\vec{p}$ is a multidimensional parameter of dim $K$where $n, x_i \in \mathbb{N}$ and $\sum x_1 \leq n$. This is the multidimensional generalization of the binomial distribution. Instead of two categories (successes and failures), there are k categories.

Let's go back to the basket problem. If $k = 3$, $n = 10$ and $p_1 = \frac{1}{4}$, $p_2 = \frac{1}{8}$, $p_3 = \frac{5}{8}$, how many mays are there to have 3 apples, 3 bananas and 4 cantaloupes?

$$\mathbb{P}(\vec{X} = \begin{bmatrix} 3 \\ 3 \\ 4 \end{bmatrix}) = \binom{10}{3,3,4} \left(\frac{1}{4}\right)^3 \left(\frac{1}{8}\right)^3 \left(\frac{5}{8}\right)^4$$

What are the parameter space of the Multinomial distribution? $n \in \mathbb{N}$. $p \in (0,1)^k$ or sets of all k-tuples such that $\vec{p} \cdot \vec{1} = 1$ where $\sum p_k = 1$.

Let $\vec{X} \sim \text{Multinom}(n, \vec{p}) := \binom{n}{x_1, x_2, \dots, x_k} p_1^{x_1} p_2^{x_2} \dots p_k^{x_k}$ where k is the number of categories to choose from.

$$\dim[X] = k$$

There is no indicator function since multichoose is 0 unless

$$\sum_{i=1}^{k} x_i = n \text{ and } \forall x_i \in \mathbb{N}_0$$

$$\text{Supp}[\vec{X}] = \{\vec{x} : \vec{1} \cdot \vec{x} = n \text{ and } \vec{x} \in \mathbb{N}_0^k\}$$

$$\text{Parameter Space } : \vec{p} \in \{\vec{p} : \vec{p} \in (0,1)^k \text{ and } \vec{p} \cdot \vec{1} = 1\}$$

What's the probability of getting 3 apples, 2 bananas and 5 cantaloupes if $p_A = \frac{1}{4}$, $p_B = \frac{1}{8}$ and $p_C = \frac{5}{8}$?

$$\mathbb{P}(\vec{X} = \begin{bmatrix} 3 \\ 2 \\ 5 \end{bmatrix}) = \binom{10}{3,2,5} \left(\frac{1}{4}\right)^3 \left(\frac{1}{8}\right)^2 \left(\frac{5}{8}\right)^5$$

Let $k = 2$, then $\vec{p} = \begin{bmatrix} p \\ 1-p \end{bmatrix}$. Thus

$$p(\vec{x}) = \mathbb{P}(x_1, x_2) = \text{Multinom}(n, \begin{bmatrix} p \\ 1-p \end{bmatrix}) = \binom{n}{x_1, x_2} p^{x_1} (1-p)^{x_2}$$

This is not binomial ($\text{Bin}(n, p)$).
Is $X_1, X_2 \stackrel{iid}{\sim}$? If so,

$$\mathbb{P}(x_1, x_2) = \mathbb{P}(x_1)\mathbb{P}(x_2) \to \mathbb{P}(x_1 \mid x_2) = \mathbb{P}(x_1) \text{ or } \mathbb{P}(x_2 \mid x_1) = \mathbb{P}(x_2)$$

This is true since $\forall x_1 \in \text{Supp}[X_1]$ and $\forall x_2 \in \text{Supp}[X_2]$,

$$\begin{aligned} \mathbb{P}(X_1 \mid X_2) &= \frac{\mathbb{P}(X_1, X_2)}{\mathbb{P}(X_2)} \stackrel{\text{if iid}}{=} \frac{\mathbb{P}(X_1)\mathbb{P}(X_2)}{\mathbb{P}(X_2)} &= \mathbb{P}(X_1) \\ \mathbb{P}(X_2 \mid X_1) &= &= \mathbb{P}(X_2) \end{aligned}$$

Thus are $X_1, X_2 \stackrel{iid}{\sim}$? No. If you know x_2, then $x_1 = n - x_2$. They are dependent on one another.

$$\begin{aligned}
\mathbb{P}(X_1 \mid X_2) &= \frac{\mathbb{P}(X_1, X_2)}{\mathbb{P}(X_2)} \\
\mathbb{P}(X_2) &= \sum_{x_1 \in \text{Supp}[X_1]} \mathbb{P}(X_1, X_2) \\
&= \sum_{x_1=0}^{n} \frac{n!}{x_1! x_2!} p^{x_1} (1-p)^{x_2} \mathbb{1}_{x_1+x_2=n} \\
&= \frac{n!}{x_2!}(1-p)^{x_2} \qquad \underbrace{\sum_{x_1=0}^{n} \frac{p^{x_1}}{x_1!} \mathbb{1}_{x_1=n-x_2}}_{\text{this is all zero except when } x_1=n-x_2} \\
&= \frac{n!}{x_2!}(1-p)^{x_2} \frac{p^{n-x_2}}{(n-x_2)!} \\
&= \binom{n}{x_2}(1-p)^{x_2} p^{n-x_2} \\
X_2 &\sim \text{Binom}(n, 1-p) \\
X_1 &\sim \text{Binomn, p}
\end{aligned}$$

This shows that the marginal distribution is a binomial distribution as well.

$$\mathbb{P}(X_1 \mid X_2) = \frac{\frac{n!}{x_1! x_2!} p^{x_1} (1-p)^{x_2} \mathbb{1}_{x_1+x_2=n}}{\frac{n!}{x_2!(n-x_2)!}(1-p)^{x_2} p^{n-x_2}} = \frac{(n-x_2)!}{x_1!} p^{x_1+x_2-n} \mathbb{1}_{x_1+x_2=n}$$

The indicator function is 0 unless $x_1 = n - x_2$. Thus

$$\mathbb{P}(x_1 = n - x_2 \mid x_2) = \frac{(n-x_2)!}{(n-x_2)!} p^0 = 1$$

This is not the same as $\text{Bin}(n, p)$.

Let $X \sim \text{Multinom}(n, \vec{p})$. Then

$$\begin{aligned}
\mathbb{P}(X_{-j} \mid X_j) &= \frac{\mathbb{P}(X_1, \dots, X_k)}{\mathbb{P}(X_j)} = \frac{\text{Multinom}(n, \vec{p})}{\text{Binom}(n, p_j)} \\
&= \frac{\frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}}{\frac{n!}{x_j!(n-x_j)!} p_j^{x_j} (1-p_j)^{n-x_j}} \\
&= \frac{(n-x_j)!}{x_1! \dots x_{j-1}! x_{j+1}! \dots x_k!} \frac{p_1^{x_1} \dots p_{j-1}^{x_{j-1}} p_{j+1}^{x_{j+1}} \dots p_k^{x_k}}{(1-p_j)^{n-x_j}} \\
&\text{Let } n' = n - x_j \\
&\text{Then } \sum_{j=1}^k x_j = n \to x_1 + \dots + x_{j-1} + x_j + x_{j+1} + \dots + x_k = n \\
&\to x_1 + \dots + x_{j-1} + x_{j+1} + \dots + x_k = n - x_j = n' \\
&= \binom{n'}{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k} \frac{p_1^{x_1} \dots p_{j-1}^{x_{j-1}} p_{j+1}^{x_{j+1}} \dots p_k^{x_k}}{(1-p_j)^{n'}} \\
&\text{Let } p_1' = \frac{p_1}{1-p_j}, p_2' = \frac{p_2}{1-p_j}, \dots, p_k' = \frac{p_k}{1-p_j} \to p' = \begin{bmatrix} p_1' \\ \dots \\ p_k' \end{bmatrix} \\
&= \binom{n'}{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k} \\
&\cdot \frac{(p_1'(1-p_1))^{x_1} \dots (p_{j-1}'(1-p_{j-1}))^{x_{j-1}} (p_{j+1}'(1-p_{j+1}))^{x_{j+1}} \dots (p_k'(1-p_k))^{x_k}}{(1-p_j)^{n'}} \\
&= \binom{n'}{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k} \\
&\cdot \frac{(p_1')^{x_1} \dots (p_{j-1}')^{x_{j-1}} (p_{j+1}')^{x_{j+1}} \dots (p_k)^{x_k} (1-p_j)^{\overbrace{x_1 + \dots + x_{j-1} + x_{j+1} + \dots + x_k}^{n'}}}{(1-p_j)^{n'}} \\
&= \text{Multinom}(n', p')
\end{aligned}$$

Recall that $\text{E}[g(X_1,\dots,X_n)] = \sum_{x_1\in\text{Supp}[X_1]}\cdots\sum_{x_n\in\text{Supp}[X_n]} g(x_1,\dots,x_n)\mathbb{P}(x_1,\dots,x_n)$.

$$\begin{aligned}
\text{E}[aX] &= a\text{E}[X] \\
\text{E}[X+c] &= \text{E}[X]+c \\
\text{E}[\sum X_i] &= \sum_{i=1}^n \text{E}[X_i] = n\mu \\
\text{E}[\prod_{i=1}^n X_i] &= \prod_{i=1}^n \text{E}[X_i] = \mu^n \text{ if } X_1,\dots,X_n \overset{iid}{\sim} \\
\sigma^2 = \text{Var}[X] &= \text{E}[\underbrace{(X-\mu)^2}_{g(x)}] = \sum_{x\in x\in\text{Supp}[X]} g(x)\mathbb{P}(x) \\
&= \sum (x-\mu)^2 p(x) \\
&= \sum x^2 p(x) + \sum(-2X\mu)p(x) + \sum \mu^2 p(x) \\
&= \text{E}[X^2] - 2\mu^2 + \mu^2 = \text{E}[X^2]-\mu^2
\end{aligned}$$

$$\begin{aligned}
\text{Var}[X+c] &= \text{Var}[X] \\
\text{Var}[cX] &= c^2\text{Var}[X] \\
\text{Var}[X_1+X_2] &= \text{E}[((X_1+X_2)-(\mu_1+\mu_2))^2] \\
&= \text{E}[X_1^2+X_2^2+\mu_1^2+\mu_2^2-2\mu_1X_1-2\mu_1X_2-2\mu_2X_1-2\mu_2X_2+2X_1X_2+2\mu_1\mu_2] \\
&= \text{E}[X_1^2]+\text{E}[X_2^2]+\mu_1^2+\mu_2^2-2\mu_1^2-2\mu_1\mu_2-2\mu_1\mu_2-2\mu_2^2+2\text{E}[X_1X_2]+2\mu_1\mu_2 \\
&= \text{Var}[X_1]+\text{Var}[X_2]+2(\text{E}[X_1X_2]-\mu_1\mu_2)
\end{aligned}$$

Define covariance as follows:

$$\text{Cov}[X_1,X_2] = \text{E}[X_1X_2]-\mu_1\mu_2$$

In fact,

$$\text{Corr}[X_1,X_2] = \frac{\text{Cov}[X_1,X_2]}{\text{SE}[X_1]\text{SE}[X_2]} \in [-1,1]$$

$$\begin{aligned}
\text{Cov}[X,X] &= \text{Var}[X] \\
\text{Cov}[aX_1,bX_2] &= ab\text{Cov}[X_1,X_2] \\
\text{Cov}[X_1+c,X_2+d] &= \text{Cov}[X_1,X_2] \\
\text{Cov}[X_2,X_1] &= \text{Cov}[X_1,X_2] \\
\text{Cov}[X+Y,Z] &= \text{E}[(X+Y-\mu_X-\mu_Y)(Z-\mu_Z)] \\
&= \text{E}[((X-\mu_X)+(Y-\mu_Y))(Z-\mu_Z)] \\
&= \text{E}[(X-\mu_X)(Z-\mu_Z)+(Y-\mu_Y)(Z-\mu_Z)] \\
&= \text{Cov}[X,Z]+\text{Cov}[Y,Z]
\end{aligned}$$

Note that

$$\begin{aligned}
\text{Var}[X_1 + X_2] &= \text{Var}[X_1] + \text{Var}[X_2] + 2\text{Cov}[X_1, X_2] \\
&= \text{Cov}[X_1, X_1] + \text{Cov}[X_2, X_2] + \text{Cov}[X_1, X_2] + \text{Cov}[X_2, X_1] \\
&= \sum_{i=1}^{2} \sum_{j=1}^{2} \text{Cov}[X_i, X_j] \\
\text{Var}[X_1 + X_2 + \cdots + X_k] &= \sum_{i=1}^{k} \sum_{j=1}^{k} \text{Cov}[X_i, X_k]
\end{aligned}$$

If $\vec{X}$ is a vector of random variables of dim k,

$$\text{E}[\vec{X}] = \text{E}[\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_k \end{bmatrix}] = \begin{bmatrix} \text{E}[X_1] \\ \text{E}[X_2] \\ \dots \\ \text{E}[X_k] \end{bmatrix}$$

Furthermore,

$$\text{Var}[\vec{X}] = \begin{bmatrix} \overbrace{\text{Var}[X_1]}^{\text{Cov}[X_1, X_1]} & \text{Cov}[X_1, X_2] & \dots & \dots \\ \text{Cov}[X_2, X_1] & \text{Var}[X_2] & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \text{Var}[X_k] \end{bmatrix}$$

This is a symmetric $k \times k$ matrix defined by

$$\text{Cov}[X_i, X_j] \quad \forall i = 1, \dots, k \text{ and } j = 1, \dots, k$$

Let $\vec{X}$ be a vector of random variables such that $\dim[X] = k$.

$$\begin{aligned}
\vec{\mu} = \text{E}[\vec{X}] &= \begin{bmatrix} \text{E}[x_1] \\ \dots \\ \text{E}[x_n] \end{bmatrix} \\
\varepsilon = \text{Var}[\vec{X}] &= \begin{bmatrix} \text{Var}[x_1] & \text{Cov}[x_1, x_2] & \dots & \dots \\ \text{Cov}[x_2, x_1] & \text{Var}[x_2] & \dots & \dots \\ \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & \text{Var}[x_k] \end{bmatrix} \\
&= \Big\{ \text{Cov}[x_i, x_i] \text{ for } i = 1, \dots, k, \ j = 1, \dots, k \Big\} \\
\varepsilon_0 = \text{Corr}[\vec{X}] &= \begin{bmatrix} 1 & & \text{Corr}[x_i, x_j] \\ & 1 & \\ \text{Corr}[x_i, x_j] & & 1 \end{bmatrix} \\
&= \Big\{ \text{Corr}[x_i, x_j] \text{ for } i = 1, \dots, k, \ j = 1, \dots, k \Big\}
\end{aligned}$$

Let $T = X_1 + \dots + X_k = T^T\vec{X} = \begin{bmatrix} 1 & \dots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ \dots \\ x_k \end{bmatrix}$.

$$\text{E}[T] = \sum_{i=1}^k \mu_i = T^T\vec{\mu}$$
$$\text{Var}[T] = \text{Var}[T^T\vec{X}] = \sum_{j=1}^k \sum_{i=1}^k \text{Cov}[X_i, X_j]$$

Let $Y = \vec{c}^T\vec{X}$. Then $\text{E}[Y] = \sum c_i\mu_i = \vec{c}^T\vec{\mu}$. What's $\text{Var}[Y] = \text{Var}[\vec{c}^T\vec{X}]$?
If $A \in \mathbb{R}^{n\times n}$ and $\vec{c} \in \mathbb{R}^n$, what is $\vec{c}^TA\vec{c}$?

$$\begin{aligned}\vec{c}^Ta\vec{c} = \vec{c}^T &\begin{bmatrix} c_1a_{11} + \dots + c_na_{1n} \\ c_1a_{21} + \dots + c_na_{2n} \\ \dots \\ c_1a_{n1} + \dots + c_na_{nn} \end{bmatrix} \\ &= c_1^2a_{11} + c_1c_2a_{12} + \dots + c_1c_na_{1n} + \\ &c_2c_1a_{21} + c_2^2a_{22} + \dots + c_2c_na_{2n} + \\ &\dots \\ &c_nc_1a_{n1} + c_nc_2a_{n2} + \dots + c_n^2a_{nn} \\ &= \sum_{j=1}^n \sum_{i=1}^n c_ic_ja_{ij}\end{aligned}$$

Thus what is $\text{Var}[\vec{c}^T\vec{X}]$?

$$\begin{aligned}\text{Var}[\vec{c}^T\vec{X}] - \text{Var}[c_1X_1 + \dots + c_kX_k] \\ &= \sum_{i=1}^k \sum_{j=1}^k \text{Cov}[c_iX_i, c_jX_j] \\ &= \sum_{i=1}^k \sum_{j=1}^k c_ic_j\text{Cov}[X_i, X_j] \\ &= \vec{c}^T\text{Var}[\vec{X}]\vec{c}\end{aligned}$$

Markovits Optimal Portfolio: Let $X_1, \dots, X_k$ be random variable models for the returns on k assets. Let $w_1, \dots, w_k$ be the weights or allocations for each. Note that $T^T\vec{w} = 1$. In addition,

$$V = \vec{w}^T\vec{X}$$
$$\text{E}[V] = \vec{w}^T\vec{\mu}$$
$$\text{Var}[V] = \vec{w}^T\sum\vec{w}$$

Given μ_0, minimize $\vec{w}^T\sum\vec{w}$ such that $T^t\vec{w} = 1(\Big\{\vec{w} : T^T\vec{w} = 1\Big\})$.

If $\vec{X} \sim \text{Multinomial}(n, \vec{p})$,

$$\text{E}[\vec{X}] = \begin{bmatrix} \text{E}[X_1] \\ \dots \\ \text{E}[X_n] \end{bmatrix} = \begin{bmatrix} np_1 \\ \dots \\ np_k \end{bmatrix} = n\vec{p}$$

$$\text{Var}[X] = \begin{bmatrix} np_1(1-p_1) & \text{Cov}[X_1, X_2] & \dots & \dots \\ & np_2(1-p_2) & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ \dots & \dots & \dots & np_k(1-p_k) \end{bmatrix}$$

Also

$$\begin{aligned} \text{Cov}[X_i, X_j] &= \text{E}[X_i, X_j] - \mu_i\mu_j \\ &= \sum_{x_i \in \text{Supp}[X_1]} \sum_{x_j \in \text{Supp}[X_2]} x_i x_j \underbrace{\mathbb{P}_{X_iX_j}(X_iX_j)}_{\text{we don't know this yet}} - \mu_i\mu_j \end{aligned}$$

Recall that if $X_1 \sim \text{Binom}(n, p_1), \dots, X_k \sim \text{Binom}(n, p_k)$, that means that $X_1 = \sum_{i=1}^n X_{i1}$ such that $X_{11}, \dots, X_{n1} \overset{iid}{\sim} \text{Bern}(p_1)$, all the way through $X_k = \sum_{i=1}^n X_{ik}$ such that $X_{1k}, \dots, X_{nk} \overset{iid}{\sim} \text{Bern}(p_k)$.
If $\vec{X} \sim \text{Multinomial}(n, \vec{p})$ then $\vec{X} = \sum_{i=1}^n \vec{X}_i$ such that $\vec{X}_1, \vec{X}_2, \dots, \vec{X}_n \overset{iid}{\sim} \text{Multinomial}(1, \vec{p})$.
Then the covariance of X_i, X_j is as follows

$$\begin{aligned} \text{Cov}[X_i, X_j] &= \text{Cov}[\sum_{l=1}^n X_{li}, \sum_{h=1}^n X_{hj}] \\ &= \sum \sum \text{Cov}[X_{li}, X_{hj}] \\ &= \sum_{l=1}^n \sum_{h=1}^n \text{E}[X_{li}, X_{hj}] - p_ip_j \\ &\text{If } l = h \\ &= \sum_{l=1}^n \text{E}[X_{li}, X_{lj}] - p_ip_j \end{aligned}$$

$$\sum_{l=1}^n -p_ip_j = -np_ip_j$$

$$\begin{aligned} &\text{If } l \neq h \\ &= \text{E}[X_{li}]\text{E}[X_{hj}] \\ &= p_ip_j \end{aligned}$$

Continuous random variable X have CDF $F(x)$ and PDF $f(x)$ such that

$$f(x) = F'(x)$$

and $\text{Supp}[X] = \Big\{x : f(x) > 0\Big\}$ and $|\text{Supp}| = |\mathbb{R}|$. Note that pmf $P(x) = 0 \forall x$.
Let $X \sim U(a, b) = \frac{1}{b-a}$ where $a, b \in \mathbb{R}$, $b > a$ and $\text{Supp}[x] = [a, b]$.

A standard uniform distribution occurs when $a = 0, b = 1$ forming $X \sum U(0,1) = 1$. Let $T_2 = X_1 + X_2$ such that $X_1, X_2 \overset{iid}{\sim} U(0,1)$. Then $\text{Supp}[T_2] = [0,2]$.
How often does T = 0? That's when $x_1 = 0$, $x_2 = 0$. None. How often does T = 2? That's when $x_1 = 1$, $x_2 = 1$. None. How often does T = 1? That's when $x_1 = 0$ and $x_2 = 1$ or $x_1 = \frac{1}{3}$ and $x_2 = \frac{2}{3}$, and so on.

$$\begin{aligned} f_T(t) &= \int_{x \in \text{Supp}[X_1]} f_{X_1}(x) f_{X_2}(t-x)\,dx \\ &= \int_0^1 1 \cdot \underbrace{\mathbb{1}_{x\in[0,1]}}_{\text{not needed}} \cdot 1 \cdot \mathbb{1}_{t-x} \in [0,1]\,dx \\ &= \int_0^1 \mathbb{1}_{x \in [t-1,t]}\,dx \\ &= \int_{\max\{0,t-1\}}^{\min\{1,t\}} dx \\ &= x \Big|_{\max\{0,t-1\}}^{\min\{1,t\}} \\ &= (\min\{1,t\} - \max\{0,t-1\})\mathbb{1}_{t\in[0,2]} \end{aligned}$$

This is the answer for $t \in [0,2]$. Alternatively,

$$f_{T_2}(t) = \begin{cases} t & \text{if } t < 1 \\ 1 - (t-1) = 2 - t & \text{if } t \geq 1 \end{cases} \mathbb{1}_{t\in[0,2]}$$

Let X, Y be continuous random variables with jdf $f_{X,Y}(x,y)$. Let $Z = g(X,Y)$ .Then

$$F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(g(X,Y) \leq z) = \int_{-\infty}^z f_Z(t)\,dt = \iint_{\left\{(x,y):g(x,y)\leq z\right\}} f_{X,Y}(x,y)\,dxdy$$

where $f_Z(t)$ is the pdf of Z.
Let $T = X + Y$. Then

$$F_Z(z) = \iint_{\left\{(x,y):x+y\leq z\right\}} f_{X,Y}(x,y)\,dxdy$$

$$\int_{\mathbb{R}} \left(\int_{\left\{y:y\leq z-x\right\}} f_{X,Y}(x,y)\,dy \right) dx$$

$$\begin{aligned} &= \int_{\mathbb{R}} \left(\int_{-\infty}^{z-x} f_{X,Y}(x,y)\,dy \right) dx \\ &= \int_{\mathbb{R}} \int_{-\infty}^{z} f_{X,Y}(x,t-x)\,dt\,dx \\ &= \int_{-\infty}^{z} \underbrace{\left(\int_{\mathbb{R}} f_{X,Y}(x,t-x)\,dx \right)}_{f_T(t)} dt \end{aligned}$$

The convolution of $f_X(x) \times f_Y(y)$ is sometime notated as $(f_X \times f_Y)(x)$.

If $X, Y \overset{iid}{\sim}$, the definition of convolution for independent random variables is as follows

$$f_T(t) = \int_{\mathbb{R}} f_X(x) f_Y(t-x)\,dx = \int_{\text{Supp}[X]} f_X(x) f_Y(t-x) \mathbb{1}_{t-x \in \text{Supp}[Y]}\,dx$$

Note that the indicator functions are included in both $f_X(x)$ and $f_Y(t-x)$.

Let $X, Y \overset{iid}{\sim} U(0,1)$ and $T = X + Y$. What's $f_T(t)$?

$$\begin{aligned} f_T(t) &= \int_{\text{Supp}[X]} f_X(x) f_Y(t-x) \mathbb{1}_{t-x \in \text{Supp}[Y]}\,dx \\ &= 1 \cdot \mathbb{1}_{x \in [0,1] \text{ and } y \in [0,1]} \\ F_T(t) &= \iint_{\left\{(x,y): x+y \leq t\right\}} f_{X,Y}(x,y)\,dxdy \\ &= \begin{cases} \frac{1}{2}t^2 & \text{if } t \in [0,1] \\ \frac{1}{2} + (\frac{1}{2} - \frac{1}{2}(2-t)^2) & \text{if } t \in [1,2] \end{cases} \end{aligned}$$

If we integrate this function to get $f_T(t)$,

$$f_T(t) = F_T'(t) = \begin{cases} t & \text{if } t \in [0,1] \\ 2-t & \text{if } t \in [1,2] \end{cases}$$

Let $X_1, X_2 \overset{iid}{\sim} U(a,b)$ and $T_2 = X_1 + X_2$. Supp$[T] = [2a, 2b]$.

$$\begin{aligned} f_{T_2}(t) &= \int_{\text{Supp}[X_1]} f_{X_1}(x) f_{X_2}(t-x)\,dx \\ &= \int_a^b \frac{1}{b-a} \frac{1}{b-a} \mathbb{1}_{t-x \in [a,b] \to x \in [t-b, t-a]}\,dx \\ &= \frac{1}{(b-a)^2} \int_{\max\{a, t-b\}}^{\min\{b, t-a\}} 1\,dx \\ &= \frac{1}{(b-a)^2} \Big(\min\{b, t-a\} - \max\{a, t-b\} \Big) \\ f_{T_2}(t) &= \begin{cases} \frac{t-2a}{(b-a)^2} & \text{if } t < a+b \\ \frac{2b-t}{(b-a)^2} & \text{if } t \geq a+b \end{cases} \mathbb{1}_{t \in [2a, 2b]} \end{aligned}$$

Recall that if $X \sim \text{Geom}(p) = (1-p)^x p$, then $F(x) = \mathbb{P}(X \leq x) = 1 - \mathbb{P}(X > x) = 1 - (1-p)^x$. If n many geometric realizations occur within each time period, then $x = tn$ and so $p(t) = (1-p)^{tn} p$. If $n \to \infty$ and $p \to 0$ but $\lambda = np$,

$$\begin{aligned} p(t) &= \Big(1 - \frac{\lambda}{n}\Big)^{tn} \frac{\lambda}{n} \\ \lim_{n \to \infty} p(t) &= \underbrace{\Big(\lim_{n \to 0} (1 - \frac{\lambda}{n})^n \Big)^t}_{e^{-\lambda t}} \underbrace{\lim_{n \to \infty} \frac{\lambda}{n}}_{0} = 0 \end{aligned}$$

Once the support is no longer discrete, the PMF vanishes. But recall that

$$\begin{aligned}
F(x) &= 1 - (1-p)^x \\
F_n(t) &= 1 - (1-p)^{nt} \\
F_n(t) &= 1 - (1 - \frac{\lambda}{n})^{nt} \\
\lim_{n\to\infty} F_n(t) &= 1 - \Big(\lim_{n\to\infty} (1 - \frac{\lambda}{n})^n\Big)^t = 1 - e^{-\lambda t} \\
\mathbb{P}(X > x) &= 1 - F(t) = e^{-\lambda t} \\
f_T(t) &= \frac{d}{dt} F_T(t) = \lambda e^{-\lambda t}
\end{aligned}$$

Let $X \sim \text{Exp}(\lambda) = \lambda e^{-\lambda x}$ where $\text{Supp}[X] = (0,\infty)$. Parameter space: $\lambda = np$ and $\lambda \in (0,\infty)$. This distribution can be used as a basic model for waiting time or failure time or survival.

If $a, b \in \mathbb{R}^+$,

$$\begin{aligned}
\mathbb{P}(x > a+b \mid x > b) &= \frac{\mathbb{P}(x > a+b \text{ and } x > b)}{\mathbb{P}(x > b)} \\
&= \frac{\mathbb{P}(x > a+b)}{\mathbb{P}(x > b)} \\
&= \frac{e^{-(a+b)x}}{e^{-bx}} \\
&= e^{-ax} \\
&= 1 - F(a) \\
&= \mathbb{P}(x > a)
\end{aligned}$$

For a continuous random variable X,

$$\text{E}[X] = \int_{\text{Supp}[X]} x f(x)\, dx$$

For the exponential distribution,

$$\int_0^\infty x\lambda e^{-\lambda x}\, dx = \lambda \int_0^\infty x e^{-\lambda x}\, dx = \dots = \frac{1}{\lambda}$$

Let $X_1, X_2, \dots \overset{iid}{\sim} \text{Exp}(\lambda)$. What's $T_2 = X_1 + X_2 \sim$? $\text{Supp}[T_2] = (0,\infty)$.

$$\begin{aligned}
f_{T_2}(t) &= \int_{\text{Supp}[X_1]} f_{X_1}(x) f_{X_2}(t-x)\, dx \\
&= \int_0^\infty \lambda e^{-\lambda x} \mathbb{1}_{x\in(0,\infty)} \cdot \lambda e^{-\lambda(t-x)} \mathbb{1}_{t-x\in(0,\infty)\to x\in(-\infty,t)}\, dx \\
&= \lambda^2 \int_0^\infty e^{-\lambda t} \mathbb{1}_{x\in(-\infty,t)}\, dx \\
&= \lambda^2 e^{-\lambda t} \int_0^\infty dt \\
&= \lambda^2 t e^{-\lambda t}
\end{aligned}$$

Let $T_3 = X_1 + X_2 + X_3 = X_3 + T_2$.

$$\begin{aligned}
f_{T_3}(t) &= \int_{\text{Supp}[X_1]} f_{X_1}(x) f_{T_2}(t-x)\,dx \\
&= \int_0^\infty \lambda e^{-\lambda x} \cdot \lambda^2 (t-x) e^{-\lambda(t-x)} \mathbb{1}_{t-x\in(0,\infty)}\,dx \\
&= \lambda^3 e^{-\lambda t} \int_0^\infty (t-x)\mathbb{1}_{t-xin(0,\infty)}\,dx \\
&= \lambda^3 e^{-\lambda t}\Big(t\int_0^\infty \mathbb{1}_{t-x\in(0,\infty)}\,dx - \int_0^\infty x\mathbb{1}_{t-x\in(0,\infty)}\,dx\Big) \\
&= \lambda^3 e^{-\lambda t}\Big(t\int_0^t dx - \int_0^t x\,dx\Big) \\
&= \lambda^3 e^{-\lambda t}(t^2 - \frac{t^2}{2}) \\
&= \frac{\lambda^3 t^2}{2} e^{-\lambda t}
\end{aligned}$$

One more time

$$\begin{aligned}
f_{T_4}(t) &= f_{X_4}(x) f_{T_3}(t) \\
&= \int_0^\infty \lambda e^{-\lambda x} \frac{\lambda^3 (t-x)^2}{2} e^{-\lambda(t-x)} \mathbb{1}_{t-x\in(0,\infty)}\,dx \\
&= \lambda^4 e^{-\lambda t}\frac{1}{2}\int_0^t (t-x)^2\,dx \\
&= \lambda^4 e^{-\lambda t}\frac{1}{3\cdot 2} t^3
\end{aligned}$$

Following this pattern, we get

$$f_{T_k}(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} = \text{Erlang}(k,\lambda)$$

Its parameter space is as follows: $\lambda \in (0,\infty)$, $k \in \mathbb{N}$. $\text{Supp}[X] = (0,\infty)$.

What's F_{T_k} of the Erlang distribution?

$$\begin{aligned}
F_{T_k} &= \int_0^x \frac{\lambda^k y^{k-1} e^{-\lambda y}}{(k-1)!}\,dy \\
&= \frac{1}{(k-1)!}\int_0^x \lambda(\lambda y)^{k-1} e^{-\lambda y}\,dy \\
&\text{Let } u = \lambda y \to \frac{du}{dy} = \lambda \to dy = \frac{du}{\lambda} \\
&= \frac{1}{(k-1)!}\int_0^{\lambda x} u^{k-1} e^{-u}\,du \\
&= \frac{\gamma(k,\lambda x)}{(k-1)!}
\end{aligned}$$

The Gamma function is as follows:

$$\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt = \underbrace{\int_0^a t^{x-1}e^{-t}\,dt}_{\gamma(x,a)} + \underbrace{\int_a^\infty t^{x-1}e^{-t}\,dt}_{\Gamma(x,a)}$$

Let $T \sim \text{Exp}(\lambda) = \lambda e^{\lambda t}$ which describes the time between Poisson events. In fact, $F_T(t) = 1 - e^{-\lambda t}$.
Let $H \sim \text{Poisson}(\lambda) = \frac{e^{-\lambda}\lambda^n}{n!}$ which describes the number of events occurring within a time interval. In fact, $F_N(n) = \sum_{i=0}^n \frac{e^{-\lambda}\lambda^i}{i!} = e^{-\lambda}\sum_{i=0}^\infty \frac{\lambda^i}{i!}$.
What is the probability that no events have occurred by $t = 1$?

$$\mathbb{P}(T > 1) = e^{-\lambda} = \mathbb{P}(N = 0) = e^{-\lambda}$$

What is the probability that at least one event occurred before $t = 1$?

$$\mathbb{P}(T < 1) = 1 - e^{-\lambda} = \mathbb{P}(N > 0) = 1 - e^{-\lambda}$$

What is the probability of no successes or one success by $t = 1$?

$$\mathbb{P}(N \leq 1) = F_N(1) = e^{-\lambda}(1 + \lambda)$$

If $T \sim \text{Erlang}(2, \lambda)$, this scenario can be computed as

$$\mathbb{P}(T > 1) = 1 - F_T(1)$$

Let $X \sim \text{Erlang}(k, \lambda) = \frac{\lambda^k x^{k-1}e^{-\lambda x}}{(k-1)!}$. Then $F_X(x) = \frac{\gamma(k,\lambda x)}{(k-1)!}$. This comes from

$$\underbrace{\Gamma(x)}_{\text{gamma function}} = \int_0^\infty t^{x-1}e^{-t}\,dt = \underbrace{\int_0^a t^{x-1}e^{-t}\,dt}_{\underbrace{\gamma(x,a)}_{\text{lower incomplete gamma function}}} + \underbrace{\int_a^\infty t^{x-1}e^{-t}\,dt}_{\underbrace{\Gamma(x,a)}_{\text{upper incomplete gamma function}}}$$

The gamma function is known as an extension of the factorial function to all real numbers.

$$\begin{aligned}
\Gamma(1) &= \int_0^\infty t^{1-1}e^{-t}\,dt \qquad = -e^{-t}\Big|_0^\infty = -(0-1) = 1 \\
\Gamma(x+1) &= \int_0^\infty t^x e^{-t}\,dt = [-t^x e^{-t}]\Big|_0^\infty - \int_0^\infty -e^{-t}xt^{x-1}\,dt = x\Gamma(x) \\
\Gamma(2) &= 1 \cdot 1 \\
\Gamma(3) &= 2\Gamma(2) = 2 \cdot 1 \\
\Gamma(4) &= 3\Gamma(3) = 3 \cdot 2 \cdot 1 \\
&\vdots \\
\Gamma(n) &= (n-1)!
\end{aligned}$$

Thus
$$F_{T_k}(x) = \frac{\gamma(k,\lambda x)}{\Gamma(k)}$$
which is called the normalized gamma function.
$$1 - F_{T_k}(x) = 1 - \frac{\gamma(k,\lambda x)}{\Gamma(k)} = \frac{\Gamma(k,\lambda x)}{\Gamma(k)} = Q(l,\lambda x)$$
which is called the regularized gamma function, a proportion of the entire gamma.

We know that $k \in \mathbb{N}$, then
$$\begin{aligned}
\Gamma(k,\lambda x) &= \int_{kx}^\infty t^{k-1}e^{-t}\,dt \\
&= -t^{k-1}e^{-t}\Big|_{\lambda x}^\infty - \int_{\lambda x}^\infty (k-1)t^{k-2}(-e^{-t})\,dt \\
&= (\lambda x)^{k-1}e^{-\lambda x} + (k-1)\Gamma(k-1,\lambda x) \\
&= (\lambda x)^{k-1}e^{-\lambda x} + (k-1)\Big((\lambda x)^{k-2}e^{-\lambda x} + (k-2)\Gamma(k-2,\lambda x)\Big) \\
&= e^{-\lambda x}\Big((\lambda x)^{k-1} + (k-1)(\lambda x)^{k-2} + (k-2)(k-1)\frac{\Gamma(k-2,\lambda x)}{e^{-\lambda x}}\Big) \\
&= e^{-\lambda x}\Big(\frac{(\lambda x)^{k-1}}{(k-1)!} + \frac{(\lambda x)^{k-2}}{(k-2)!} + \dots + \underbrace{1}_{\Gamma(1,\lambda x)=\int_{\lambda x}^\infty t^{1-1}e^{-t}\,dt = e^{-\lambda x}}\Big) \\
&= e^{-\lambda x}(k-1)!\sum_{i=0}^{k-1}\frac{(\lambda x)^i}{i!}
\end{aligned}$$
Then
$$1 - F_{T_k}(x) = \frac{e^{-\lambda x}(k-1)!\sum_{i=0}^{k-1}\frac{(\lambda x)^i}{i!}}{(k-1)!} = e^{-\lambda x}\sum_{i=0}^{k-1}\frac{(\lambda x)^i}{i!}$$
Let $T \sim \text{Erlang}(2,\lambda)$, then
$$\mathbb{P}(T > 1) = 1 - F_{T_2}(1) = e^{-\lambda}\sum_{i=0}^{1}\frac{(\lambda\cdot 1)^i}{i!} = e^{-\lambda}(1+\lambda)$$
What is the probability of k successes or less by $t = 1$?
$$\mathbb{P}(N \leq k) = F_X(k) = e^{-\lambda}\sum_{i=0}^{k}\frac{\lambda^i}{i!}$$
If successes come exponentially, what is the probability of seeing k or fewer successes by 1 hr? Let $T \sim \text{Erlang}(k+1,\lambda)$. Then
$$\mathbb{P}(T > 1) = 1 - F(1) = e^{-\lambda}\sum_{i=0}^{k}\frac{\lambda^i}{i!}$$

Poisson Process: in every unit time, there are $X \sim \text{Poisson}(\lambda)$ "hits" and each hit occurs after $T \sim \text{Exp}(\lambda)$.

$$e^{-\lambda}\sum_{i=0}^{l}\frac{\lambda^i}{i!} = \frac{\Gamma(k+1,\lambda)}{\Gamma(k)} = Q(K+1,\lambda)$$

If we let $k \to \infty$ and $Q \to 1$, then

$$\sum_{i=0}^{k}\frac{a^i}{i!} = e^a Q(k+1,a)$$
$$e^a = \sum_{i=0}^{k}\frac{a^i}{i!}$$

Running experiments	fixed time, measure number of successes	require at least 1 success	require 1 success
discretely	Binomial	Negative Binomial	Geometric
continuously	Poisson	Erlang	Exponential

What is the probability that there has been 2 successes or less by $t = 50$?

$$\begin{aligned} N &\sim \text{Binom}(50,p) \\ \mathbb{P}(N \leq 2) = F_N(2) &= \binom{50}{0}(1-p)^{50} + \binom{50}{1}p(1-p)^{49} + \binom{50}{2}p^2(1-p)^{48} \\ T &\sim \text{NegBInom}(3,p) \end{aligned}$$

$$\begin{aligned} \mathbb{P}(T \geq 48) &= 1 - F_T(47) \\ &= 1 - \sum_{i=0}^{47}\binom{i+2}{2}p^3(1-p)I6 \end{aligned}$$

Let $N \sim \text{Binom}(n,p)$ and $T \sim \text{NegBinom}(k+1,p)$, then

$$F_N(K) = 1 - F_T(n-k-1)$$
$$\sum_{i=0}^{l}\binom{n}{i}p^i(1-p)^{n-i} = 1 - \sum_{i=0}^{n-k-1}\binom{i+k}{k}p^{k+1}(1-p)^i$$

Let $X_1, X_2 \overset{iid}{\sim} \text{Poisson}(\lambda)$. What is $\mathbb{P}(X_1 \mid X_1 + X_2)$?
What is $\mathbb{P}(X_1)$? This is $\mathbb{P}(X_1 = x) = \mathbb{P}_X(x)$. What is $\mathbb{P}(X_1 + X_2)$? This is the same as

$\mathbb{P}(X_1 + X_2 = n)$. Then

$$\begin{aligned}\mathbb{P}(X_1 = x \mid X_1 + X_2 = n) &= \frac{X_1 = x \text{ and } X_1 + X_2 = n}{\mathbb{P}(X_1 + X_2 = n)} \\ &= \frac{\mathbb{P}_{X_1, X_2}(x, n-x)}{\mathbb{P}_Y(n)} \\ &= \frac{\frac{e^{-\lambda}\lambda^x}{x!} \cdot \frac{e^{-\lambda}\lambda^{n-x}}{(n-x)!}}{\frac{e^{-2\lambda}(2\lambda)^n}{n!}} \\ &= \binom{n}{x}\Big(\frac{\lambda}{2\lambda}\Big)^n \\ &= \binom{n}{x}\Big(\frac{1}{2}\Big)^2 \\ &= \text{Binom}\Big(n, \frac{1}{2}\Big)\end{aligned}$$

START OF MIDTERM 2 MATERIAL

Transformation of Discrete Random Variables
Let $X \sim \text{Bern}(p) = p^x(1-p)^{1-x}\mathbb{1}_{x\in[0,1]} = \mathbb{P}_X(x)$. Let

$$Y = 3 + x \sim \begin{cases} 4 & \text{up } p \\ 3 & \text{up } 3 - p \end{cases} = p^{y-x}(1-p)^{1-(y-3)}\mathbb{1}_{y\in[3,4]} = \mathbb{P}_Y(y)$$

$\text{Supp}[Y] = \{y : y - 3 \in \text{Supp}[x]\}$ The pmf of Y looks like the pmf of X is replaced with $y - 3$.
Let $Y = c + aX = g(x)$. Then $x = \frac{y-c}{a} = g^{-1}(y)$.

$$\begin{aligned}\text{Supp}[Y] &= \{y : \frac{y-c}{a} \in \text{Supp}[X]\} \\ &= \{y : \frac{y-c}{a} \in [0,1]\} \\ &= \{c, a + c\}\end{aligned}$$

Let $\mathbb{P}_Y(y) = p^{\frac{y-c}{a}}(1-p)^{1-\frac{y-c}{a}}\mathbb{1}_{y\in\{c,a+c\}} = \mathbb{P}_X(g^{-1}(y))$. This is the modeling support.

Let $X \sim \text{Binom}(n,p)$. Let $Y = a + cX$. Then

$$\begin{aligned}\mathbb{P}_Y(y) &= \binom{n}{g^{-1}(y)}p^{g^{-1}(y)}(1-p)^{n-g^{-1}(y)}\mathbb{1}_{y\in g(\text{Supp}[X])} \\ &= \binom{n}{\frac{y-c}{a}}p^{\frac{y-c}{a}}(1-p)^{n-\frac{y-c}{a}}\mathbb{1}_{y\in\{c,a+c,2a+c,\dots,na+c\}}\end{aligned}$$

Let $X \sim \text{Binom}(n,p)$ and $Y = X^3$. Then

$$\mathbb{P}_Y(y) = \binom{4}{\sqrt[3]{y}}p^{\sqrt[3]{y}}(1-p)^{1-\sqrt[3]{y}}\mathbb{1}_{y\in\{0,1,2^3,3^3,\dots,n^3\}}$$

Let $X \sim \text{Geom}(p)$ and $Y = \max\{3, x\}$. This looks like

X	Y
0	3
1	3
2	3
3	3
4	4
5	5
$\vdots$	$\vdots$

There is no $g^{-1}(y)$ function because g is not 1-1. Note that $\mathbb{P}_Y(4) = \mathbb{P}_X(4)$, $\mathbb{P}_Y(5) = \mathbb{P}_X(5)$, but $\mathbb{P}_Y(3) \neq \mathbb{P}_X(3)$. In fact $\mathbb{P}_Y(3) = \mathbb{P}_X(0) + \mathbb{P}_X(1) + \mathbb{P}_X(2) + \mathbb{P}_X(3)$. Thus $\mathbb{P}_Y(y) \neq \mathbb{P}_X(g^{-1}(y))$. From this, conclude that this only works for g functions which are 1-1. In general, the formula for discrete random variable function

$$\mathbb{P}(Y)y = \sum_{\{x:g(x)=y} \mathbb{P}_X(x) = \sum_{\{x:x\in g^{-1}(y)} \mathbb{P}_X(x) = \mathbb{P}_X(g^{-1}(y))$$

In this example,

$$\begin{aligned}\mathbb{P}_Y(y) &= \Big(\mathbb{P}_X(0) + \mathbb{P}_X(1) + \mathbb{P}_X(2) + \mathbb{P}_X(3)\Big)\mathbb{1}_{y=3} + p(1-p)^y \mathbb{1}_{y\in\{4,5,\dots\}} \\ &= \Big(p + (1-p)p + (1-p)^2 p + (1-p)^3 p)\mathbb{1}_{y=3} + \underbrace{p(1-p)^y}_{\text{Geom}(p)} \mathbb{1}_{y\in\{4,5,\dots\}}\end{aligned}$$

Note that $F_Y(y) = \sum_{x:g(x)\leq y} \mathbb{P}_X(x)$.

Let $X_1, X_2 \overset{iid}{\sim} \text{Poisson}(\lambda)$ and $Y = -X_2$. Then $\mathbb{P}_Y(y) = \mathbb{P}_X(-y) = \frac{e^{-\lambda}\lambda^{-y}}{(-y)!}\mathbb{1}_{y\in\{0,-1,-2,\dots\}}$. Let $D = X_1 - X_2 = X_1 + Y$. $\text{Supp}[D] = \mathbb{Z}$. Then

$$\mathbb{P}_D(d) = \sum_{x\in\text{Supp}[X_1]} \mathbb{P}_{X_1}(x)\mathbb{P}_Y(d-x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} \frac{e^{-\lambda}\overbrace{\lambda^{-(d-x)}}^{\lambda^x\lambda^{-d}}}{\underbrace{(-(d-x))!}_{(x-d)!}} \mathbb{1}_{\underbrace{d-x\in\{0,-1,-2,\dots\}}_{\underbrace{x-d\in\{0,1,2,\dots\}}_{x\in\{d,d+1,d+2,\dots\}}}}$$

If $d > 0$, the sum begins at d; if $d \leq 0$, the sum begins at 0. Thus $\max\{0, d\}$.

$$\mathbb{P}_D(d) = e^{-2\lambda} \begin{cases} \sum_{x=d}^{\infty} \frac{\lambda^{2x-d}}{x!(x-d)!} & \text{if } d \leq 0 \text{ (upper)} \\ \sum_{x=0}^{\infty} \frac{\lambda^{2x-d}}{x!(x-d)!} & \text{if } d < 0 \text{ (lower)} \end{cases}$$

Let $x' = x - d \to x = x' + d$

$$= \sum_{x'=0}^{\infty} \frac{\lambda^{\overbrace{2(x'+d)-d}^{2x'-d}}}{(x'+d)!x'!} = \sum_{i=0}^{\infty} \frac{\left(\frac{2\lambda}{2}\right)^{2i-d}}{\Gamma(i+d-1)\Gamma(i-1)}$$

This is the modified Bessel function of the 1st kind denoted $I_D(2\lambda)$

Let $d' = -d$

$$= \sum_{x=0}^{\infty} \frac{\lambda^{2x+d'}}{x!(x+d')!} = \underbrace{\sum_{i=0}^{\infty} \frac{\left(\frac{2\lambda}{2}\right)^{2i-d'}}{\Gamma(i+d'-1)\Gamma(i-1)}}_{I_{d'}(2\lambda)}$$

If $d < 0 \to d' = |d|$

If $d > 0 \to d = |d|$

Thus

$$\mathbb{P}_D(d) = e^{-2\lambda} I_{|d|}(2\lambda) = \text{Skellam}(\lambda, \lambda)$$

This distribution is used to model point spreads in baseball, soccer, hockey, differences in photon noise, etc.

Let $X \sim U(0,1)$ and $Y = aX + c = g(X)$ such that g is 1-1. Can we use the formula $\mathbb{P}_Y(y) = \mathbb{P}_X(g^{-1}(y))$? No because there is no $\mathbb{P}_X(x)$ (pmf). It will not generalize for continuous random variables..

Consider $Y = g(X)$ where g is 1-1. Find $f_Y(y)$ given $f_X(x)$. If it's 1-1, it's either strictly increasing or strictly decreasing.

If g is increasing,

$$F_y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$$

To find the cdf of Y, just differentiate!

$$f_Y(y) = F_y'(y) = \frac{d}{dy}[F_X(g^{-1}(y))] = F_X'(g^{-1}(y))\frac{d}{dy}[g^{-1}(y)] = f_X(g^{-1}(y))\frac{d}{dy}[g^{-1}(y)]$$

On the other hand, if g is decreasing,

$$F_Y(y) = \mathbb{P}(g^{-1}(y) \leq y) = \mathbb{P}(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y))$$

Then

$$f_Y(y) = F_Y'(y) = \frac{d}{dy}[1 - F_X(g^{-1}(y))] = -\underbrace{f_X(g^{-1}(y))}_{\geq 0}\underbrace{\frac{d}{dy}[g^{-1}(y)]}_{\leq 0}$$

In general,

$$f_Y(y) = f_X(g^{-1}(y))\Big|\frac{d}{dt}[g^{-1}(y)]\Big|$$

Note: $\text{Supp}[Y] = g(\text{Supp}[X]) = \{g(x) : x \in \text{Supp}[Y]\} = \{y : g^{-1}(y) \in \text{Supp}[X]\}$.

If $Y = aX + c = g(X)$ where $a, c \in \mathbb{R}$ and $a \neq 0$ (the linear tranformation), then

$$y = ax + c \to x = \frac{y-c}{a} = g^{-1}(y) \to \Big|\frac{d}{dy}[g^{-1}(y)]\Big| = \frac{1}{|a|}$$

Thus

$$f_Y(y) = \frac{1}{|a|}f_X(\frac{y-c}{a})$$

Common Linear Transformations:

$$\text{If } Y = -X \to f_Y(y) = f_X(-y)$$

$$\text{If } Y = X + c \to f_Y(y) = f_X(y-c)$$

Let $X \sim U(0,1)$ and $Y = aX + c$. Then

$$f_Y(y) = \frac{1}{|a|}f_X(\frac{y-c}{a}) = \frac{1}{|a|}(1) = \frac{1}{|a|} \text{ where } \text{Supp}[Y] = [c, a+c] \text{ and so } Y \sim U(c, a+c)$$

Let $X \sim \text{Exp}(\lambda)$ and $Y = aX + c$. In fact, $\text{Supp}[Y] = (c, \infty)$.

$$f_Y(y) = \frac{1}{|a|}f_Y(\frac{y-c}{a}) = \frac{1}{|a|}\lambda e^{-\lambda(\frac{y-c}{a})}$$

Letting $c = 0$ and $a > 0$, this becomes

$$f_Y(y) = \frac{\lambda}{a}e^{-\frac{\lambda}{a}y} = \text{Exp}(\frac{\lambda}{a})$$

Let $X \sim U(0,1)$ and $Y = 1 - X$. Then $Y \sim U(0,1) = f_Y(y) = f_Y(y-1) = 1$ where $\text{Supp}[Y] = 1 - [0,1] = [0,1]$.

Let $Y = aX$, then $f_Y(y) = \frac{1}{|a|}f_X(\frac{y}{a})$.

Let $X \sim U(0,1)$ and $Y = -\ln(x)$. Then

$$f_Y(y) = \underbrace{f_X(g^{-1}(y))}_{1}\Big|\frac{d}{dy}[g^{-1}(y)]\Big| = \frac{d}{dy}[-e^{-y}] = e^{-y} = \text{Exp}(1)$$

Let $X \sim \text{Exp}(1)$ and $Y = -\ln\left(\frac{e^{-x}}{1-e^{-x}}\right) = \ln\left(\frac{1-e^{-x}}{e^{-x}}\right) = \ln(e^x - 1)$. Since $x \in (0,\infty)$, then $e^x \in (1,\infty)$ and so $e^x - 1 \in (0,\infty)$ and therefore $\ln(e^x - 1) \in (-\infty,\infty)$. Thus $\text{Supp}[Y] = \mathbb{R}$. To find $g^{-1}(y)$

$$\begin{aligned} y &= \ln(e^x - 1) \\ e^y &= e^x - 1 \\ e^x &= e^y + 1 \\ x &= \underbrace{\ln(e^y + 1)}_{g^{-1}(y)} \end{aligned}$$

Thus

$$\begin{aligned}
f_Y(y) &= f_X(g^{-1}(y))\Big|\frac{d}{dy}[g^{-1}(y)]\Big| \\
&= f_X(\ln(e^y+1))\Big|\frac{e^y}{e^y+1}\Big| \\
&= e^{-\ln(e^y+1)}\frac{e^y}{e^y+1} \\
&= e^{\ln(\frac{1}{e^y+1})}\frac{e^y}{e^y+1} \\
&= \frac{e^y}{(e^y+1)^2} \\
&= \text{Logistic}(0,1)
\end{aligned}$$

Let $X \sim U(0,1)$ and $Y = \ln(\frac{1}{x}-1) = g(x)$.

$$\begin{aligned}
x &\in [0,1] \\
\frac{1}{x} &\in (1,\infty) \\
\frac{1}{x} - 1 &\in (0,\infty) \\
\ln(\frac{1}{x}-1) &\in \mathbb{R} \\
-\ln(\frac{1}{x}-1) &\in \mathbb{R} \\
\text{Supp}[Y] &= \mathbb{R}
\end{aligned}$$

If $y = -\ln(\frac{1}{x}-1)$ then $g^{-1}(y) = \frac{1}{1+e^{-y}}$
Let

$$f(x) = \frac{L}{1+e^{-k(x-x_0)}}$$

be the logistic function where L is the max, k is the steepness and x_0 is the midpoint. If we let $L=1$, $x_0=0$ and $k=1$, we get the standard logistic function

$$f(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{1+e^x} = g(x)$$

Thus

$$f_Y(y) = f_X(g^{-1}(y))\Big|\frac{d}{dy}[g^{-1}(y)]\Big| = f_X\Big(\frac{1}{1+e^{-y}}\Big)\frac{e^{-y}}{(1+e^{-y})^2} = \frac{e^{-y}}{(1+e^{-y})^2} = \text{Logistic}(0,1)$$

By integrating this to get the CDF, we get

$$F_Y(y) = \frac{1}{1+e^{-y}}$$

This distribution looks like the normal distribution but has heavier tails.

Let $X \sim \text{Exp}(1)$ and $Y = ke^X$ such that $k \in (0,\infty)$. Supp$[X] = (0,\infty)$. If $k = 1$, Supp$[Y] = (1,\infty)$; otherwise for general k, Supp$[Y] = (k,\infty)$.

$$y = ke^x \to g^{-1}(y) = \ln\Big(\frac{y}{k}\Big)$$

Then

$$\begin{aligned}
f_Y(y) &= f_X\Big(\ln\frac{y}{k}\Big)y^{-1} \\
&= \lambda e^{-\lambda \ln\frac{y}{k}}y^{-1} \\
&= \lambda e^{\ln\left(\frac{k}{y}\right)^{\lambda}}y^{-1} \\
&= \lambda\Big(\frac{k}{y}\Big)^{\lambda}\frac{1}{y} \\
&= \frac{\lambda k^d}{y^{d+1}} \\
&= \text{Pareto}(k,d)
\end{aligned}$$

Then

$$F_Y(y) = \int_k^y \frac{\lambda k^d}{t^{d+1}}\,dt = 1 - \Big(\frac{k}{y}\Big)^d$$

This distribution is used to model

- population spreads - towns/cities
- survivals, hard drive failures
- surge of sand particles
- file size/ packet size in Internet traffic
- "Pareto Principle" - 1896 - 80% of the land in Italy was owned by 20% of the population

Let $X \sim \text{Pareto}(1, \overbrace{\log_4(5)}^{1.16})$.
What values of x has $p = \mathbb{P}(X \leq x)$ if continuous if $F_X^{-1}(p)$? Quantile$[x,p] = \overset{inf}{x}\{F(x) \geq p\}$.

$$\begin{aligned}
p = F_Y(p) &= 1 - (\frac{k}{y})^{\lambda} \\
1 - p &= \Big(\frac{k}{y}\Big)^{\lambda} \\
\Big(1-p\Big)^{\frac{1}{\lambda}} &= \frac{k}{y} \\
y &= k(1-p)^{-\frac{1}{\lambda}} = F_Y^{-1}(p)
\end{aligned}$$

For $X \sim \text{Pareto}(1, \log_4 5)$,

$$\begin{aligned} F_X^{-1}(p) &= (1-p)^{-0.86} \\ F_X^{-1}(0.8) = (1-0.8)^{-0.86} &= 4 \\ 1 - F_X(4) &= 1 - \left(\frac{1}{4}\right)^{1.16} = 0.8 \end{aligned}$$

Let $X, Y \overset{iid}{\sim} \text{Exp}(1)$ and $D = X - Y$. Let $Z = -Y$ such that $f_Z(z) = f_Y(-z) = e^z$. Then

$$\begin{aligned} D &= X + Z \\ &\sim \int_{\text{Supp}[X]} f_X(x) f_Z(d-x)\,dx \\ &= \int_0^\infty e^{-x} e^{d-x} \mathbb{1}_{d-x \in (-\infty, 0)}\,dx \\ &= e^d \int_0^\infty e^{-2x}\,dx \\ &= e^d \left[-\frac{1}{2} e^{-2x} \right]_{\max\{0,d\}}^\infty \\ &= \frac{1}{2} \begin{cases} e^d & \text{if } d \leq 0 \\ e^{-d} & \text{if } d > 0 \end{cases} \\ &= \frac{1}{2} e^{-|d|} = \text{Laplace}(0,1) \end{aligned}$$

The Laplace distribution is a "double Exponential" distribution.
1774 - "First Law of.. " - Imagine you're measuring a value V. Your measuring instrument is not perfect so you measure $Y \neq V$ but close so $Y = V + \varepsilon$ where ε is the error. It seems reasonable that $\text{E}[\varepsilon] = 0$ and so $\text{E}[Y] = V$. If $\text{Med}(\varepsilon) = 0$ then $\text{Med}(Y) = V$.

$$f_\varepsilon(\varepsilon) = f_\varepsilon(-\varepsilon)$$

Over/under numbers of the same magnitude are equiprobable.

$$f'(\varepsilon) < 0 \text{ if } \varepsilon > 0$$

and so

$$f'(\varepsilon) = f'(-\varepsilon) \to f(\varepsilon) = ce^{-mx}$$

It was figured out that $f(\varepsilon) \propto e^{-\varepsilon^2} = \text{Normal}$ when Gauss was 2 years old. This became the Second Law of Errors.

Let $X \sim \text{Exp}(1) = e^{-x}$ and $Y = -\ln X$ where $\text{Supp}[Y] = \mathbb{R}$.

$$y = \ln \frac{1}{x} \to g^{-1}(y) = e^{-y}$$

Then

$$\begin{aligned}
\Big|\frac{d}{dy}[g^{-1}(y)]\Big| &= e^{-y} \\
f_Y(y) &= f_X(e^{-y})e^{-y} \\
&= e^{-e^{-y}}e^{-y} \\
&= \exp\Big(-(y+e^{-y})\Big) \\
&= \text{Gumbel}(0,1)
\end{aligned}$$

This is the standard Gumbel distribution.

Let $X \sim \text{Gumbel}(0,1)$ and

$$Y = \mu + \beta X \sim \frac{1}{|\beta|} f_X\Big(\frac{y-\mu}{\beta}\Big) = \frac{1}{|\beta|}\exp\Big(-\Big(\frac{y-\mu}{\beta} + e^{-\frac{y-\mu}{\beta}}\Big)\Big) = \text{Gumbel}(\mu,\beta)$$

Parameter Space: $\beta > 0$, $\mu \in \mathbb{R}$.

$$\text{Gumbel}(\mu,\beta) = \frac{1}{\beta}\exp\Big(-\Big(\frac{y-\mu}{\beta} + e^{-\frac{(y-\mu)}{\beta}}\Big)\Big)$$

Let $X \sim \text{Exp}(1)$ and $Y = -\ln(X) = \ln(\frac{1}{X}) \sim \text{Gumbel}(0,1) = e^{-(y+e^{-y})} = e^{-y}e^{-e^{-y}}$ which is the standard Gumbel. Find the CDF of Gumbel. Let $Y \sim \text{Gumbel}(0,1)$.

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(-Y \geq -y) = \mathbb{P}(e^{-Y} \geq e^{-y}) = \mathbb{P}(X \geq e^{-y}) = 1 - F_X(e^{-y}) = e^{-e^{-y}}$$

If $X \sim \text{Gumbel}(0,1)$, then

$$Y = \mu + \beta X \sim \text{Gumbel}(\mu,\beta) = \frac{1}{\beta}e^{-(\frac{y-\mu}{\beta} + e^{-(\frac{y-\mu}{\beta})})}$$

Find the CDF of Gumbel.

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \leq y) \\
&= \mathbb{P}(\frac{Y-\mu}{\beta} \leq \frac{y-\mu}{\beta}) \\
&= \mathbb{P}(X \leq \frac{y-\mu}{\beta}) \\
&= F_X(\frac{y-\mu}{\beta}) \\
&= e^{-e^{-(\frac{y-\mu}{\beta})}}
\end{aligned}$$

Let $X \sim \text{Gumbel}(\mu, \beta)$ and $Y = e^{-X}$

$$\begin{aligned}
\text{Supp}[Y] &= (0,\infty) \\
x &= -\ln(y) = g^{-1}(y) \\
|\frac{d}{dy}g^{-1}(y)| &= y^{-1} \\
f_Y(y) &= f_X(-\ln(y))y^{-1} \\
&= \frac{1}{\beta}\exp\Big(-\Big(\frac{-\ln(y)-\mu}{\beta}\Big)\Big)\exp\Big(-\exp\Big(-\Big(\frac{-\ln y-\mu}{\beta}\Big)\Big)\Big) \\
&\text{Note } -(\frac{-\ln(y)-\mu}{\beta}) = \frac{\ln(y)+\mu}{\beta} \\
&\text{Let } k = \frac{1}{\beta} \text{ and } \mu = \ln(\lambda) \text{ where } \lambda \in (0,\infty) \\
\frac{\ln(y)+\mu}{\beta} &= k(\ln(y)+\ln(\lambda)) = \ln((y\lambda)^k) \\
f_Y(y) &= k\ \underbrace{(y\lambda)^k}_{y^k\ \underbrace{\lambda^k}_{\lambda\lambda^{k-1}}}\ e^{-(y\lambda)^k}y^{-1} \\
&= (k\lambda)(y\lambda)^{k-1}e^{-(y\lambda)^k} \\
&= \text{Weibull}(k,\lambda)
\end{aligned}$$

Note: If $k=1$, (thus $\beta = 1$ on the Gumbel), $\text{Weibull}(1,\lambda) = \lambda e^{-\lambda y} = \text{Exp}(\lambda)$. In addition,

$$\begin{aligned}
F_Y(y) &= \mathbb{P}(Y \le y) \\
&= \mathbb{P}(\ln(Y) \le \ln(y)) \\
&= \mathbb{P}(-\ln(Y) \ge -\ln(y)) \\
&= \mathbb{P}(X \ge -\ln(y)) \\
&= 1 - F_X(-\ln(y)) \\
&= 1 - \exp\Big(-\exp\Big(-(\frac{-\ln(y)-\mu}{\beta})\Big)\Big) \\
&= 1 - \exp\Big(-\exp\Big(\frac{\ln(y)+\mu}{\beta}\Big)\Big) \\
&= 1 - e^{-e^{\frac{\mu}{\beta}}y^{\frac{1}{\beta}}} \\
&= 1 - e^{-e^{\ln(y)^k}} \\
&= 1 - e^{-(\lambda x)^k}
\end{aligned}$$

If $\lambda = 1$ ($n = 0$ on the Gumbel), Weibull(1,1) = Exp(1).

The Weibull distribution is used to model survival time / failure times; it's a generalization of the exponential.

- If $k \neq 1$, then it is not memoryless
- If $k > 1$, $\mathbb{P}(X \geq a + b \mid X \geq a)$ gets smaller with a (dies quicker)
- If $k < 1$, $\mathbb{P}(X \geq a + b \mid X \geq a)$ gets larger with a (dies slower)
- If $k = 1$, no change

Let's say $k > 1$ (e.g. $k = 2$):
If $X \sim \text{Weibull}(2, \lambda)$, then $F_X(x) = 1 - e^{-(\lambda x)^2}$.

$$\mathbb{P}(X \geq b) > \mathbb{P}(X \geq a + b \mid X \geq a) = \frac{\mathbb{P}(X \geq a + b)}{\mathbb{P}(X \geq a)} = \frac{e^{-(\lambda(a+b))^2}}{e^{-(\lambda a)^2}} = \frac{e^{-(\lambda a)^2} e^{-2\lambda^2 ab} e^{-(\lambda b)^2}}{e^{-(\lambda a)^2}}$$

Then

$$e^{-\lambda^2 b^2} > e^{-2\lambda^2 ab} e^{-(\lambda b)^2}$$

This is

$$-\lambda b^2 > -\lambda(2ab + b^2) \to b^2 < 2ab + b^2$$

which is valid.

Let's say $k < 1$ (e.g. $k = \frac{1}{2}$), then $F_X(x) = 1 - e^{-(\lambda x)^{\frac{1}{2}}}$. Then

$$\mathbb{P}(X \geq b) < \mathbb{P}(X \geq a + b \mid X \geq a) = \frac{\mathbb{P}(X \geq a + b)}{\mathbb{P}(X \geq a)} = \frac{e^{-(\lambda(a+b))^{\frac{1}{2}}}}{e^{-(\lambda a)^{\frac{1}{2}}}} = e^{-(\lambda(a+b))^{\frac{1}{2}} + (\lambda a)^{\frac{1}{2}}}$$

Then

$$\begin{aligned}
e^{-(\lambda b)^{\frac{1}{2}}} = e^{-\lambda^{\frac{1}{2}} b^{\frac{1}{2}}} &< e^{-\lambda^{\frac{1}{2}}((a+b)^{\frac{1}{2}} - a^{\frac{1}{2}})} \\
-\lambda^{\frac{1}{2}} b^{\frac{1}{2}} &< -\lambda^{\frac{1}{2}}((a+b)^{\frac{1}{2}} - a^{\frac{1}{2}}) \\
b^{\frac{1}{2}} &> (a+b)^{\frac{1}{2}} - a^{\frac{1}{2}} \\
a^{\frac{1}{2}} + b^{\frac{1}{2}} &> (a+b)^{\frac{1}{2}} \\
(a^{\frac{1}{2}} + b^{\frac{1}{2}})^2 &> a + b \\
a + b + 2a^{\frac{1}{2}} b^{\frac{1}{2}} &> a + b
\end{aligned}$$

which is valid.

Let $X \sim$ Weibull and $Y = \frac{1}{X}$ (inverse waiting time).

$$\begin{aligned}
x &= \frac{1}{y} = g^{-1}(y) \\
|\frac{d}{dy} g^{-1}(y)| &= \frac{1}{y^2} \\
\text{Supp}[Y] &= (0, \infty) \\
f_Y(y) &= f_X(\frac{1}{y})\frac{1}{y^2} = (k\lambda)(\frac{\lambda}{y})^{k-1} e^{-(\frac{\lambda}{y})^k} \\
&= k\lambda^k \frac{1}{y^{\underbrace{k-1+2}_{k+1}}} e^{-\frac{\lambda^k}{y^k}} \\
&= \frac{k}{\lambda}(\frac{y}{\lambda})^{-(k+1)} e^{-(\frac{y}{\lambda})^{-k}} \\
&= \text{Frechet}(k, \lambda, \underbrace{0}_{\text{centered}})
\end{aligned}$$

Parameter space: $k \in (0, \infty)$, $\lambda \in (0, \infty)$.
If $X \sim \text{Frechet}(k, \lambda, 0)$, then $Y = X + c \sim \text{Frechet}(k, \lambda, c)$.

Note: Gumbel, Weibull and Frechet belong to a special family called the Generalized Extreme Value Distribution.
Units:

- Weibull: waiting time
- Frechet: inverse waiting time
- Gumbel: log inverse waiting time

Recall that $X \sim \text{Erlang}(k, \lambda) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{\Gamma(k)}$ and $X \sim \text{NegBinom}(k, p) = \underbrace{\binom{x+k-1}{k-1}}_{\frac{(x+k-1)!}{x!(k-1)!}} p^k (1-p)^x = \frac{\Gamma(x+k)}{\Gamma(x+1)\Gamma(k)} p^k (1-p)^x$. For both distributions, $k \in \mathbb{N}$ since it is a number of successes. What's wrong with allowing $k \in (0, \infty)$ i. e. all positive reals? You can show that the PDF of Erlang and PMF of negative binomial would still be valid. Conceptually? Wait for a fractional number of successes? Imagine "success" is initially continuous (such as success measured in dollars). If $k \in (0, \infty)$ these distributions got different names.

$X \sim \text{Gamma}(k, \lambda)$ useful due to flexible waiting time ditribution

$X \sim \text{ExtNegBinom}(k, \lambda)$ ignore this

The supports are $(0, \infty)$.

Let $X \sim \text{Gamma}(k_1, \lambda)$ and $Y \sim \text{Gamma}(k_2, \lambda)$. Then

$$\begin{aligned}
f_{X+Y}(t) &= \int_0^\infty \frac{\lambda^{k_1} x^{k_1-1} e^{-\lambda x}}{\Gamma(k_1)} \frac{\lambda^{k_2}(t-x)^{k_2-1} \overbrace{e^{-\lambda(t-x)}}^{e^{-\lambda t}e^{\lambda x}}}{\Gamma(k_2)} \mathbb{1}_{\underbrace{t-x}_{x \leq t} \in (0,\infty)} \, dx \\
&= \frac{\lambda^{k_1+k_2} e^{-\lambda t}}{\Gamma(k_1)\Gamma(k_2)} \int_0^t x^{k_1-1}(t-x)^{k_2-1} \, dx \\
&\text{Let } u = \frac{x}{t} \to \frac{du}{dx} = \frac{1}{t} \to dx = t\, du \\
&x = ut \to x_l = 0 \to u_l = 0, \; x_u = t \to u_u = 1 \\
&= \frac{\lambda^{k_1+k_2} e^{-\lambda t}}{\Gamma(k_1)\Gamma(k_2)} \int_0^1 (ut)^{k_1-1}(t-ut)^{k_2-1} \, du \\
&= \frac{\lambda^{k_1+k_2} e^{-\lambda t}}{\Gamma(k_1)\Gamma(k_2)} \int_0^1 t^{k_1-1} t^{k_2-1} u^{k_1-1}(1-u)^{k_2-1} t \, du
\end{aligned}$$

Let $X \sim \text{Gamma}(k_1, \lambda)$ and $Y \sim \text{Gamma}(k_2, \lambda)$ ($X, Y \overset{iid}{\sim}$). The Gamma distribution describes waiting time for k Exponential(λ) timed events where k could be fractional. Then

$$X + Y \sim \text{Gamma}(k_1 + k_2, \lambda)$$

$$\begin{aligned}
X + Y &\sim f_X(x) \times f_Y(y) \\
&= \frac{\lambda^{k_1+k_2} e^{-\lambda t}}{\Gamma(k_1)\Gamma(k_2)} \int_0^1 t^{k_1-1} t^{k_2-1} u^{k_1-1}(1-u)^{k_2-1} t \, du \\
&= \frac{\lambda^{k_1+k_2} e^{-\lambda t} t^{k_1+k_2-1}}{\Gamma(k_1)\Gamma(k_2)} \int_0^1 u^{k_1-1}(1-u)^{k_2-1} \, du
\end{aligned}$$

Recall that $X \sim \text{Exp}(\lambda) = f(x) = \lambda e^{-\lambda x}$ and $\int_{\text{Supp}[X]} f(x)\, dx = 1$. Note

$$f(x) = \lambda e^{-\lambda x} \propto e^{-\lambda x} = k(x)$$

Here, $k(x)$ is called the kernel of the Exponential distribution and is proportional to $f(x)$.

$$k(x) = cf(x) \to f(x) = \frac{1}{c} k(x) \text{ where } c \text{ is not a function of } x$$

$$1 = \int_{\text{Supp}[X]} f(x)\, dx = \int_{\text{Supp}[X]} \frac{1}{c} k(x)\, dx \to c = \int_{\text{Supp}[X]} k(x)\, dx$$

In this case, $\frac{1}{c} = \lambda$ and so $c = \frac{1}{\lambda}$

$$\int e^{-\lambda x}\, dx = \frac{1}{\lambda}$$

$k(x)$ can be restored to $f(x)$ by multiplying it by $\frac{1}{c}$.

Let $X \sim \text{Binom}(n, p) = p(x) = \binom{n}{x} p^x (1-p)^{n-x}$. Then

$$p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^n (1-p)^{-x} \propto \underbrace{(x!(n-x)!)^{-1} \left(\frac{p}{1-p}\right)^x}_{\text{identifies the Binomial}} = k(x)$$

Let $X \sim \text{Weibull}(k,\lambda) = f(x) = k\lambda(x\lambda)^{k-1}e^{-(x\lambda)^k}$. Then

$$p(x) \propto \underbrace{xe^{-(x\lambda)^k}}_{identifies the Weibull} = k(x)$$

Let $X \sim \text{Gamma}(k,\lambda) = f(x) = \frac{\lambda^k e^{\lambda x} x^{k-1}}{\Gamma(k)}$. Then

$$p(x) \propto e^{\lambda x}x^{k-1} = k(x)$$

Therefore,

$$f_{X+Y}(t) = \frac{\lambda^{k_1+k_2}e^{-\lambda t}t^{k_1+k_2-1}}{\Gamma(k_1)\Gamma(k_2)} \int_0^1 u^{k_1-1}(1-u)^{k_2-1}\,du \propto e^{-\lambda t}t^{k_1+k_2-1} \propto \text{Gamma}(k_1+k_2,\lambda)$$

As a corollary,

$$\begin{aligned} f_{X+Y}(t) &= \frac{\lambda^{k_1+k_2}e^{-\lambda t}t^{k_1+k_2-1}}{\Gamma(k_1+k_2)} \\ &= \frac{\lambda^{k_1+k_2}e^{-\lambda t}t^{k_1+k_2-1}}{\Gamma(k_1)\Gamma(k_2)} \int_0^1 u^{k_1-1}(1-u)^{k_2-1}\,du \\ \frac{1}{\Gamma(k_1)\Gamma(k_2)}\int_0^1 u^{k_1-1}(1-u)^{k_2-1}\,du &= \frac{1}{\Gamma(k_1+k_2)} \\ \int_0^1 u^{k_1-1}(1-u)^{k_2-1}\,du &= \frac{\Gamma(k_1)\Gamma(k_2)}{\Gamma(k_1+k_2)} \end{aligned}$$

Let $B(\alpha,\beta)$ be the beta function. Then

$$\begin{aligned} B(\alpha,\beta) &= \int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt \\ &= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} \\ &= \frac{\int_0^\infty t^{\alpha-1}e^{-t}\,dt \int_0^\infty t^{\beta-1}e^{-t}\,dt}{\int_0^\infty t^{\alpha+\beta-1}e^{-t}\,dt} \end{aligned}$$

Let $X_1, X_2, \dots, X_n$ be a sequence of continuous random variables. Then $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ denotes the order statistics where

$$\begin{aligned} X_{(1)} &= \min\Big\{X_1,\dots,X_n\Big\} \\ X_{(n)} &= \max\Big\{X_1,\dots,X_n\Big\} \\ X_{(k)} &= \Big\{k^{\text{th}} \text{ largest of } X_1,\dots,X_n\Big\} \end{aligned}$$

For example,

$$\begin{aligned} X_1 &= 9 = X_{(3)} \\ X_2 &= 2 = X_{(1)} \\ X_3 &= 12 = X_{(4)} \\ X_4 &= 7 = X_{(2)} \end{aligned}$$

Let $R = X_{(n)} - X_{(1)}$ be the range of the set under the assumption of $\stackrel{iid}{\sim}$ of $X_1, \dots X_n$. Let's first derive the distribution of the maximum.

$$\begin{aligned} 12 &= \max\{2,7,9,12\} \\ X_{(n)} &= \max\{X_1, \dots, X_n\} \end{aligned}$$

This means that all X_i's are less than $X_{(n)}$.

$$\begin{aligned} F_{X_{(n)}}(x) &= \mathbb{P}(X_{(n)} < x) \\ &= \mathbb{P}(X_1 < x, X_2 < x, \dots, X_n < x) \\ &= \prod_{i+1}^{n} \mathbb{P}(X_i < x) \\ &= \mathbb{P}(X_1 < x)^n \\ &= F(x)^n \end{aligned}$$

Then

$$f_{X_{(n)}}(x) = F'_{X_{(n)}}(x) = nf(x)F(x)^{n-1}$$

On the other side,

$$\begin{aligned} 2 &= \min\{2,7,9,12\} \\ X_{(1)} &= \min\{X_1, \dots, X_n\} \end{aligned}$$

This means that all X_i's are greater than $X_{(1)}$.

$$\begin{aligned} F_{X_{(1)}}(x) &= \mathbb{P}(X_{(1)} \leq x) \\ &= 1 - \mathbb{P}(X_{(1)} \geq x) \\ &= 1 - \mathbb{P}(X_1 \geq x, X_2 \geq, \dots, X_n \geq n) \\ &= 1 - \prod_{i=1}^{n} \mathbb{P}(X_i \geq n) \\ &= 1 - \mathbb{P}(X_i \geq x)^n \\ &= 1 - (1 - F(x))^n \end{aligned}$$

Then

$$f_{X_{(1)}} = n(-f(x))(-1)(1-F(x))^{n-1} = nf(x)(1-F(x))^{n-1}$$

What about $X_{(k)}$, the k^{th} largest of $X_1, \dots, X_n$? In our example, 9 is the third largest of $\Big\{2,7,9,12\Big\}$, and so $X_{(3)} = 9$.

Goal: $F_{X_{(k)}}(x)$, the CDF of the k^{th} largest random variable of $X_1, \dots, X_n$.
Consider $n = 10$. What is the $\mathbb{P}(X_1, \dots, X_4 \in (-\infty, x)$ and $X_5, \dots, X_{10} \in (x, \infty))$? It is

$$\begin{aligned} &\mathbb{P}(X_1 \leq x, \dots, X_4 \leq x, X_5 > x, \dots, X_{10} > x) \\ &\mathbb{P}(X_1 \leq x) \dots \mathbb{P}(X_4 \leq x)\mathbb{P}(X_5 > x) \dots \mathbb{P}(X_{10} > x) \\ &F(x)^4(1-F(x))^6 \end{aligned}$$

More generally, what is the $\mathbb{P}(\text{any } 4 \in (-\infty, x) \text{ and the other } 6 \in (x, \infty))$?

$$\begin{aligned}
&\mathbb{P}(\underbrace{X_1 \leq x, \dots, X_4 \leq x}_{\text{these 4 below}}, \underbrace{X_5 > x, \dots, X_{10} > x}_{\text{these 6 above}}) \\
&+ \mathbb{P}(\underbrace{X_{10} \leq x, X_7 \leq x, X_3 \leq x, X_9 \leq x}_{\text{these 4 below}}, \underbrace{X_1 > x, X_3 > x, \dots, X_8 > x}_{\text{these 6 above}}) \\
&+ \text{ all other possibilities} \\
&= \binom{10}{4} F(x)^4 (1 - F(x))^6
\end{aligned}$$

This looks like the binomial where $n = 10$ and $p = F(x)$. Then

$$\begin{aligned}
F_{X_{(4)}}(x) &= \mathbb{P}(X_{(4)} \leq x) \\
&= \binom{10}{4} F(x)^4 (1 - F(x))^6 + \binom{10}{5} F(x)^5 (1 - F(x))^5 + \dots + \binom{10}{10} F(x)^{10} (1 - F(x))^0 \\
&= \sum_{j=4}^{10} \binom{10}{j} F(x)^j (1 - F(x))^{10-j}
\end{aligned}$$

Generalizing this to arbitrary n and k:

$$F_{X_{(k)}}(x) = \sum_{j=k}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j}$$

Verify that this works for the max and min:

$$\begin{aligned}
F_{X_{(n)}}(x) &= \sum_{j=n}^{n} F(x)^j (1 - F(x))^{n-j} = \binom{n}{n} F(x)^n (1 - F(x))^{n-n} = F(x)^n \\
F_{X_{(1)}}(x) &= \sum_{j=1}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j} \\
&= \left(\sum_{j=0}^{n} \binom{n}{j} F(x)^j (1 - F(x))^{n-j} \right) - \binom{n}{0} F(x)^0 (1 - F(x))^{n-0} \\
&= \Big(F(x) + (1 - F(x)) \Big)^n - (1 - F(x))^n \\
&= 1 - (1 - F(x))^n
\end{aligned}$$

Note that

$$\begin{aligned}
f_{X_{(k)}}(x) &= F'_{X_{(k)}}(x) \\
&= \frac{d}{dt}\Big[\sum_{j=k}^{n}\binom{n}{j}F(x)^j(1-F(x))^{n-j}\Big] \\
&= \sum_{j=k}^{n}\frac{n!}{j!(n-j)!}\underbrace{\frac{d}{dx}[F(x)^j(1-F(x))^{n-j}]}_{F(x)^j(n-j)(1-F(x))^{n-j-1}(-f(x))+(1-F(x))^{n-j}jF(x)^{j-1}f(x)} \\
&= \sum_{j=k}^{n}\frac{n!}{(j-1)!(n-j)!}f(x)F(x)^{j-1}(1-F(x))^{n-j} \\
&- \sum_{j=k}^{n}\frac{n!}{j!(n-j-1)!}f(x)F(x)^{j}(1-F(x))^{n-j-1}
\end{aligned}$$

We can reindex this to end at $n-1$ since at n it is 0

$$\begin{aligned}
&= \sum_{j=k}^{n}\frac{n!}{(j-1)!(n-j)!}f(x)F(x)^{j-1}(1-F(x))^{n-j} \\
&- \sum_{j=k}^{n-1}\frac{n!}{j!(n-j-1)!}f(x)F(x)^{j}(1-F(x))^{n-j-1}
\end{aligned}$$

Reindex this again so that it sums from $k+1$ to n. Let $l=k+1$ so that $j=l-1$

$$\begin{aligned}
&= \sum_{j=k}^{n}\frac{n!}{(j-1)!(n-j)!}f(x)F(x)^{j-1}(1-F(x))^{n-j} \\
&- \sum_{l=k+1}^{n}\frac{n!}{(l-1)!\underbrace{(n-(l-1)-1)!}_{(n-l)!}}f(x)F(x)^{l-1}(1-F(x))^{\overbrace{n-(l-1)-1}^{n-l}}
\end{aligned}$$

Let $j=l$

$$\begin{aligned}
&= \sum_{j=k}^{n}\frac{n!}{(j-1)!(n-j)!}f(x)F(x)^{j-1}(1-F(x))^{n-j} \\
&- \sum_{j=k+1}^{n}\frac{n!}{(j-1)!(n-j)!}f(x)F(x)^{j-1}(1-F(x))^{n-j} \\
&= (a_k+a_{k+1}+\cdots+a_n)-(a_{k+1}+\cdots+a_n) \\
&= a_k
\end{aligned}$$

$$f_{X_{(k)}} = \frac{n!}{(k-1)!(n-k)!}f(x)F(x)^{k-1}(1-F(x))^{n-k}$$

Let $X_1,\dots,X_n \overset{iid}{\sim} \text{U}(0,1)$. Note: $f(x)=1$ and $F(x)=x$.

$$f_{X_n}(x) = n\underbrace{f(x)}_{1}\underbrace{F(x)^{n-1}}_{x^{n-1}} = nx^{n-1} = \text{Beta}(k,n-k+1)$$

In fact, $\text{Supp}[X_{(k)}] = \text{Supp}[X] = [0,1]$.

$$f_{X_{(1)}} = nf(x)(1-F(x))^{n-1} = n(1-x)^{n-1} = \text{Beta}(1,n)$$

Then

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} \underbrace{f(x)}_{1} \underbrace{(F(x))^{k-1}}_{x} (1 - \underbrace{F(x)}_{x})^{n-k} \propto \underbrace{x^{k-1}(1-x)^{n-k}}_{k(x)}$$

Recall that

$$f(x) = \frac{1}{c}k(x) \to \int_{\text{Supp}[X]} k(x)\,dx = c$$

Therefore

$$\int_0^1 x^{k-1}(1-x)^{n-k}\,dx = \int_0^1 x^{k-1}(1-x)^{(n-k+1)-1}\,dx = B(k, n-k+1)$$

thus

$$f_{X_{(k)}}(x) = \frac{1}{B(\alpha,\beta)} x^{k-1}(1-x)^{n-k+1-1} = \text{Beta}(k, n-k+1)$$

In general, $X \sim \text{Beta}(\alpha,\beta) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$ where $\text{Supp}[X] = (0,1)$ when $\alpha > 0$ and $\beta > 0$.

$$\int_{\text{Supp}[X]} f(x) = 1 \to \int_0^1 \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\,dx = \frac{B(\alpha,\beta)}{B(\alpha,\beta)} = 1$$

In fact,

$$F(x) = \mathbb{P}(X \le x) = \frac{1}{B(\alpha,\beta)} \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,dt = \frac{\overbrace{B(x,\alpha,\beta)}^{\text{incomplete beta function}}}{B(\alpha,\beta)} = \underbrace{I_X(\alpha,\beta)}_{\text{regularized incomplete beta function}}$$

What's the expected value of a Beta distribution?

$$\begin{aligned} \text{E}[X] &= \frac{1}{B(\alpha,\beta)} \int_0^1 x^{\alpha+1-1}(1-x)^{\beta-1}\,dx = \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} \\ &= \frac{\frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)}}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} \\ &= \frac{\alpha}{\alpha+\beta} \end{aligned}$$

Let $X \sim f(x)$. What if we know $x \in A$ where $A \subseteq \text{Supp}[X]$? Call this distribution Y. What is the distribution of this new random variable Y? To get $f_Y(y)$, let $z = \mathbb{1}_{x \in A} \sim \text{Bern}(\mathbb{P}(x \in A))$. In fact,

$$f_{X,Z}(x,z) = f(x)\mathbb{1}_{x\in A}^{z}\mathbb{1}_{x \notin A}^{1-z}$$

Then

$$f_{X|Z}(x,z) = \frac{f_{X,Z}(x,z)}{p_Z(z)} = \frac{f(x)\mathbb{1}_{x\in A}^{z}\mathbb{1}_{z\notin A}^{1-z}}{\mathbb{P}(x\in A)^z(1-\mathbb{P}(x\in A))^{1-z}}$$

Let $Y = X|Z = 1$. Then

$$f_Y(x) = f_{X|Z}(x,1) = \frac{f(x)}{\mathbb{P}(x\in A)}\mathbb{1}_{x\in A}$$

Is this a PDF?

$$\int_{\text{Supp}[Y]} \frac{f(x)}{\mathbb{P}(x\in A)}\mathbb{1}_{x\in A}\,dx = \int_A \frac{f(x)}{\mathbb{P}(x\in A)}\,dx = \frac{\mathbb{P}(x\in A)}{\mathbb{P}(x\in A)} = 1$$

Typical Truncations:
If $x \geq 9,\ f_Y(x) = \frac{f(x)}{1-F(9)}\mathbb{1}_{x\geq 9}$
If $x \leq 9,\ f_Y(x) = \frac{f(x)}{F(9)}\mathbb{1}_{x\leq 9}$
If $x \in (a,b),\ f_Y(x) = \frac{f(x)}{F(b)-F(a)}\mathbb{1}_{x\in(a,b)}$

Let $X \sim \text{Exp}(\lambda)$ and $x \geq 9$. Then

$$f_Y(y) = \frac{\lambda e^{-\lambda x}}{e^{-9\lambda}}\mathbb{1}_{x\geq 9} = \lambda e^{-\lambda(x-9)}\mathbb{1}_{x\geq 9}$$

Let $g : \mathbb{R}^n \to \mathbb{R}^n$ and be $1-1$. Let $\vec{X}$ be a vector random variable with dim $= n$ and $\vec{Y}$ be a vector random variable with dim $= n$.
If $f_{\vec{X}}(\vec{x}) = f_{X_1,\dots,X_n}(x_1,\dots,x_n)$ is known and $\vec{Y} = g(\vec{X})$, find $f_{\vec{Y}}(\vec{y}) = f_{Y_1,\dots,Y_n}(y_1,\dots,y_n)$.

$$\begin{aligned}
Y_1 &= g_1(X_1,\dots,X_n) \\
Y_2 &= g_2(X_1,\dots,X_n) \\
&\ \vdots \\
Y_n &= g_n(X_1,\dots,X_n) \\
&\text{Since } g \text{ is } 1-1, \exists h_1,\dots,h_n \text{ where} \\
X_1 &= h_1(Y_1,\dots,Y_n) \\
X_2 &= h_2(Y_1,\dots,Y_n) \\
&\ \vdots \\
X_n &= h_n(Y_1,\dots,Y_n)
\end{aligned}$$

Thus

$$f_{Y_1,\dots,Y_n}(y_1,\dots,y_n) = f_{X_1,\dots,X_n}\Big(h_1(y_1,\dots,y_n),\dots,h_n(y_1,\dots,y_n)\Big)\Big|J_h(y_1,\dots,y_n)\Big|$$

where

$$J_n = \det\left(\begin{pmatrix} \frac{\partial h_1}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \cdots & \frac{\partial h_n}{\partial y_n} \end{pmatrix}\right)$$

In the one dimensional case, $Y = g(X)$ and so $X = g^{-1}(y)$ and thus $J_h = \det\left([\frac{\partial g^{-1}(y)}{\partial y}]\right) = \frac{\partial g^{-1}(y)}{\partial y}$ and so $f_Y(y) = f_X(g^{-1}(y))|\frac{d}{dy}[g^{-1}(y)]|$.

Given $X_1, X_2 \overset{iid}{\sim}$

$$\begin{aligned} Y_1 &= \frac{X_1}{X_2} = g_1(X_1, X_2) \\ Y_2 &= X_2 = g_2(X_1, X_2) \\ X_1 &= Y_1 Y_2 = h_1(Y_1, Y_2) \\ X_2 &= Y_2 = h_2(Y_1, Y_2) \end{aligned}$$

Find $f_{Y_1}(y_1)$.

$$\begin{aligned} \frac{\partial h_1}{\partial y_1} &= y_2 \\ \frac{\partial h_1}{\partial y_2} &= y_1 \\ \frac{\partial h_2}{\partial y_1} &= 0 \\ \frac{\partial h_2}{\partial y_2} &= 1 \\ J_h &= \det\left(\begin{pmatrix} y_2 & y_1 \\ 0 & 1 \end{pmatrix}\right) = y_2 \cdot 1 - 0 \cdot y_1 = y_2 \end{aligned}$$

Then

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(y_1 y_2, y_2)|y_2|$$

and so

$$\begin{aligned} f_{Y_1}(y_1) &= \int_{\text{Supp}[Y_2]} f_{Y_1,Y_2}(y_1, y_2)\, dy_2 \\ &= \int_{\text{Supp}[Y_2]} f_{X_1,X_2}(y_1 y_2, y_2)|y_2|\, dy_2 \\ &\text{If } X_1, X_2 \text{ independent and positive} \\ &= \int_{\text{Supp}[X_2]} x_2 f_{X_1}(y_1, x_2) f_{X_2}(x_2)\, dx_2 \end{aligned}$$

Given $X_1, X_2 \overset{iid}{\sim}$

$$\begin{aligned} Y_1 &= \frac{X_1}{X_1 + X_2} = g_1(X_1, X_2) \\ Y_2 &= X_1 + X_2 = g_2(X_1, X_2) \\ X_1 &= Y_1 Y_2 = h_1(Y_1, Y_2) \\ X_2 &= Y_2 - Y_1 Y_2 = h_2(Y_1, Y_2) \end{aligned}$$

Then

$$\frac{\partial h_1}{\partial y_1} = y_2$$
$$\frac{\partial h_1}{\partial y_2} = y_1$$
$$\frac{\partial h_2}{\partial y_1} = -y_1$$
$$\frac{\partial h_2}{\partial y_2} = 1 - y_1$$
$$J_h = \det\left(\begin{pmatrix} y_2 & y_1 \\ -y_2 & 1-y_1 \end{pmatrix}\right) = y_2(1-y_1) - y_1(-y_2) = y_2$$

Therefore

$$f_{Y_1,Y_2}(y_1,y_2) = f_{X_1,X_2}(y_1y_2, y_2(1-y_1))|y_2|$$

and so

$$f_{Y_1}(y_1) = \int_{\underbrace{\text{Supp}[Y_2]}_{\text{Supp}[X_2]}} f_{X_1,X_2}(y_1y_2, y_2(1-y_1))|y_2|\,dy_2$$

If X_1, X_2 are independent and positive

$$f_Y(y_1) = \int_{\text{Supp}[Y_2]} y_2 f_{X_1}(y_1y_2) f_{X_2}(y_2(1-y_1))\,dy_2$$

Let $X_1 \sim \text{Gamma}(\alpha,\lambda)$ be independent of $X_2 \sim \text{Gamma}(\beta,\lambda)$. Let $Y_1 = \frac{X_1}{X_1+X_2}$. $\text{Supp}[Y_1] = (0,1)$ and $\text{Supp}[Y_2] = (0,\infty)$. What's the distribution of Y_1?

$$\begin{aligned}
f_{Y_1}(y_1) &= \int_0^\infty f_{X_1}(y_1y_2) f_{X_2}(y_2 - y_1y_2) y_2\,dy_2 \\
&= \int_0^\infty \frac{\lambda^\alpha (y_1y_2)^{\alpha-1} e^{-\lambda y_1 y_2}}{\Gamma(\alpha)} \frac{\lambda^\beta (y_2(1-y_1))^{\beta-1} e^{-\lambda y_2(1-y_1)}}{\Gamma(\beta)} y_2\,dy_2 \\
&= \frac{\lambda^{\alpha+\beta} y_1^{\alpha-1}(1-y_1)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty y_2^{\alpha+\beta-1} e^{\overbrace{-\lambda y_1 y_2 - \lambda y_2(1-y_1)}^{-\lambda(y_1y_2+y_2-y_1y_2)}}\,dy_2 \\
&= \frac{\lambda^{\alpha+\beta} y_1^{\alpha-1}(1-y_1)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty y_2^{\alpha+\beta-1} e^{-\lambda y_2}\,dy_2
\end{aligned}$$

Let $u = \lambda y_2 \to \frac{du}{dy_2} = \lambda \to dy_2 = \frac{1}{\lambda} du$ and note that $y_2 = \frac{u}{\lambda}$.

$$\begin{aligned}
f_{Y_1}(y_1) &= \frac{\lambda^{\alpha+\beta} y_1^{\alpha-1}(1-y_1)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \int_0^\infty \underbrace{\underbrace{(\frac{u}{\lambda})^{\alpha+\beta-1} e^{-u} \frac{1}{\lambda}}_{\frac{u^{\alpha+\beta-1}}{\lambda^{\alpha+\beta-1}} e^{-u} \frac{1}{\lambda}}}\, du \\
&= \frac{\lambda^{\alpha+\beta} y_1^{\alpha-1}(1-y_1)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)} \frac{1}{\lambda^{\alpha+\beta}} \underbrace{\int_0^\infty u^{\alpha+\beta-1} e^{-u}\, du}_{\Gamma(\alpha+\beta)} \\
&= \underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{\frac{1}{B(\alpha,\beta)}} y_1^{\alpha-1}(1-y_1)^{\beta-1} \\
&= \text{Beta}(\alpha, \beta)
\end{aligned}$$

Let $X \sim \text{U}(0,1)$ and $Y|X = x \sim \text{U}(0,x)$. What does $Y \sim f_Y(y)$ look like? Note: $\text{Supp}[Y] = [0,1]$.

$$\begin{aligned}
f_Y(y) &= \int_{\text{Supp}[X]} f_{X,Y}(x,y)\, dx \\
&= \int_{\text{Supp}[X]} f_{Y|X}(y,x) f_X(x)\, dx \\
&= \int_{\mathbb{R}} \frac{1}{x} \underbrace{\underbrace{\mathbb{1}_{y\in[0,x]}}_{0 \le y \le x}}_{x \ge y} (1) \mathbb{1}_{x\in[0,1]}\, dx \\
&= \int_y^1 \frac{1}{x}\, dx = \ln(x)\Big|_y^1 = -\ln(y)
\end{aligned}$$

Check:

$$\begin{aligned}
\int_0^1 f_Y(y)\, dy &= -\int_0^1 \ln(y)\, dy \\
&= -[y\ln(y) - y]\Big|_0^1 \\
&= [y - y\ln(y)]\Big|_0^1 \\
&= (1-0) - (0-0) = 1
\end{aligned}$$

What is $f_Y(y)$? It is the marginal density.

A download either takes on average of 10 mins with no network traffic or an average of 20 mins with network traffic. Network traffic occurs with probability of $\frac{2}{3}$.

$$Y \sim \begin{cases} \text{Exp}(\frac{1}{10}) & \text{with probability } \frac{1}{3} \\ \text{Exp}(\frac{1}{20}) & \text{with probability } \frac{2}{3} \end{cases}$$

A familiar way to describe this is as follows:
Let $X = \mathbb{1}_{\text{network traffic}} = \text{Bern}(\frac{2}{3}) = (\frac{2}{3})^x(\frac{1}{3})^{1-x}$. Then

$$Y|X \sim \text{Exp}((\frac{1}{20})^x(\frac{1}{10})^{1-x}) = (\frac{1}{20})^x(\frac{1}{10})^{1-x}e^{-(\frac{1}{20})^x(\frac{1}{10})^{1-x}y}$$

How long does a download take? Traffic isn't mentioned here so; we want to use unconditional probability $f_Y(y)$. Note: $\text{Supp}[Y] = (0,\infty)$. In general, $f_Y(y) = \int_{\text{Supp}[X]} f_{X,Y}(x,y)\,dx$. Here, X is discrete so:

$$\begin{aligned} f_Y(y) &= \sum_{x\in\text{Supp}[X]} f_{X,Y}(x,y) \\ &= \sum_{x\in\text{Supp}[X]} f_{Y|X}(y,x)p_X(x) \\ &= \sum_{x\in\left\{0,1\right\}} (\frac{1}{20})^x(\frac{1}{10})^{1-x}e^{-(\frac{1}{20})^x(\frac{1}{10})^{1-x}y}(\frac{2}{3})^x(\frac{1}{3})^{1-x} \\ &= \frac{2}{3}(\frac{1}{30}e^{-\frac{1}{20}y}) + \frac{1}{3}(\frac{1}{10}e^{-\frac{1}{10}y}) \\ &= \frac{2}{3}\text{Exp}(\frac{1}{20}) + \frac{1}{3}\text{Exp}(\frac{1}{10}) \end{aligned}$$

This is a mixture distribution or mixture model.

If the download took 25 mins, what is the probability there was network traffic?

$$\mathbb{P}(X = 1 \mid Y = 25 \text{ minutes}) = ?$$

$$p_{X|Y}(x,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{f_{Y|X}(y,x)p_X(x)}{f_Y(y)}$$

$$\begin{aligned} p_{X|Y}(1,25) &= \frac{\overbrace{(\frac{1}{20})e^{-\frac{1}{20}(25)}}^{29.6026}(\frac{2}{3})}{\frac{1}{3}\underbrace{(\frac{1}{10}e^{-\frac{1}{10}(25)})}_{1.21825} + \frac{2}{3}\underbrace{(\frac{1}{20}e^{-\frac{1}{20}(25)})}_{29.6826}} \\ &\approx 90\% \end{aligned}$$

Here, since Y is a mixture of uncountably many values from X. It's called a compound distribution. On the other hand, a mixture distribution is for at most countably many elements. If X is uncountable, then compound distribution.

Let car accidents be distributed as $Y \sim \text{Poisson}(\lambda)$. But λ is not the same for all drivers. It is drawn from a $\text{Gamma}(\alpha,\beta)$. $\text{Supp}[Y] = \mathbb{N}$ since all values of λ are valid. Here we will use compound distribution since X is a continuous random variable.

$$\begin{aligned} p_Y(y) &= \int_{\text{Supp}[X]} p_{Y|X}(y,x) f_X(x)\, dx \\ &= \int_0^\infty \frac{e^{-x}x^y}{y!} \frac{\beta^\alpha x^{\alpha-1} e^{-\beta x}}{\Gamma(\alpha)}\, dx \\ &= \frac{\beta^\alpha}{y!\Gamma(\alpha)} \int_0^\infty x^{y+\alpha-1} e^{-(\beta+1)x}\, dx \end{aligned}$$

Let $u = (\beta+1)x$. Then $x = \frac{u}{\beta+1}$. Furthermore, $\frac{du}{dx} = \beta + 1$ and $dx = \frac{1}{\beta+1}\, du$.

$$\begin{aligned} p_Y(y) &= \frac{\beta^\alpha}{y!\Gamma(\alpha)} \int_0^\infty (\frac{u}{\beta+1})^{y+\alpha-1} e^{-u} \frac{1}{\beta+1}\, du \\ &= \frac{\beta^\alpha}{y!\Gamma(\alpha)(\beta+1)^{y+\alpha}} \int_0^\infty y^{y+\alpha-1} e^{-u}\, du \\ &= \frac{\beta^\alpha \Gamma(y+\alpha)}{y!\Gamma(\alpha)(\beta+1)^{y+\alpha}} \end{aligned}$$

Let $k = \alpha$ and $= \frac{\beta}{1+\beta}$. Then $1 - p = \frac{1}{1+\beta}$. Thus

$$p_Y(y) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y-1)} p^k (1-p)^y = \text{ExtNegBinom}(p,k)$$

If $k \in \mathbb{N}$,

$$p_Y(y) = \binom{y+k+1}{k} p^k (1-p)^y = \text{NegBinom}(p,k)$$

The Negative Binomial distribution is an over-dispersed Poisson model.

Let $Y|X \sim \text{Binom}(n,x)$ and $X \sim \text{Beta}(\alpha,\beta)$. $\text{Supp}[Y] = \Big\{0, 1, 2, \dots, n\Big\}$. Then

$$\begin{aligned} p_Y(y) &= \int_{\text{Supp}[X]} p_{Y|X}(y,x) f_X(x)\, dx \\ &= \int_0^1 \binom{n}{y} x^y (1-x)^{n-y} \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}\, dx \\ &= \frac{\binom{n}{y}}{B(\alpha,\beta)} \int_0^1 x^{y+\alpha-1}(1-x)^{n-y+\beta-1}\, dx \\ &= \frac{\binom{n}{y}}{B(\alpha,\beta)} B(y+\alpha, n-y+\beta) \\ &= \text{BetaBinomial}(n,\alpha,\beta) \end{aligned}$$

The BetaBinomial distribution is an over-dispersed binomial distribution.

Let $X \sim \text{Gamma}(\alpha, \beta)$ and $Y|X \sim \text{Exp}(x)$. Then

$$\begin{aligned}
f_Y(y) &= \int_{\text{Supp}[X]} f_{Y|X}(y,x) f_X(x)\,dx \\
&= \int_0^\infty x e^{-xy} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}\,dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha+1-1} e^{-(\beta+y)x}\,dx \\
&\text{Let } u = (\beta+y)x \to x = \frac{u}{\beta+y} \to \frac{du}{dx} = \beta + y \to dx = \frac{1}{\beta+y}\,du \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty \frac{u^{\alpha+1-1}}{(\beta+y)^\alpha} e^{-u} \frac{1}{\beta+y}\,du \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)(\beta+y)^{\alpha+1}} \underbrace{\int_0^\infty u^{\alpha+1-1} e^{-u}\,du}_{\Gamma(\alpha+1)} \\
&= \frac{\Gamma(\alpha+1)}{\Gamma(\alpha)} \frac{\beta^{\alpha+1}}{\beta} \frac{1}{(\beta+y)^{\alpha+1}} \\
&= \frac{\alpha}{\beta}(1 + \frac{y}{\beta})^{-(\alpha+1)} \\
&= \text{Lomax}(\beta, \alpha)
\end{aligned}$$

The Lomax distribution is a survival distribution.

Let $a, b \in \mathbb{R}$. Then $z = a + bi \in \mathbb{C}$ (the set of complex numbers) where

$$i = \sqrt{-1} \to i^2 = -1 \to i^3 = -i \to i^4 = 1$$

Note: $\text{Re}[z] = a$, $\text{Im}[z] = b$.

$$\begin{aligned}
e^x &= \sum_{k=0}^\infty \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots \\
e^{itx} &= \sum_{k=0}^\infty \frac{(itx)^k}{k!} = 1 + itx - \frac{t^2x^2}{2!} - \frac{it^3x^3}{3!} + \frac{t^4x^4}{4!} + \frac{it^5x^5}{5!} + \dots \\
\sin(x) &= x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots \\
\cos(x) &= 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \dots \\
i\sin(x) &= itx - \frac{it^3x^3}{3!} + \frac{it^5x^5}{5!} + \dots \\
i\cos(x) &= 1 - \frac{t^2x^2}{2!} + \frac{t^4x^4}{4!} + \dots \\
e^{itx} &= \cos(tx) + i\sin(tx) \\
&\text{If } \pi = tx \to e^{-i\pi} = -1 \to e^{i\pi} + 1 = 0 \text{ (Euler's Identity)}
\end{aligned}$$

In the complex number system, $|z| = \sqrt{a^2+b^2} \in [0,\infty)$ is the complex norm and $\theta = \arctan\left(\frac{b}{a}\right) \in [-\pi,\pi]$ is the argument of z, or $\text{Arg}(z)$.

$$z = |z|e^{i\theta}$$

Define

$$L^1 := \left\{ f : \int_{\mathbb{R}} |f(x)|\,dx < \infty \right\}$$

Note that all PDFs are L^1 because they integrate to 1.
If $f \in L^1$, then there exists $\hat{f}$ defined as

$$\hat{f}(t) = \int_{\mathbb{R}} e^{-2\pi itx} f(x)\,dx$$

This is known as the Fourier transform of f. Note that $\hat{f}$ doesn't necessarily $\in L^1$. $f(x)$ is called the time domain and $\hat{f}(t)$ is called the frequency domain. In fact, $f(x)$ can be written as a sum of sines and cosines. Note that

$$\begin{aligned} \text{Re}[\hat{f}(t)] &= \text{ amplitude of frequency} \\ \text{Arg}[\hat{f}(0)] &= \text{ phase shift of wave} \end{aligned}$$

Let $\phi(t) = \hat{f}(-\frac{t}{2\pi}) = \int_{\mathbb{R}} e^{-itx} f(x)\,dx = \text{E}[e^{itx}]$ which is the expectation if $f(x)$ is a PDF of a random variable X.
Note: if $\hat{f} \in L^1$ then $f(x) = \int_{\mathbb{R}} e^{2\pi itx}\hat{f}(t)\,dt$ which is the inverse Fourier transform.
If $\phi(t) \in L^1$, let $u = -2\pi t$ and so $t = \frac{-u}{2\pi}$ and thus $\frac{du}{dt} = -2\pi$ and $dt = -\frac{1}{2\pi}\,du$. Therefore

$$f(x) = \int_{\mathbb{R}} e^{2\pi i(\frac{-u}{2\pi})x}\hat{f}(\frac{-u}{2\pi})(-\frac{1}{2\pi})\,du$$

When $t = \infty, u = -\infty$ and when $t = -\infty, u = \infty$. Then

$$\begin{aligned} f(x) &= \int_{\infty}^{-\infty} e^{2\pi i(\frac{-u}{2\pi})x}\hat{f}(-\frac{u}{2\pi})(-\frac{1}{2\pi})\,du \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iux}\hat{f}(\frac{-u}{2\pi})\,du \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iux}\phi(u)\,du \end{aligned}$$

$\phi_X(t)$ is the characteristic function of a random variable X.

$$\phi_X(t) = \text{E}[-itx] = \begin{cases} \sum_{x\in\text{Supp}[X]} e^{itx}p(x) & \text{if } x \text{ discrete} \\ \int_{\text{Supp}[X]} e^{itx}f(x)\,dx & \text{if } x \text{ continuous} \end{cases}$$

Properties:

- $\phi(0) = 1$ since $\text{E}[e^{i(0)x}] = \text{E}[1] = 1$.

- If X_1, X_2 independent and $Y = X_1 + X_2$, then

$$\begin{aligned}\phi_Y(t) &= \phi_{X_1+X_2}(t) \\ &= \mathrm{E}[e^{it(X_1+X_2)}] \\ &= \mathrm{E}[e^{itX_1}e^{itX_2}] \\ &= \mathrm{E}[e^{itX_1}]\mathrm{E}[e^{itX_2}] \\ &= \phi_{X_1}(t)\phi_{X_2}(t)\end{aligned}$$

- If $Y = aX + b$, $a, b \in \mathbb{R}$, then

$$\begin{aligned}\phi_Y(t) &= \mathrm{E}[e^{itY}] \\ &= \mathrm{E}[e^{it(aX+b)}] \\ &= \mathrm{E}[e^{itaX}e^{itb}] \\ &= e^{itb}\mathrm{E}[e^{itaX}] \\ &= e^{itb}\phi_X(at)\end{aligned}$$

- $\phi_X(t)$ is bounded by 1 and thus always exists

$$\begin{aligned}|\phi_X(t)| = |\mathrm{E}[e^{itx}]| &= |\int_{\mathbb{R}} e^{itx} f(x)\,dx| \\ &\leq \int_{\mathbb{R}} |e^{itx} f(x)|\,dx \leq \int_{\mathbb{R}} |e^{itx}||f(x)|\,dx \\ &= \int |f(x)|\,dx = 1\end{aligned}$$

Define $M_X(t) = \phi(\frac{t}{i}) = \mathrm{E}[e^{tx}]$. This is the moment generating function. It is not granted to exist for all functions thus characteristic functions are more powerful.

Consider $\phi'_X(t) = \frac{d}{dt}[\mathrm{E}[e^{itx}]] = \frac{d}{dt}[\int_{\mathbb{R}} e^{itx} f(x)\,dx] = \int_{\mathbb{R}} f(x)\frac{d}{dt}[e^{itx}]\,dx$. Does

$$\frac{d}{dt}[\int g(x,t)\,dx] \stackrel{?}{=} \int_{\mathbb{R}} \frac{\partial}{\partial t}[g(x,t)]\,dx$$

Conditions:

1. There exists $t \in A$ such that $\int_{\mathbb{R}} g(x,t)\,dx$ converges where $A = [a,b] \subset \mathbb{R}$
2. $g(x,t)$ continuous for all $t \in A$
3. $g'(x,t)$ continuous for all $t \in \mathbb{R}$
4. For all $t \in A$, $\int_{\mathbb{R}} \frac{\partial}{\partial t} g(x,t)\,dt$ converges uniformly

$$\phi'_X(t) = \int_{\mathbb{R}} f(x) i x e^{itx}\, dx$$

Consider $\phi'_X(0) = \int_{\mathbb{R}} f(x) i x\, dx = i \int_{\mathbb{R}} x f(x)\, dx = i\mathrm{E}[X]$ Then

$$\begin{aligned} \phi''_X(t) &= \int_{\mathbb{R}} f(x) i^2 x^2 e^{itx}\, dx \\ \phi''_X(0) &= i^2 \int_{\mathbb{R}} x^2 f(x)\, dx = i^2 \mathrm{E}[X^2] \\ \phi'''_X(0) &= i^3 \mathrm{E}[X^3] \end{aligned}$$

More Properties:

- $\mathrm{E}[X^n] = \frac{\phi_X^{(n)}(0)}{i^n}$
- For all $a < b$, $\mathbb{P}(x \in (a,b)) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t)\, dt$ (Inversion Theorem)
  Motivation if $\phi_X \in L^1$:

$$\begin{aligned} f(x) &= \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_X(t)\, dt \\ \mathbb{P}(X \in (a,b)) &= \int_a^b f(x) = \int_a^b \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_X(t)\, dt\, dx \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \Big(\int_a^b e^{-itx}\, dx \Big) \phi_X(t)\, dt \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t)\, dt \end{aligned}$$

- $\phi_X(t) = \phi_Y(t) \leftrightarrow X \overset{d}{=} Y$
- $\phi_{X_n}(t)$ is the characteristic function for X_n
  If for all t, $\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)$ then

$$\begin{aligned} \lim_{n\to\infty} F_{X_n}(x) &= F_X(x) \\ \lim_{n\to\infty} X_n &= X \\ X_n &\xrightarrow{d} X \end{aligned}$$

Let $X \sim \mathrm{Gamma}(k, \lambda)$.

$$\phi_X(t) = \int_0^\infty e^{itx} \frac{\lambda^k e^{-\lambda x} x^{k-1}}{\Gamma(k)}\, dx = \frac{\lambda^k}{\Gamma(k)} \int_0^\infty x^{k-1} e^{(it-\lambda)x}\, dx$$

Let $u = (\lambda - it)x \to x = \frac{u}{\lambda - it}$ and so $dx = \frac{1}{\lambda - it}\,du$. Then

$$\begin{aligned}
\phi_X(t) &= \frac{\lambda^k}{\Gamma(k)} \int_0^\infty \frac{u^{k-1}}{(\lambda - it)^{k-1}} e^{-u} \frac{1}{\lambda - it}\,du \\
&= \frac{\lambda^k}{\Gamma(k)(\lambda - it)^k} \underbrace{\int_0^\infty u^{k-1} e^{-u}\,du}_{\Gamma(k)} \\
&= \Big(\frac{\lambda}{\lambda - it}\Big)^k \\
&= \Big(1 - \frac{it}{\lambda}\Big)^{-k}
\end{aligned}$$

Let $X_1 \sim \text{Gamma}(k_1, \lambda)$ and $X_2 \sim \text{Gamma}(k_2, \lambda)$. Then $X_1 + X_2 \sim \text{Gamma}(k_1 + k_2, \lambda)$.

$$\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t) = \Big(\frac{\lambda}{\lambda - it}\Big)^{k_1}\Big(\frac{\lambda}{\lambda - it}\Big)^{k_2} = \Big(\frac{\lambda}{\lambda - it}\Big)^{k_1+k_2}$$

Let $X \sim \text{Poisson}(\lambda)$.

$$\begin{aligned}
\phi_X(t) &= \sum_{x=0}^\infty e^{itx} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=0}^\infty \frac{(e^{it})^x \lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=0}^\infty \frac{(\lambda e^{it})^x e^{-\lambda}}{x!} \cdot \frac{e^{-\lambda e^{it}}}{e^{-\lambda e^{it}}} \\
&= \frac{e^{-\lambda}}{e^{-\lambda e^{it}}} \underbrace{\sum_{x=0}^\infty \frac{(xe^{it})^x e^{-\lambda e^{it}}}{x!}}_{\text{PMF of Poisson } (\lambda e^{it})} \\
&= e^{\lambda(e^{it}-1)}
\end{aligned}$$

Let $X_1 \sim \text{Poisson}(\lambda_1)$ and $X_2 \sim \text{Poisson}(\lambda_2)$. Then $X_1 + X_2 \sim \text{Poisson}(\lambda_1 + \lambda_2)$

$$\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t) = e^{\lambda_1(e^{it}-1)} e^{\lambda_2(e^{it}-1)} = e^{(\lambda_1+\lambda_2)(e^{it}-1)}$$

Let $X_1, \dots, X_n \overset{iid}{\sim}$ with same distribution and finite mean μ and finite variance σ^2.

$$\begin{aligned}
\bar{X}_n &= \frac{1}{n}\sum_{i=1}^n x_i \\
\text{E}[\bar{X}] &= \mu \\
\text{Var}[\bar{X}] &= \frac{\sigma^2}{n}
\end{aligned}$$

Let $Z_n = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ (standardization). Here, $\text{E}[Z_n] = 0$ and $\text{Var}[Z_n] = 1$. What happens as $n \to \infty$?

$$\phi_{\bar{X}}(t) = \phi_{\sum X_i}(\frac{t}{n}) = (\phi_X(\frac{t}{n}))^n$$

$$\begin{aligned}
\phi_{Z_n}(t) &= \phi_{\bar{X}_n}(\frac{t}{\frac{\sigma}{\sqrt{n}}})e^{it(\frac{-\mu}{\frac{\sigma}{\sqrt{n}}})} \\
&= \phi_{\bar{X}_n}(\frac{t\sqrt{n}}{\sigma})e^{-\frac{it\mu\sqrt{n}}{\sigma}} \\
&= \phi_{\bar{X}_n}(\frac{t\sqrt{n}}{\sigma})e^{-\frac{it\mu\sqrt{n}}{\sigma}\frac{n}{n}} \\
&= \phi_{\bar{X}_n}(\frac{t\sqrt{n}}{\sigma})e^{-\frac{it\mu n}{\sigma\sqrt{n}}} \\
&= \left(\phi_X(\frac{t}{\sigma\sqrt{n}})\right)^n e^{-\frac{it\mu n}{\sigma\sqrt{n}}} \\
\lim_{n\to\infty}\phi_{Z_n}(t) &= \lim_{n\to\infty} e^{\ln\left(\phi_X(\frac{t}{\sigma\sqrt{n}})^n e^{-\frac{it\mu n}{\sigma\sqrt{n}}}\right)} \\
&= \lim_{n\to\infty} e^{n\ln\left(\phi_X(\frac{t}{\sigma\sqrt{n}})-\frac{it\mu n}{\sigma\sqrt{n}}\right)} \\
&= \lim_{n\to\infty} e^{n\left(\ln(\phi_X(\frac{t}{\sqrt{n}})-\frac{it\mu}{\sigma\sqrt{n}})\right)} \\
&= e^{\lim_{n\to\infty} n\left(\ln(\phi_X(\frac{t}{\sigma\sqrt{n}})-\frac{it\mu}{\sigma\sqrt{n}})\right)} \\
&= e^{\lim_{n\to\infty}\frac{\ln(\phi_X(\frac{t}{\sigma\sqrt{n}})-\frac{it\mu}{\sigma\sqrt{n}})}{\frac{1}{n}}\cdot\frac{\frac{t^2}{\sigma^2}}{\frac{t^2}{\sigma^2}}} \qquad = e^{\frac{t^2}{\sigma^2}\lim_{n\to\infty}\frac{\ln(\phi_X(\frac{t}{\sigma\sqrt{n}}))-\frac{it\mu}{\sigma\sqrt{n}}}{(\frac{t}{\sigma\sqrt{n}})^2}}
\end{aligned}$$

Let $u = \frac{t}{\sigma\sqrt{n}}$, then as $n \to \infty$, $u \to 0$.

$$\begin{aligned}
\lim_{n\to\infty}\phi_{Z_n}(t) &= e^{\frac{t^2}{\sigma^2}}\lim_{u\to 0}\frac{\ln(\phi_X(u)) - i\mu(1)}{u^2} \\
&= e^{\frac{t^2}{2\sigma^2}\lim_{u\to 0}\frac{\frac{\phi'(u)}{\phi(u)}-i\mu}{u}} \qquad\qquad = e^{\frac{t^2}{2\sigma^2}\lim_{u\to 0}\frac{d}{du}[\frac{\phi'(u)}{\phi(u)}]} \\
\lim_{u\to 0}\frac{d}{du}[\frac{\phi'(u)}{\phi(u)}] &= \lim_{n\to 0}\frac{\phi''(u)\phi(u) - (\phi'(u))^2}{\phi(u)^2} \\
&= \frac{\overbrace{\phi''(0)}^{i^2\mathrm{E}[X^2]}\overbrace{\phi(0)}^{1} - \overbrace{(\phi'(0))^2}^{(i\mu)^2}}{\underbrace{\phi(0)^2}_{1^2}} \\
&= i^2(\mathrm{E}[X^2] - \mu^2) = -\mu^2 \\
\phi_Z(t) &= e^{\frac{t^2}{2\sigma^2}-\sigma^2} = e^{-\frac{t^2}{2}}
\end{aligned}$$

Then

$$\begin{aligned}
f(x) &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}\phi(t)\,dt = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-itx}e^{-\frac{t^2}{2}}\,dt \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-(itx+\frac{t^2}{2})}\,dt
\end{aligned}$$

Note that $\frac{t^2}{2}+itx = (\frac{t}{\sqrt{2}}+\frac{\sqrt{2}ix}{2})^2 = (\frac{t}{\sqrt{2}}+\frac{\sqrt{2}ix}{2})^2+\frac{x^2}{2}$. Therefore

$$\begin{aligned}
f(x) &= \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\left((\frac{t}{\sqrt{2}+\frac{\sqrt{2}ix}{2})^2+\frac{x^2}{2}} \right)}\,dt \\
&= \frac{1}{2\pi}e^{-\frac{x^2}{2}}\int_{\mathbb{R}} e^{-\left(\frac{t}{\sqrt{2}}+\frac{\sqrt{2}ix}{2}\right)^2}\,dt
\end{aligned}$$

Let $y = \frac{t}{2}+\frac{\sqrt{2}ix}{2}$. Then $\frac{dy}{dt} = \frac{1}{\sqrt{2}}$ and $dt = \sqrt{2}\,dy$ Then

$$\begin{aligned}
f(x) &= \frac{1}{2\pi}e^{-\frac{x^2}{2}}\int_{\mathbb{R}} e^{-y^2}\sqrt{2}\,dy \\
&= \frac{1}{\pi\sqrt{2}}e^{-\frac{x^2}{2}}\underbrace{\int_{\mathbb{R}} e^{-y^2}\,dy}_{\sqrt{\pi}} \\
&= \frac{1}{\pi\sqrt{2}}e^{-\frac{x^2}{2}}\sqrt{\pi} \\
&= \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \\
&= N(0,1)
\end{aligned}$$

This is the standard normal distribution and what we just showed is the Central Limit Theorem. The Central Limit Theorem says that if $X_1,\dots,X_n \overset{iid}{\sim} f(\mu,\sigma^2)$ then

$$\sum_{i=1}^n X_i \overset{d}{\approx} N(n\mu, n\sigma^2)$$

if n is large enough, and

$$\bar{X} \overset{n}{\approx} N(\mu, \frac{\sigma^2}{n})$$

Let $X \sim N(0,1) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Then $\phi_X(t) = e^{-\frac{t^2}{2}}$. Furthermore, $\mathrm{E}[X] = 0$ and $\mathrm{Var}[X] = 1$. This is because $\lim_{n\to\infty} Z_n \overset{d}{=} X$ where $Z_n = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$. Then $\mathrm{E}[Z_n] = 0$ and $\mathrm{Var}[Z_n] = 1$ for all n.

Let $Y = \mu + \sigma X$, assuming $\sigma \in (0,\infty)$.

$$f_Y(y) = \frac{1}{|\sigma|}f_X(\frac{y-\mu}{\sigma}) = \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}}e^{-\frac{(\frac{y-\mu}{\sigma})^2}{2}} = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2\sigma^2}(y-\mu)^2} = N(\mu,\sigma^2)$$

This is the general Normal random variable where

$$\begin{aligned}
\mathrm{E}[Y] &= \mu + \sigma\overbrace{\mathrm{E}[X]}^{0} = \mu \\
\mathrm{Var}[Y] &= \sigma^2\underbrace{\mathrm{Var}[X]}_{1} = \sigma^2
\end{aligned}$$

$$\phi_Y(t) = e^{it\mu}\phi_X(\sigma t) = e^{it\mu}e^{-\frac{(\sigma t)^2}{2}} = e^{it\mu - \frac{\sigma^2 t^2}{2}}$$

Let $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$. What's $Y = X_1 + X_2$?

$$\begin{aligned}\phi_Y(t) &= \phi_{X_1}(t)\phi_{X_2}(t) \\ &= e^{it\mu_1 - \frac{\sigma_1^2 t^2}{2}} e^{it\mu_2 - \frac{\sigma_2^2 t^2}{2}} \\ &= e^{it\mu_1 + it\mu_2 - \left(\frac{\sigma_1^2 t^2}{2} + \frac{\sigma_2^2 t^2}{2}\right)} \\ &= e^{it(\mu_1+\mu_2) - \frac{t^2}{2}(\sigma_1^2 + \sigma_2^2)} \\ &= N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)\end{aligned}$$

On the other hand,

$$\begin{aligned}Y = X_1 + X_2 &= f_{X_1}(x) \cdot f_{X_2}(x) \\ &= \int_{\mathbb{R}} f_{X_1}(x) f_{X_2}(t-x)\,dx \\ &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{1}{2\sigma_1^2}(x-\mu_1)^2} \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{1}{2\sigma_2^2}(t-x-\mu_2)^2}\,du \\ &\text{Note: no indicator function because all values are valid} \\ &= \text{ lots of work} \\ &= N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)\end{aligned}$$

Let $X \sim N(\mu, \sigma^2)$ and $Y = e^X = g(X)$. Then $g^{-1}(y) = \ln(y) \to |\frac{d}{dy}[g^{-1}(y)]| = \frac{1}{y}$ where $\text{Supp}[Y] = (0, \infty)$.

$$f_Y(y) = f_X(g^{-1}(y))|\frac{d}{dy}[g^{-1}(y)]| = \frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{y}e^{-\frac{1}{2\sigma^2}(\ln(y)-\mu)^2} = \text{LogNormal}(\mu, \sigma^2)$$

If $X \sim \text{LogN}(\mu, \sigma^2)$, then $Y = \ln(X) \sim N(\mu, \sigma^2)$.

The LogN distribution is really cool. Consider the following situation. You have an amount of money Y_0. Every time period, Y changes based on a proportional change R_t. For example, $Y_1 = Y_0(1 + R_1)$. If $R_1 = 0.3$ and $Y_0 = 10$, then $Y_1 = 13$, an increase of 30%. Then, $Y_2 = Y_1(1 + R_2) = Y_0(1 + R_1)(1 + R_2)$, and so on.

$$Y_t = Y_0 \prod_{i=1}^{t}(1 + R_i) = Y_0 e^{\ln(\prod_{i=1}^t 1 + R_i)} = Y_0 e^{\sum_{i=1}^t \ln(1+R_i)}$$

Let $X_i = \ln(1 + R_I)$ and so $Y_t = Y_0 e^{\sum_{i=1}^t X_i}$. If t is large, $X = \sum_{i=1}^r X_i \overset{d}{\approx} N(t\mu_X, t\sigma_X^2)$ by the CLT.

$$e^X \approx \text{LogN}(t\mu_X, t\sigma_X^2) \approx \text{LogN}(t\mu_R, t\sigma_R^2)$$

What is μ_X? Let $R = 3$. Then

$$\ln(1 + 0.03) = 0.0296 \approx 0.03$$

If $R = -5$,
$$\ln(1 + -0.05) = -0.051 \approx -0.05$$
Thus
$$\ln(1 + x) \approx x$$
This is because of the Taylor series
$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$
and so $\mu_X \approx \mu_R$ and $\sigma_X \approx \sigma_R$.

Start off with \$1000. Assume starting number is $\overset{iid}{\sim} N(10\%, 10\%^2)$. What is the probability after 5 years that you have more that \$1650?
Let $Y_t = Y_0 e^X$. We need to scale the LogN.
Let $X \sim$ LogN. and
$$\begin{aligned} Y &= aX \\ &\sim \frac{1}{a} f_X(\frac{y}{a}) \\ &= \frac{1}{a} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\frac{y}{a}} e^{-\frac{1}{2\sigma^2}(\ln(\frac{y}{a}) - \mu)^2} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\ln(y) - (\mu + \ln(a)))^2} \\ &= \text{LogN}(\mu + \ln(a), \sigma^2) \end{aligned}$$
If $Y_t = 1000e^X$ and $X \approx N(50\%, 5(10\%)^2)$ $(\ln(1000) = 6.91)$. Then
$$\begin{aligned} Y_5 &\sim \text{LogN}(\overbrace{50\% + 6.91}^{7.41}, \overbrace{5(10\%)^2}^{.5^2}) \\ \mathbb{P}(Y_5 > 1650) &= 1 - F_{Y_5}(1650) = 1 - \text{plnorm}(1650, 7.41, 0.5) \approx 51.2\% \end{aligned}$$

Let $Z \sim N(0, 1)$. What's $Y = Z^2 \sim ?$ Note that $\text{Supp}[Y] = [0, \infty)$. Also note that $g(Z)$ is not a 1-1 function.
$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(Z \leq y) \\ &= \mathbb{P}(Z \in [-\sqrt{y}, \sqrt{y}]) \\ &= 2\mathbb{P}(Z \in [0, \sqrt{y}]) \\ &= 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx \\ &= 2(F_Z(\sqrt{y}) - \frac{1}{2}) \\ &= 2F_Z(\sqrt{y}) - 1 \end{aligned}$$

If so, then

$$\begin{aligned}
f_Y(y) + \frac{d}{dy}&[2F_Z(\sqrt{y}) - 1] \\
&= 2\frac{d}{dy}[F_Z(\sqrt{y})] \\
&= 2\frac{1}{2}y^{-\frac{1}{2}}F_Z'(\sqrt{y}) \\
&= \frac{1}{\sqrt{y}}\frac{1}{\sqrt{2\pi}}e^{-\frac{(\sqrt{y})^2}{2}} \\
&= \frac{1}{\sqrt{2}\sqrt{\pi}\sqrt{y}}e^{-\frac{y}{2}} \\
&\sim \chi_1^2
\end{aligned}$$

This is the Chi-Square distribution with degree of freedom of 1.
Recall that $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$ and so, if we let $u = \sqrt{t}$ and thus $\frac{du}{dt} = \frac{1}{2}\frac{1}{\sqrt{t}} \to dt = 2\sqrt{t}\,du = 2u\,du$.

$$\Gamma(\frac{1}{2}) = \int_0^\infty t^{-\frac{1}{2}}e^{-t}\,dt = \int_0^\infty \frac{1}{u}e^{-u^2}2u\,du = 2\int_0^\infty e^{-u^2}\,du = 2\frac{\sqrt{\pi}}{2} = \sqrt{\pi}$$

Then we can transform $f_Y(y)$ above into the following

$$f_Y(y) = \frac{(\frac{1}{2})^{\frac{1}{2}}y^{-\frac{1}{2}}e^{-\frac{y}{2}}}{\Gamma(\frac{1}{2})} = \text{Gamma}(\frac{1}{2}, \frac{1}{2})$$

Recall that $\text{Gamma}(\alpha, \beta) = \frac{\beta^\alpha, x^{\alpha-1}e^{-\beta x}}{\Gamma(x)}$.

Let $X_1 \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$ and $X_2 \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$. Then $X_1 + X_2 \sim \text{Gamma}(1, \frac{1}{2})$. Furthermore, if $X_1, \dots, X_k \overset{iid}{\sim} \text{Gamma}(\frac{1}{2}, \frac{1}{2})$ then

$$\sum_{i=1}^k X_i \sim \text{Gamma}(\frac{k}{2}, \frac{1}{2}) = \frac{(\frac{1}{2})^{\frac{k}{2}}x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{\Gamma(\frac{k}{2})} = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})}x^{\frac{k}{2}-1}e^{-\frac{x}{2}} = \chi_k^2$$

Note: $\chi_2^2 = \text{Exp}(\frac{1}{2})$.

If $X_1, \dots, X_k \overset{iid}{\sim} N(0,1)$., then $\sum_{i=1}^k X_i^2 \sim \chi_k^2$.

Let $X \sim \chi_k^2$ and $Y = \sqrt{X}$. $\text{Supp}[Y] = (0, \infty)$. Then $g^{-1}(y) = y^2$ and so $|\frac{d}{dy}[g^{-1}(y)]| = 2y$.

$$f_Y(y) = 2yf_X(y^2) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})}(y^2)^{\frac{k}{2}-1}e^{-\frac{y^2}{2}}2y = \frac{1}{2^{\frac{k}{2}-1}\Gamma(\frac{k}{2})}y^{k-1}e^{-\frac{y^2}{2}} \sim \chi_k$$

This is the Chi distribution with degree of freedom k.

Let $X \sim N(0,1)$. What's $|X| \sim$? Well, $X^2 \sim \chi^2$,

$$\sqrt{X^2} = |X| \approx \chi_1 = \sqrt{\frac{2}{\pi}}e^{-\frac{x^2}{2}} = 2(\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}})$$

Let $X \sim \chi^2_k$ and $Y = \frac{X}{k}$. What is its distribution? Let's do scales of Gammas $c \in (0, \infty)$. If $X \sim \text{Gamma}(\alpha, \beta)$ and $Y = cX$, then

$$\begin{aligned}
f_Y(y) &= \frac{1}{c} f_X(\frac{y}{c}) \\
&= \frac{\beta^\alpha (\frac{y}{c})^{\alpha-1} e^{-\frac{\beta y}{c}}}{c\Gamma(\alpha)} \\
&= \frac{\beta^\alpha y^{\alpha-1} e^{-\frac{\beta}{c} y}}{c^{x-1} c\Gamma(\alpha)} \\
&= \frac{(\frac{\beta}{c})^\alpha y^{\alpha-1} e^{-(\frac{\beta}{c})y}}{\Gamma(\alpha)} \\
&= \text{Gamma}(\alpha, \frac{\beta}{c})
\end{aligned}$$

Therefore if $X \sim \chi^2_k$, then $Y = \frac{X}{k} \sim \text{Gamma}(\frac{k}{2}, \frac{k}{2})$.

Let $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$ What's $R = \frac{X_1/k_1}{X_2/k_2} \sim$? $\text{Supp}[R] = (0, \infty)$. Note that the ratio of $\text{Gamma}(\frac{k_1}{2}, \frac{k_1}{2})$ to $\text{Gamma}(\frac{k_2}{2}, \frac{k_2}{2})$ is both independent and positive.
Recall that

$$R = \frac{V_1}{V_2} \sim \int_{\text{Supp}[V_2]} t f_{V_1}(rt) f_{V_2}(t)\, dt$$

Let $a = \frac{v_1}{2}$, then

$$V_1 \sim \frac{a^a x^{a-1} e^{-ax}}{\Gamma(a)}$$

Let $b = \frac{k_2}{2}$, then

$$v_2 \sim \frac{b^b x^{b-1} e^{-bx}}{\Gamma(b)}$$

Then $f_R(r)$ is as follows:

$$\begin{aligned}
R &\sim \int_0^\infty t\frac{a^a(rt)^{a-1}e^{-art}}{\Gamma(a)}\frac{b^bt^{b-1}e^{-bt}}{\Gamma(b)}\,dt \\
&= \frac{a^ab^br^{a-1}}{\Gamma(a)\Gamma(b)}\int_0^\infty t^{a+b-1}e^{-(ar+b)t}\,dt \\
&\text{Let } u=(ar+b)t \to t=\frac{1}{ar+b}u \to dt=\frac{1}{ar+b}\,du \\
&= \frac{a^ab^br^{a-1}}{\Gamma(a)\Gamma(b)}\int_0^\infty \frac{u^{a+b-1}}{(ar+b)^{a+b-1}}e^{-u}\frac{1}{ar+b}\,du \\
&= \frac{a^ab^br^{a-1}}{\Gamma(a)\Gamma(b)(ar+b)^{a+b}}\underbrace{\int_0^\infty u^{a+b-1}e^{-u}\,du}_{\Gamma(a+b)} \\
&= \frac{a^ab^b}{B(a,b)}r^{a-1}\quad\underbrace{\underbrace{(ar+b)^{-(a+b)}}_{b^{-(a+b)}(1+\frac{a}{b}r)^{-(a+b)}}}_{b^{-a}} \\
&= \frac{\left(\frac{a}{b}\right)^a}{B(a,b)}r^{a+1}(1+\frac{a}{b}r)^{-(a+b)} \\
&= \frac{\left(\frac{k_1}{k_2}\right)^{\frac{k_1}{2}}}{B\left(\frac{k_1}{2},\frac{k_2}{2}\right)}r^{\frac{k_1}{2}-1}\left(1+\frac{k_1}{k_2}r\right)^{-\frac{k_1+k_2}{2}} \\
&= F(k_1,k_2)
\end{aligned}$$

This is the F distribution with parameters called degrees of freedoms k_1 and k_2. The F distribution, or Fisher-Snedecor distribution, or F for Fisher, comes up all over statistics especially when testing effects in linear models $k_1,k_2\in\mathbb{N}$ but the distribution is defined for $k_1,k_2\in(0,\infty)$ due to the Gamma function.

Consider $Z\sim N(0,1)$ and $V\sim\chi^2_k$. Let $W=\frac{Z}{\sqrt{\frac{V}{k}}}$. What's its distribution? Consider $W^2=\frac{Z^2}{\frac{V}{k}}$. $Z^2\sim\chi^2_1$ and so $\frac{Z^2}{1}\sim\text{Gamma}(\frac{1}{2},\frac{1}{2})$. $\frac{V}{k}\sim\text{Gamma}(\frac{k}{2},\frac{k}{2})$. Therefore

$$W^2\sim F(1,k)=\frac{\left(\frac{1}{k}\right)^{\frac{1}{2}}}{B\left(\frac{1}{2},\frac{k}{2}\right)}w^{-\frac{1}{2}}(1+\frac{1}{k}w)^{-\frac{1+k}{2}}=\frac{1}{\sqrt{k}B\left(\frac{1}{2},\frac{k}{2}\right)}w^{-\frac{1}{2}}(1+\frac{w}{k})^{-\frac{k+1}{2}}$$

So to get the distribution of W, find the square root of $F(1,k)$.
Let $X\sim F(1,k)$ and $Y=\pm\sqrt{X}$ (not a simple 1-1 function). But Y is symmetric around 0.

$$F_Y(y)=\mathbb{P}(Y\in[-y,y])=\mathbb{P}(Y^2\le y^2)=\mathbb{P}(X\le y^2)=F_X(y^2)$$

Take $\frac{d}{dy}$ of both sides.

$$\frac{d}{dy}[F_Y(y) - F_Y(-y)] = \frac{d}{dy}[F_X(y^2)]$$

Then

$$\begin{aligned} f_Y(y) - -f_Y(-y) &= f_X(y^2) \cdot 2y \\ 2f_Y(y) &= f_X(y^2) \cdot 2y \\ f_Y(y) &= f_X(y^2)y \end{aligned}$$

Therefore

$$\begin{aligned} f_Y(y) &= f_X(y^2)y \\ &= \frac{1}{\sqrt{k}B\left(\frac{1}{2}, \frac{k}{2}\right)} \underbrace{(y^2)^{-\frac{1}{2}}}_{\frac{1}{y}} (1 + \frac{y^2}{k})^{-\frac{k+1}{2}} y \\ &= \frac{1}{\sqrt{k}B\left(\frac{1}{2}, \frac{k}{2}\right)} (1 + \frac{y^2}{k})^{-\frac{k+1}{2}} \\ &= T_k \end{aligned}$$

This is the Student's T distribution with k degrees of freedom.

Let $Z \sim N(0,1)$ and $V \sim \chi_k^2$, then

$$\frac{Z}{\sqrt{\frac{V}{k}}} \sim T_k$$

If $V \sim T_k$, what's $\lim_{k\to\infty} V$?

$$\lim_{k\to\infty} \frac{1}{\sqrt{k}B\left(\frac{1}{2}, \frac{k}{2}\right)} (1 + \frac{y^2}{k})^{-\frac{k+1}{2}} = \lim_{k\to\infty} \frac{1}{\sqrt{k}B\left(\frac{1}{2}, \frac{k}{2}\right)} \lim_{k\to\infty} (1 + \frac{y^2}{k})^k \lim_{k\to\infty} (1 + \frac{y^2}{k})^{-\frac{1}{2}}$$

$$\lim_{k\to\infty} (1 + \frac{y^2}{k})^k \lim_{k\to\infty} (1 + \frac{y^2}{k})^{-\frac{1}{2}} = (e^{y^2})^{-\frac{1}{2}} \cdot 1 = e^{-\frac{y^2}{2}}$$

$$\lim_{k\to\infty} \frac{1}{\sqrt{k}B\left(\frac{1}{2}, \frac{k}{2}\right)} = \lim_{k\to\infty} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{k}{2}\right)} = \frac{1}{\sqrt{\pi}} \lim_{k\to\infty} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k}\Gamma\left(\frac{k}{2}\right)}$$

If n gets large, $\Gamma(n) \approx \sqrt{2\pi(n-1)}\left(\frac{n-1}{e}\right)^{n-1}$. Note that $\frac{k+1}{2} - 1 = \frac{k-1}{2}$ and $\frac{k}{2} - 1 = \frac{k-2}{2}$.

Thus

$$\begin{aligned}
\frac{1}{\sqrt{\pi}}\lim_{k\to\infty}\frac{\Gamma\Big(\frac{k+1}{2}\Big)}{\sqrt{k}\Gamma\Big(\frac{k}{2}\Big)} &= \frac{1}{\sqrt{\pi}}\lim_{k\to\infty}\frac{\sqrt{2\pi\Big(\frac{k-1}{2}\Big)}\Big(\frac{k-1}{2e}\Big)^{\frac{k-1}{2}}}{\sqrt{k}\sqrt{2\pi\Big(\frac{k-2}{2}\Big)}\Big(\frac{k-2}{2e}\Big)^{\frac{k-2}{2}}} \\
&= \frac{1}{\sqrt{\pi}}\lim_{k\to\infty}\frac{(k-1)^{\frac{k}{2}}(k-1)^{-\frac{1}{2}}(2e)^{\frac{k}{2}-1-\frac{k}{2}+\frac{1}{2}}}{(k-2)^{\frac{k}{2}}(k-2)^{-1}} \\
&= \frac{1}{\sqrt{\pi}}\lim_{k\to\infty}\sqrt{\frac{k-2}{k}}\Big(\frac{k-1}{k-2}\Big)^{\frac{k}{2}}(2e)^{-\frac{1}{2}} \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{e}}\underbrace{\lim_{k\to\infty}\sqrt{1-\frac{2}{k}}}_{1}\Big(\lim_{k\to\infty}\Big(\frac{k-1}{k-2}\Big)^k\Big) \\
&\text{Let } l = k-2 \to k = l+2 \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{e}}\Bigg(\lim_{l\to\infty}\Big(\frac{l+1}{l}\Big)^{l+2}\Bigg)^{\frac{1}{2}} \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{e}}\Bigg(\underbrace{\lim_{l\to\infty}\Big(1+\frac{1}{l}\Big)^l}_{e}\underbrace{\lim_{l\to\infty}\Big(1+\frac{1}{l}\Big)^2}_{1}\Bigg)^{\frac{1}{2}} \\
&= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{e}}\sqrt{e} = \frac{1}{\sqrt{2\pi}}
\end{aligned}$$

Thus

$$\lim_{k\to\infty} T_k = \frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}} = N(0,1)$$

Let $X_1 \sim N(0,1)$ and $X_2 \sim N(0,1)$. What's $R = \frac{X_1}{X_2} \sim ?$ Let $Z \sim N(0,1)$ and $V \sim \chi_k^2$, then

$$\frac{Z}{\sqrt{\frac{V}{k}}} \sim T_k$$

Note that $X_2^2 = \chi_1^2$. Then

$$\frac{X_1}{X_2} = \frac{X_1}{\sqrt{\frac{X_2^2}{1}}} \sim T_1 = \frac{\overbrace{\Gamma\Big(\frac{1+1}{2}\Big)}^{\Gamma(1)=1}}{\sqrt{(1)\pi}\underbrace{\Gamma\Big(\frac{1}{2}\Big)}_{\sqrt{\pi}}}\Big(1+\frac{X^2}{1}\Big)^{\overbrace{-\frac{1+1}{2}}^{-1}} = \frac{1}{\pi}\frac{1}{1+X^2} = \text{Cauchy}(0,1)$$

The Cauchy distribution is a special case of the T distribution.

Let $X \sim \text{Cauchy}(0,1)$. Then $Y = c+\sigma X \sim \frac{1}{\sigma}f_X(\frac{y-c}{\sigma}) = \frac{1}{\pi\sigma}\frac{1}{1+(\frac{y-c}{\sigma})^2} = \text{Cauchy}(c,\sigma)$.

$$\text{E}[X] = \int_{\mathbb{R}} x\frac{1}{\pi}\frac{1}{1+x^2}\,dx = \frac{1}{\pi}\int_{\mathbb{R}}\frac{x}{1+x^2}\,dx = \frac{1}{\pi}\Big[\frac{1}{2}\ln(x^2+1)\Big]_{-\infty}^{\infty} = \infty$$

This means that μ doesn't exist. Likewise, the variance does not exist and no moments exists.
$$M_X(t) = \text{E}[e^{tX}] = \int_{\mathbb{R}} e^{tx}\frac{1}{\pi}\frac{1}{1+x^2}\,dx = \infty$$
and so no moment generating function. The characteristic function is difficult to prove but
$$\phi_X(t) = e^{-|t|}$$
and
$$\phi_X'(t) = \frac{-t}{|t|}e^{-|t|}$$
but $\phi_X'(0)$ does not exist.

It is also called the Lorentz distribution. Why? Imagine you have a source of light at $y = 1$ above the origin and it shines light equally in all directions. What does the light density look like on the x-axis? The light shines on all angles so $\theta \sim U(\pi, 2\pi) = \frac{1}{\pi}$. $\tan(\theta) = \frac{x}{1}$. If $X = \tan(\theta) = g(\theta)$, then $\theta = \arctan(x) = g^{-1}(x)$ and $|\frac{d}{dx}[g^{-1}(x)]| = \frac{1}{1+x^2}$. Then
$$f_X(x) = f_\theta(g^{-1}(x))\frac{d}{dx}[g^{-1}(x)] = \frac{1}{\pi}\frac{1}{1+x^2}$$
Proof of Cauchy: Let $R = \frac{X_1}{X_2} \sim \int_{\text{Supp}[X_2]} |x_2| f_{X_1}(x_2 r) f_{X_2}(x_2)\,dx_2$,
$$\begin{aligned}
f_R(r) &= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}e^{-\frac{x_x^2 r^2}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{x_2^2}{2}}\,dx_2 \\
&= \frac{1}{2\pi}\int_{\mathbb{R}} |x_2| e^{-\frac{1}{2}x_2^2(r^2+1)}\,dx_2 \\
&= \frac{1}{2\pi}\left(\int_{-\infty}^0 -x_2 e^{-\frac{1}{2}x_2^2(r^2+1)}\,dx_2 + \int_0^\infty x_2 e^{-\frac{1}{2}x_2^2(r^2+1)}\,dx_2\right) \\
&= \frac{1}{2\pi}\left(\int_0^\infty x_2 e^{-\frac{1}{2}x_2^2(r^2+1)}\,dx_2 - \int_{-\infty}^0 x_2 e^{-\frac{1}{2}x_2^2(r^2+1)}\,dx_2\right)
\end{aligned}$$
Let $u = -\frac{1}{2}x_2^2(r^2+1)$. Then $\frac{du}{dx_2} = -x_2(r^2+1) \to dx = -\frac{1}{x_2(r^2+1)}\,du$. Note that at $x_0 = 0$, $u = 0$, at $x_0 = \infty$, $u = -\infty$ and at $x_0 = -\infty$, $u = -\infty$. So the integral becomes
$$\begin{aligned}
R &\sim \frac{1}{2\pi}\left(\int_0^\infty x_2 e^{-\frac{1}{2}x_2^2(r^2+1)}\,dx_2 - \int_{-\infty}^0 x_2 e^{-\frac{1}{2}x_2^2(r^2+1)}\,dx_2\right) \\
&= \frac{1}{2\pi}\left(\int_0^{-\infty} x_2 e^u\Big(-\frac{1}{x_2}(r^2+1)\Big)\,du - \int_{-\infty}^0 x^2 e^u\Big(-\frac{1}{x_2(r^2+1)}\Big)\right) \\
&= \frac{1}{2\pi}\Big(-\frac{1}{r^2+1}\Big)\left(\underbrace{[e^u]\Big|_0^{-\infty} - [e^u]\Big|_{-\infty}^0}_{(0-1)-(1-0)}\right) \\
&= -\frac{1}{2\pi}\frac{1}{r^2+1}(-2) \\
&= \frac{1}{\pi}\frac{1}{r^2+1}
\end{aligned}$$

START OF FINAL MATERIAL

Let $X_1, \dots, X_n \overset{iid}{\sim} f(\mu, \sigma^2)$ (some distribution). $\bar{X}$ is the average random variable. It is often used as an estimator for μ. It has nice properties, such as $\mathrm{E}[\bar{X}] = \mu$ (unbiased: on average, it is spot on). $\bar{x}$ is a realization from $\bar{X}$. $\bar{x}$ is an estimate of μ. This is why you use the sample average to estimate the mean.
How to estimate σ^2? More rare but definitely common.

$$s^2 = \frac{1}{n-1}\sum(x_i - \bar{x})^2 \text{ the sample variance estimate}$$

$$S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2 \text{ the sample variance estimator}$$

Therefore s^2 is a realization from S^2 and $\mathrm{E}[S^2] = \sigma^2$ which is also unbiased.

Assume $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. We know that $X_1 + \cdots + X_n \sim N(n\mu, n\sigma^2)$ from using characteristic functions. If $\bar{X} = \frac{X_1 + \cdots + X_n}{n} \sim N(\mu, \frac{\sigma^2}{n})$ then what's the distribution of $S^2 = \frac{1}{n-1}((X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2)$?

Let $Z_1, \dots, X_n \overset{iid}{\sim} N(0,1)$ and $\vec{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$. Note that $\vec{Z}^T\vec{Z} = \sum_{i=1}^n Z_i^2 \sim \chi_n^2$.

$$\sum \chi_i^2 = \sum \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma^2} \sim \chi_k^2$$

Note that

$$\begin{aligned}
\sum(X_i - \mu)^2 &= \sum(X_i - \bar{X} + \bar{X} - \mu)^2 \\
&= \sum((X_i - \bar{X}) + (\bar{X} - \mu))^2 \\
&= \sum((X_i - \bar{X})^2 + 2(X_i - \bar{X})(\bar{X} - \mu) + (\bar{X} - \mu)^2) \\
&= \sum(X_i - \bar{X})^2 + 2\sum(X_i\bar{X} - \bar{X}^2 - \mu X_i + \bar{X}\mu) + n(\bar{X} - \mu)^2 \\
&= \sum(X_i - \bar{X})^2 + 2(n\bar{X}^2 - n\bar{X}^2 - \mu n\bar{X} + n\bar{X}\mu) + n(\bar{X} - \mu)^2 \\
&= \sum(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2
\end{aligned}$$

Hence

$$\underbrace{\frac{\sum(X_i - \bar{X})^2}{\sigma^2}}_{\frac{(n-1)S^2}{\sigma^2}} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} \sim \chi_n^2$$

Furthermore,

$$\frac{n(\bar{X} - \mu)^2}{\sigma^2} = \Bigg(\underbrace{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}_{N(0,1)}\Bigg)^2 = Z^2 \sim \chi_1^2$$

If $X_1 \sim \chi^2_{k_1}$ and $X_2 \sim \chi^2_{k_2}$, then $X_1 + X_2 \sim \chi^2_{k_1+k_2}$.
It would be nice if $\frac{\sum(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$; then $\chi^2_{n-1} + \chi^2_1 = \chi^2_n$. Then χ^2_{n-1} needs to be independent of $\bar{X}$.

Cochran's Theorem: Let $Z_1, \dots, Z_n \overset{iid}{\sim} N(0,1)$. Let $Q_1, \dots, Q_k$ be scalar random variables with a quadratic form: $Q_j = \vec{Z}^t B_j \vec{Z}$ and $B_1, \dots, B_k$ are positive semidefinite matrices (matrix A is positive semidefinite if for all $\vec{v}$, $\vec{v}^T A \vec{v} \geq 0$) such that

1. $n = \sum$ rank (B_j)
2. Q_js are independent
3. $Q_j \sim \chi^2_{\text{rank}(B_j)}$

Note that

$$\begin{aligned} \vec{Z}^T\vec{Z} &= Q_1 + \dots + Q_k \\ &= \vec{Z}^t B_1 \vec{Z} + \dots + \vec{Z}^t B_k \vec{Z} \\ &= \vec{Z}^t (B_1 + \dots + B_k) \vec{Z} \\ I_n &= B_1 + \dots + B_k \end{aligned}$$

Note that

$$\begin{aligned} \sum Z_i^2 &= \sum((Z_i - \bar{Z}) + (\bar{Z}))^2 \\ &= \sum (Z_i - \bar{Z})^2 + 2\sum(Z_i - \bar{Z})\bar{Z} + \sum \bar{Z}^2 \\ &= \sum (Z_i - \bar{Z})^2 + 2(\sum Z_i\bar{Z} - \sum \bar{Z}^2) + n\bar{Z}^2 \\ &= \sum (Z_i - \bar{Z})^2 + 2(n\bar{Z}^2 - n\bar{Z}^2) + n\bar{Z}^2 \\ &= \underbrace{\sum (Z_i - \bar{Z})^2}_{Q_1} + \underbrace{n\bar{Z}^2}_{Q_2} \end{aligned}$$

Let $Q_2 = n\bar{Z}^2 = \vec{Z}^T B_2 \vec{Z}$. Let $J_n = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}$ ($n \times n$ matrix). Note that

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} = 2J$$

Then

$$\begin{aligned} Q_2 &= \vec{Z}^T(\frac{1}{n}J_n)\vec{Z} \\ &= \vec{Z}^T \begin{pmatrix} \frac{1}{n} & \dots & \frac{1}{n} \\ \vdots & \ddots & \vdots \\ \frac{1}{n} & \dots & \frac{1}{n} \end{pmatrix} \\ &= \vec{Z}^T \begin{pmatrix} \bar{Z} \\ \vdots \\ \bar{Z} \end{pmatrix} \\ &= z_1\bar{Z} + \dots + z_n\bar{Z} \\ &= \bar{Z}(\sum Z_i) \\ &= \bar{Z}n\bar{Z} \end{aligned}$$

What does the first term look like?

$$\begin{aligned} Q_1 &= \sum(Z_i - \bar{Z})^2 \\ &= \sum Z_i^2 - 2\sum Z_i\bar{Z} + \sum \bar{Z}^2 \\ &= \sum Z_i^2 - 2n\bar{Z}^2 + n\bar{Z}^2 \\ &= \sum Z_i^2 - n\bar{Z}^2 \\ &= \vec{Z}^T\vec{Z} - \frac{1}{n}\vec{Z}^T J_n \vec{Z} \\ &= \vec{Z}^T I_n \vec{Z} - \vec{Z}^T \frac{1}{n} J_n \vec{Z} \\ &= \vec{Z}^T(I_n - \frac{1}{n}J_n)\vec{Z} \end{aligned}$$

Therefore

$$\sum Z_i^2 = \overbrace{\vec{Z}^T(I_n - \frac{1}{n}J_n)\vec{Z}}^{Q_1} + \overbrace{\vec{Z}^T(\frac{1}{n}J_n)\vec{Z}}^{Q_2}$$

Then $B_1 = I_n - \frac{1}{n}J_n$ and $B_2 = \frac{1}{n}J_n$. Note also, $B_1B_2 = (I_n - \frac{1}{n}J_n)(\frac{1}{n}J_n) = \frac{1}{n}J_n - \frac{1}{n}J_nJ_n = 0$.
Theorem: If matrix A is both symmetric and independent, $\text{tr}(A) = \text{rank}(A)$.
$\frac{1}{n}J_n$ is clearly symmetric (each entry is $\frac{1}{n}$). Is it independent? $AA = A$.

$$\frac{1}{n}J_n\frac{1}{n}J_n = \frac{1}{n^2}J_nJ_n = \frac{1}{n^2}\begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}\begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix} = \frac{1}{n^2}\begin{pmatrix} n & \dots & n \\ \vdots & \ddots & \vdots \\ n & \dots & n \end{pmatrix} = \frac{1}{n^2}nJ_n = \frac{1}{n}J_n$$

Also, $\text{rank}(\frac{1}{n}IJ_n) = \text{tr}(\frac{1}{n}J_n) = \sum_{i=1}^n \frac{1}{n} = n \cdot \frac{1}{n} = 1$.

$$I_n - \frac{1}{n}J_n = \begin{pmatrix} 1-\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & \dots & -\frac{1}{n} \\ \vdots & \ddots & \ddots & \vdots \\ -\frac{1}{n} & \dots & \dots & 1-\frac{1}{n} \end{pmatrix}$$

This matrix is clearly symmetric. Is it independent?

$$\begin{aligned}(I_n - \frac{1}{n}J_n)(I_n - \frac{1}{n}J_n) &= I_nI_j - \frac{1}{n}J_nI_n - \frac{1}{n}I_nJ_n + \frac{1}{n^2}J_nJ_n \\ &= I_n - \frac{1}{n}J_n - \frac{1}{n}J_n + \frac{1}{n}J_n \\ &= I_n - \frac{1}{n}J_n\end{aligned}$$

Therefore $\text{rank}(I_n - \frac{1}{n}J_n) = \text{tr}(I_n - \frac{1}{n}J_n) = \sum_{i=1}^n 1 - \frac{1}{n} = n(1-\frac{1}{n}) = n-1$.
We still need to prove that B_1 and B_2 are positive semidefinite. Matrix A is positive semidefinite if for all $\vec{v} \neq \vec{0}$, $\vec{v}^TA\vec{v} \geq 0$. Well, $\vec{Z}^TB_2\vec{Z} = n\bar{Z}^2 \geq 0$; and $\vec{Z}^TB_1\vec{Z} = \sum(Z_i - \bar{Z})^2 \geq 0$. Therefore it is and we can apply Cochran's theorem.

1. $\sum(Z_i - \bar{Z})^2 \sim \chi^2_{n-1}$ and $n\bar{Z}^2 \sim \chi^2_1$
2. $\sum(Z_i - \bar{Z})^2$ is independent of $n\bar{Z}^2$

Therefore $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.

$$\begin{aligned}\sum Z_i^2 &= \underbrace{\sum\left(\frac{X_i - \mu}{\sigma}\right)^2}_{\chi^2_k} \\ &= \frac{\sum(X_i - \bar{X})^2}{\sigma^2} + \frac{n(\bar{X} - \mu)^2}{\sigma^2} \\ &= \left(\frac{\vec{X} - \vec{\mu}}{\sigma}\right)^T\left(\frac{1}{n}J_n\right)\left(\frac{\vec{X} - \mu}{\sigma}\right) \\ &= \left(\frac{\vec{X} - \vec{\mu}}{\sigma}\right)^T(I_n - \frac{1}{n}J_n)\left(\frac{\vec{X} - \vec{\mu}}{\sigma}\right)\end{aligned}$$

Using Cochran's theorem,

1. $\frac{\sum(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$ and $\frac{n(\bar{X}-\mu)^2}{\sigma^2} \sim \chi^2_1$
2. $\frac{\sum(X_i - \bar{X})^2}{\sigma^2}$ and $\frac{n(\bar{X}-\mu)^2}{\sigma^2}$ are independent

Since $\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X_i - \bar{X})^2}{\sigma^2}$, then $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$ and so $S^2 \sim \frac{\sigma^2}{n-1}\chi^2_{n-1} = \text{Gamma}(\frac{n-1}{2}, \frac{n-1}{2\sigma^2})$. Thus $\frac{\sqrt{n-1}}{\sigma}S \sim \chi_{n-1}$. Also, $\frac{(n-1)s^2}{\sigma^2}$ is independent of $n\left(\frac{\bar{X}-\mu}{\sigma^2}\right)^2$.
Since $n-1$, n, μ, σ^2 are constants, S^2 and $\bar{X}$ are independent.

Here is where this is all important, allowing us to use the z-test. Let $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Consider $\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$. Then $\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim$ close to $N(0,1)$ since $S \approx \sigma$. This allows the z-test to work. Furthermore,

$$\begin{aligned}
\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} &= \frac{\bar{X}-\mu}{\frac{1}{\sqrt{n}}\sqrt{S^2}} \\
&= \frac{\bar{X}-\mu}{\frac{1}{\sqrt{n}}\sqrt{\frac{\sigma^2}{n-1}\frac{n-1}{\sigma^2}S^2}} \\
&= \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{\frac{n-1}{\sigma^2}S^2}{n-1}}} \\
&= \frac{\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\frac{n-1}{\sigma^2}S^2}{n-1}}} \\
&\sim T_{n-1}
\end{aligned}$$

The numerator of this is $N(0,1)$ and the denominator is χ^2_{n-1} and they are both independent of each other. This gives rise to the t-test.

Let $\vec{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \in \mathbb{R}^{n\times 1}$. Then $\text{E}[\vec{X}] = \vec{\mu}$ and $\text{E}[\vec{X}^T] = \vec{\mu}^T$. Let the following be a matrix of random variables: $X = \begin{pmatrix} X_{11} & \dots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{nm} \end{pmatrix} \in \mathbb{R}^{n\times m}$. Furthermore, $\text{E}[X] = \begin{pmatrix} \mu_{11} & \dots & \mu_{1m} \\ \vdots & \ddots & \vdots \\ \mu_{n1} & \ddots & \mu_{nm} \end{pmatrix} = \mu \in \mathbb{R}^{n\times m}$. Let's define the covariance.

$$\begin{aligned}
\Sigma = \text{Cov}[\vec{X}] &= \text{E}[(\vec{X}-\vec{\mu})(\vec{X}-\vec{\mu})^T] \\
&= \text{E}[\begin{pmatrix} X_1-\mu_1 \\ \vdots \\ X_n-\mu_n \end{pmatrix} \begin{pmatrix} X_1-\mu_1 & \dots & X_n-\mu_n \end{pmatrix}] \\
&= \begin{pmatrix} \text{E}[(X_1-\mu_1)^2] & \text{E}[(X_1-\mu_1)(X_2-\mu_2)] & \dots & \text{E}[(X_1-\mu_1)(X_n-\mu_n)] \\ \text{E}[(X_2-\mu_2)(X_1-\mu_1)] & \text{E}[(X_2-\mu_2)^2] & \dots & \dots \\ \vdots & \ddots & \ddots & \vdots \\ \text{E}[(X_n-\mu_n)(X_1-\mu_1)] & \text{E}[(X_n-\mu_2)(X_2-\mu_2)] & \dots & \text{E}[(X_n-\mu_n)^2] \end{pmatrix} \\
&= \begin{pmatrix} \text{Var}[X_1] & \text{Cov}[X_1,X_2] & \dots \\ \text{Cov}[X_1,X_2] & \text{Var}[X_2] & \dots \\ \vdots & \ddots & \vdots \\ \dots & \dots & \text{Var}[X_n] \end{pmatrix}
\end{aligned}$$

Furthermore,

$$\begin{aligned}\Sigma &= \text{Cov}[\vec{X}] \\ &= \text{E}[(X-\mu)(X^T-\mu^T)] \\ &= \text{E}[XX^T - \mu X^T - X\mu^T + \mu\mu^T] \text{ each of which is } (n\times 1)(1\times n)\end{aligned}$$

Let $X \in \mathbb{R}^{n\times m}$ and $A \in \mathbb{R}^{p\times n}$; then $AX \in \mathbb{R}^{n\times m}$.

$$\begin{aligned}\text{E}[AX] &= \text{E}[\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{p1} & \dots & a_{pn}\end{pmatrix}\begin{pmatrix} X_{11} & \dots & X_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1} & \dots & X_{nm}\end{pmatrix}] \\ &= \begin{pmatrix} a_{1\dots}\mu_{\dots 1} & a_{1\dots}\mu_{\dots 2} & \dots & a_{1\dots}\mu_{\dots m} \\ a_{2\dots}\mu_{\dots 1} & a_{2\dots}\mu_{\dots 2} & \dots & a_{2\dots}\mu_{\dots m} \\ \vdots & \ddots & \ddots & \vdots \\ a_{p\dots}\mu_{\dots 1} & \dots & \dots & a_{p\dots}\mu_{\dots m}\end{pmatrix} \underbrace{\begin{pmatrix} a_{1\dots} & \\ a_{2\dots} & \\ & \vdots \\ a_{p\dots} & \end{pmatrix}}_{A} \underbrace{\begin{pmatrix}\mu_{\dots 1} & \mu_{\dots 2} & \dots & \mu_{\dots m}\end{pmatrix}}_{\text{E}[X]} \\ &= A\text{E}[X]\end{aligned}$$

Let $X \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n\times m}$.

$$\begin{aligned}\text{E}[X+B] &= \text{E}[\begin{pmatrix} X_{11}+B_{11} & \dots & X_{1m}+B_{1m} \\ \vdots & \ddots & \vdots \\ X_{n1}+B_{n1} & \dots & X_{nn}+B_{nn}\end{pmatrix}] \\ &= \begin{pmatrix} \mu_{11}+B_{11} & \dots & \mu_{1m}+B_{1m} \\ \vdots & \ddots & \vdots \\ \mu_{n1}+B_{n1} & \dots & \mu_{nm}+B_{nm}\end{pmatrix} \\ &= \mu + B = \text{E}[X] + B\end{aligned}$$

in the same manner, $\text{E}[AX+B] = A\text{E}[X]+B$ if dimensions conform, otherwise not defined. Similarly, $\text{E}[B+XA] = B + \text{E}[X]A$ if dimensions conform.
Hence

$$\begin{aligned}\Sigma &= \text{Cov}[\vec{X}] \\ &= \text{E}[XX^T] + \text{E}[-\mu X^T] + \text{E}[-X\mu^T] + \text{E}[\mu\mu^T] \\ &= \text{E}[XX^T] - \mu\underbrace{\text{E}[X^T]}_{\mu^T} - \underbrace{\text{E}[X]}_{\mu}\mu^T + \mu\mu^T \\ &= \text{E}[XX^T] - \mu\mu^T \\ &= \text{E}[XX^T] - \text{E}[X]\text{E}[X^T]\end{aligned}$$

Consider the following:

$$\begin{aligned}\text{Cov}[A^TX] &= \text{E}[(A^TX)(A^TX)] - \text{E}[A^TX]\text{E}[(A^TX)^T] \\ &= \text{E}[A^TXX^TA] - \text{E}[A^TX]\text{E}[X^TA] \\ &= A^T\text{E}[XX^T]A - A^T\text{E}[X]\text{E}[X^T]A \\ &= A^T(\text{E}[XX^T] - \mu\mu^T)A \\ &= A^T\text{Cov}[X]A \\ &= A^T\Sigma A\end{aligned}$$

Note that $\text{E}[A^TX] = A^T\mu$ and $\text{E}[(A^TX)^T] = (A^T\mu)^T = \mu^TA^{T^T} = \mu^TA$.
Hence $\text{Cov}[AX] = A\Sigma A^T$.

Let $\vec{Z} = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$ where $Z_1, \dots, Z_n \overset{iid}{\sim} N(0,1)$. Then $\vec{Z} \sim N_n(0, I_n)$ (the multivariate normal distribution of dimension n). Furthermore, $\text{E}[\vec{Z}] = \vec{0}$, how about $\text{Cov}[\vec{Z}]$? Well, all $\text{Cov}[Z_i, Z_j] = 0$ if $i \neq j$ and $\text{Var}[Z_i] = 1$ for all i; hence, $\text{Cov}[\vec{Z}] = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} = I_n$, the identity matrix of size n. What's its PDF?

$$f_{\vec{Z}}(\vec{z}) = f_{\vec{Z}}(z_1, \dots, z_n) = f_{Z_1}(z_1)\cdot\dots\cdot f_{Z_n}(z_n) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}}e^{-\frac{Z_i^2}{2}} = \frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}\sum_{i=1}^n Z_i^2} = \frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}\vec{Z}^T\vec{Z}}$$

This all happens because each Z_I is independent from each other.
Let $\vec{X} = \vec{Z} + \vec{c}$, where $\vec{c} \in \mathbb{R}^n$, a constant. Then $\text{E}[\vec{X}] = \text{E}[Z] + \vec{c} = \vec{0} + \vec{c} = \vec{c}$. In addition, $\text{Var}[\vec{X}] = I_n$; hence, $\vec{X} \sim N_n(\vec{c}, I_n)$.

$$f_{\vec{X}}(\vec{x}) = f_{X_1}(x)\cdot\dots\cdot f_{X_n}(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(X_i - c_i)^2} = \frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}\sum(X_i - c_i)^2} = \frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}(\vec{X}-\vec{c})^T(\vec{X}-\vec{c})}$$

Let $\vec{X} = A\vec{Z}$ where $\vec{X} \in \mathbb{R}^{m\times 1}$, $A \in \mathbb{R}^{m \times n}$ and $\vec{Z} \in \mathbb{R}^{n \times 1}$. Then

$$\vec{X} = A\vec{Z} = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix} = \begin{pmatrix} a_{11}Z_1 + a_{12}Z_2 + \dots + a_{1n}Z_n \\ a_{21}Z_1 + a_{22}Z_22 + \dots + a_{2n}Z_n \\ \vdots \\ a_{m1}Z_1 + a_{m2}Z_2 + \dots + a_{mn}Z_n \end{pmatrix}$$

Each of these row is $\sim N(0, \sum a_{xi}^2)$. Therefore, $\text{E}[\vec{X}] = A\text{E}[\vec{Z}] = A\vec{0}_n = \vec{0}_m \in \mathbb{R}^m$ and $\Sigma = \text{Cov}[\vec{X}] = A\text{Cov}[\vec{Z}]A^T = AI_nA^T = AA^t \in \mathbb{R}^{m\times n}$.
Is Σ symmetric?

$$\Sigma = \Sigma^T = (AA^T)^T = A^{T^T}A^T = AA^T$$

Is $\text{Cov}[X_1, X_2] = 0$? No, they are dependent since they contain the same Z_is. What's $f_{\vec{X}}(\vec{x})$? Note that $\vec{X} = A\vec{Z} = g(\vec{Z})$, a multivariable change of variable problem. To do this, we must have $\dim(\vec{X}) = \dim(\vec{Z})$, so $m = n$. There exists h such that $\vec{X} = h(\vec{Z})$. What is it?

$$\begin{aligned} \vec{X} &= A\vec{Z} \\ \vec{Z} &= A^{-1}\vec{X} = h(\vec{Z}) \end{aligned}$$

Therefore A must be an invertible matrix and $h(\vec{x}) = A^{-1}\vec{X}$.

$$f_{\vec{X}}(\vec{x}) = f_{\vec{Z}}(A^{-1}\vec{X})|J_h(\vec{x})|$$

Let $B = A^{-1}$, then $h(\vec{X}) = \begin{pmatrix} h_1(\vec{x}) \\ h_2(\vec{x}) \\ \vdots \\ h_n(\vec{x}) \end{pmatrix} = B\vec{X}$ and

$$J_n = \begin{pmatrix} \frac{\partial}{\partial x_1} h_1(\vec{x}) & \dots & \frac{\partial}{\partial x_n} h_1(\vec{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_1} h_n(\vec{x}) & \dots & \frac{\partial}{\partial x_n} h_n(\vec{x}) \end{pmatrix}$$

Note that

$$B = \begin{pmatrix} \vec{b}_{1\cdot} \\ \vdots \\ \vec{b}_{n\cdot} \end{pmatrix} = \begin{pmatrix} \vec{b}_{\cdot 1} & \dots & \vec{b}_{\cdot n} \end{pmatrix} = \begin{pmatrix} b_{11} & \dots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nn} \end{pmatrix}$$

Then $h_1(\vec{x}) = \vec{b}_{1\cdot}.\vec{x} = b_{11}x_1 + \dots + b_{1n}x_n$. Furthermore,

$$\begin{aligned} \frac{\partial}{\partial x_1}[h_1(\vec{x})] &= b_{11} \\ \frac{\partial}{\partial x_2}[h_1(\vec{x})] &= b_{12} \\ &\vdots \\ \frac{\partial}{\partial x_n}[h_1(\vec{x})] &= b_{1n} \end{aligned}$$

These are the elements of the first row of J_n. Following this pattern, we see that

$$J_n = \det \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \dots & b_{nn} \end{pmatrix} = \det B = \det A^{-1}$$

Note that $\frac{\partial}{\partial \vec{x}}[C\vec{x}] = C$.
Therefore

$$f_{\vec{X}}(\vec{x}) = f_{\vec{Z}}(A^{-1}\vec{X})|\det A^{-1}|$$

Recall that $\det A^{-1} = \frac{1}{\det A}$ because if $AA^{-1} = I$, then $\det AA^{-1} = \det I$ and so $\det A \det A^{-1} = 1$. Hence

$$f_{\vec{X}}(\vec{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\det A|} e^{-\frac{1}{2}(A^{-1}\vec{X})^T(A^{-1}\vec{X})} = \frac{1}{(2\pi)^{\frac{n}{2}}|\det A|} e^{-\frac{1}{2}\vec{X}^T(A^{-1})^T A^{-1}\vec{X}}$$

Recall that $\Sigma = AA^T$ and so $\Sigma^{-1} = (AA^T)^{-1} = (A^T)^{-1}A^{-1}$. Does $(A^T)^{-1} = (A^{-1})^T$? Yes.

$$\begin{aligned} AA^{-1} &= I \\ (AA^{-1})^T &= I^T = I \\ (A^{-1})^T A^T &= I \\ (A^T)^{-1} A^T &= I \\ (A^T)^{-1} &= (A^{-1})^T \end{aligned}$$

Hence $\Sigma^{-1} = (A^{-1})^T A^{-1}$, which is symmetric.

$$f_{\vec{X}}(\vec{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\det A|} e^{-\frac{1}{2}\vec{X}^T\Sigma^{-1}\vec{X}}$$

Note that: $|\det \Sigma| = |\det AA^T| = |\det A \det A^T| = |\det A^2|$ and so $\sqrt{|\det \Sigma|} = |\det A|$. This says that

$$f_{\vec{X}}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n|\det \Sigma|}} e^{-\frac{1}{2}\vec{X}^t\Sigma^{-1}\vec{X}} = N_n(\vec{0}, \Sigma)$$

If $\vec{X} = A\vec{Z} + \vec{\mu}$, then $\vec{X} \sim N_n(\vec{\mu}, \Sigma) = \frac{1}{\sqrt{(2\pi)^n|\det \Sigma|}} e^{-\frac{1}{2}(\vec{X}-\vec{\mu})^T\Sigma^{-1}(\vec{X}-\vec{\mu})}$.
If $\vec{X} \sim N_n(\vec{\mu}, \Sigma)$, $B \in \mathbb{R}^{m\times n}$, then $B\vec{X} \sim ?$

Recall that $\phi_X(t) = \text{E}[itX]$. If X is a vector, $\phi_{\vec{X}}(\vec{t}) = \text{E}[e^{i\vec{t}^t\vec{X}}]$.
Properties:

- $\phi_{\vec{X}_1+\vec{X}_2}(\vec{t}) = \text{E}[e^{i\vec{t}^T(\vec{X}_1+\vec{X}_2)}] = \text{E}[e^{i\vec{t}^T\vec{X}_1+i\vec{t}^T\vec{X}_2}] = \text{E}[e^{i\vec{t}^T\vec{X}_1}e^{i\vec{t}^T\vec{X}_2}] = \phi_{\vec{X}_1}(\vec{t}) \cdot \phi_{\vec{X}_2}(\vec{t})$
- $\phi_{A\vec{X}+\vec{c}}(\vec{t}) = \text{E}[e^{i\vec{t}^T(A\vec{X}+\vec{c})}] = \text{E}[e^{i\vec{t}^TA\vec{X}+i\vec{t}^T\vec{c}}] = e^{i\vec{t}^T\vec{c}}\text{E}[e^{i\vec{t}^TA\vec{X}}] = e^{i\vec{t}^T\vec{c}}\phi_{\vec{X}}(A^T\vec{t})$

What's the characteristic function for MVN $\vec{Z} \sim N_n(\vec{0}, I_n)$?

$$\begin{aligned}
\mathrm{E}[e^{i\vec{t}^T\vec{Z}}] &= \underbrace{\int_{\mathbb{R}} \dots \int_{\mathbb{R}}}_{\mathbb{R}^n} e^{i\vec{t}^T\vec{Z}} f_{\vec{Z}}(\vec{z})\,d\vec{z} \\
&= \int \dots \int e^{i(t_1z_1+t_2z_2+\dots+t_nz_n)} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}(z_1^2+\dots+z_n^2)}\,dz_1\dots dz_n \\
&= \prod_{i=1}^n \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{it_iz_i-\frac{1}{2}z_i^2}\,dz_i \\
-\frac{1}{2}z_i^2 + it_iz_i &= -\frac{1}{2}(z_i^2 - 2it_iz_i) = -\frac{1}{2}((z_i-it_i)^2 - i^2t_i^2) \\
&= -\frac{1}{2}((z_i-it_i)^2 + t_i^2) = -\frac{1}{2}(z_i-it_i)^2 - \frac{t_i^2}{2} \\
\mathrm{E}[e^{i\vec{t}^T\vec{Z}}] &= \prod_{i=1}^n e^{-\frac{t_i^2}{2}} \underbrace{\underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z_i-it_i)^2}\,dz_i}_{N(it_i,1)}}_{1} \\
&= e^{-\frac{1}{2}\sum t_i^2} = e^{-\frac{1}{2}\vec{t}^T\vec{t}} = \phi_{\vec{Z}}(\vec{t})
\end{aligned}$$

If $\vec{X} = A\vec{Z} + \vec{\mu}$ (where $A \in \mathbb{R}^{n\times n}$), then $\vec{X} \sim N_n(\vec{\mu}, \Sigma)$ such that $\Sigma = AA^T$ and so

$$\phi_{\vec{X}}(\vec{t}) = e^{i\vec{t}^T\vec{\mu}}\phi_{\vec{Z}}(A^T\vec{t}) = e^{i\vec{t}^T\vec{\mu}}e^{-\frac{1}{2}\vec{t}^TAA^t\vec{t}} = e^{i\vec{t}^T\vec{\mu}-\frac{1}{2}\vec{t}^T\Sigma\vec{t}}$$

Let $\vec{Y} = B\vec{X}$ such that $B \in \mathbb{R}^{m\times n}$ where $m \neq n$. Then

$$\phi_{\vec{Y}}(\vec{t}) = \phi_{\vec{X}}(B^T\vec{t}) = e^{i\vec{t}^TB\vec{\mu}-\frac{1}{2}\vec{t}^TB\Sigma B^T\vec{t}} \to Y \sim N_m(B\vec{\mu}, B\Sigma B^T)$$

Let $\vec{Z} \sim N_n(\vec{0}, I_n)$. Let $\vec{X} = B\vec{Z} + \vec{c}$ where $B \in \mathbb{R}^{m\times n}$ $(m \neq n)$ and $\vec{c} \in \mathbb{R}^m$. Then

$$\phi_{\vec{X}}(\vec{t}) = e^{i\vec{t}^T\vec{c}}\phi_{\vec{Z}}(B^T\vec{t}) = e^{i\vec{t}^T\vec{c}-\frac{1}{2}\vec{t}^TBB^T\vec{t}} \to X \sim N_m(\vec{c}, \Sigma)$$

where $\Sigma = BB^T$. Note that Σ must be full rank.

Given $\vec{X}$, how do we standardize back to $\vec{Z}$?
Let $\vec{X} = A\vec{Z} + \vec{\mu}$. Then $\vec{X} - \vec{\mu} = A\vec{Z}$ and so $A^{-1}(\vec{X} - \vec{\mu}) = \vec{Z}$. This can only happen if A is invertible Then $\vec{Z} = A^{-1}\vec{X} - A^{-1}\vec{\mu}$. Furthermore, since $\vec{Z}^T\vec{Z} = \sum_{i=1}^n Z_i^2 \sim \chi_n^2$,

$$\begin{gathered}
(A^{-1}(\vec{X} - \vec{\mu}))^T(A^{-1}(\vec{X} - \vec{\mu}) = \vec{Z}^T\vec{Z} \sim \chi_n^2 \\
(\vec{X}^T - \vec{\mu}^T)(A^{-1})^TA^{-1}(\vec{X} - \vec{\mu}) = \vec{Z}^T\vec{Z} \sim \chi_n^2 \\
(\vec{X}^t - \vec{\mu}^T)\Sigma^{-1}(\vec{X} - \vec{\mu}) \sim \chi_n^2
\end{gathered}$$

Let $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Then $\vec{X} \sim N_n(\mu\vec{1}, \sigma^2 I_n)$. This says that $\Sigma = \sigma^2 I = AA^T$ and so $A = \sigma I$.

Let $\vec{X} = \sigma I\vec{Z} + \vec{\mu} = \sigma\vec{Z} + \vec{\mu}$, then

$$(\vec{X} - \vec{\mu})^T \frac{1}{\sigma^2}(\vec{X} - \vec{\mu}) \sim \chi_n^2$$
$$\frac{1}{\sigma^2}(\vec{X} - \vec{\mu})^T(\vec{X} - \vec{\mu}) \sim \chi_n^2$$
$$\frac{1}{\sigma^2}\sum(X_i - \mu)^2 \sim \chi_n^2$$

Let X be a nonnegative random variable with finite expectation μ Consider $a > 0$ a constant. Consider the inequality

$$a\mathbb{1}_{x\geq a} \leq x$$

Is this intuitive? Yes because
If $x \geq a$, $a(1) \leq x \to x \geq a$
If $x < a$, $a(0) \leq x \to x \geq 0$.
Note that $\text{E}[a\mathbb{1}_{x\geq a}] \leq \mu$ and $a\text{E}[\mathbb{1}_{X\geq a}] \leq \mu$. This is $a\mathbb{P}(X \geq a) \leq \mu$ and therefore

$$\mathbb{P}(X \geq a) \leq \frac{\mu}{a}$$

This is Markov's inequality.
Corollaries:
Let $a = a'\mu$

$$\mathbb{P}(X \geq a'\mu) \leq \frac{1}{a'}$$

Let h be a monotonically increasing function: $h(a)\mathbb{1}_{h(X)>h(a)} \leq h(X)$. Then

$$\mathbb{P}(h(X) \geq h(a)) \leq \frac{\text{E}[h(X)]}{h(a)}$$
$$\mathbb{P}(X \geq a) \leq \frac{\text{E}[h(X)]}{h(a)}$$

Let $h(X) = X^p$ such that $p > 1$.

$$\mathbb{P}(X \geq a) \leq \frac{\text{E}[X^p]}{a^p}$$

Recall that Quantile$[X,p] = F_X^{-1}(p)$ (if F is continuous). Then

$$\mathbb{P}(X \geq a) \leq \frac{\mu}{a}$$
$$1 - \mathbb{P}(X \leq a) \leq \frac{\mu}{a}$$
$$1 - F(a) \leq \frac{\mu}{a}$$
$$\text{Let } a = F_X^{-1}(p)$$
$$1 - F(F_X^{-1}(p)) \leq \frac{\mu}{F_X^{-1}(p)}$$
$$1 - p \leq \frac{\mu}{F_X^{-1}(p)}$$
$$\text{Quantile}[X,p] \leq \frac{\mu}{1-p}$$

Note that $\text{med}[X] \leq 2\mu$.

Consider any random variable X. $|X|$ is nonnegative.

$$\mathbb{P}(|X| \geq a) \leq \frac{\text{E}[|X|]}{a}$$

Let X be any random variable with finite μ and finite σ^2. Let $Y = (X-\mu)^2$. Note that Y is nonnegative.

$$\begin{aligned}
\mathbb{P}(Y \geq a^2) &\leq \frac{\text{E}[Y]}{a^2} \\
&= \frac{\text{E}[(X-\mu)^2]}{a^2} \\
&= \frac{\sigma^2}{a^2} \\
\mathbb{P}((X-\mu)^2 \geq a^2) &\leq \frac{\sigma^2}{a^2} \\
\mathbb{P}(|X-\mu| \geq a) &\leq \frac{\sigma^2}{a^2}
\end{aligned}$$

This is Chebyshev's inequality.

Let X be any random variable. Let $Y = e^{tX}$ (Y is nonnegative.)

$$\begin{aligned}
\mathbb{P}(Y \geq c) &\leq \frac{\text{E}[Y]}{c} \\
\mathbb{P}(e^{tX} \geq c) &\leq \frac{\text{E}[e^{tX}]}{c} \\
&\text{Let } c = e^{ta} \\
\mathbb{P}(e^{tX} \geq e^{ta}) &\leq \frac{\text{E}[e^{tX}]}{e^{ta}} \\
&= \frac{M_X(t)}{e^{ta}}
\end{aligned}$$

Note that $M_X(t) = \text{E}[e^{tX}]$ is a moment generating function.
If $t > 0$, $\mathbb{P}(X \geq a) \leq e^{-ta}M_x(t)$. If $t < 0$, $\mathbb{P}(X \leq a) \leq e^{-ta}M_X(t)$.
Therefore,

$$\mathbb{P}(X \geq a) \leq t \overset{\min}{>} 0 \, \{e^{-ta}M_X(t)\}$$

$$\mathbb{P}(X \leq a) \leq t \overset{\min}{<} 0 \, \{e^{-ta}M_X(t)\}$$

This is Chernoff's Inequality.

Let $X \sim \text{Binom}\Big(n, \frac{1}{4}\Big)$. Then $\mu = \frac{1}{4}n$ and $\sigma^2 = \frac{3}{16}n$. What's $\mathbb{P}(X \geq \frac{3}{4}n)$? If n is large, $X \approx N\Big(\frac{1}{4}n, \Big(\sqrt{\frac{3}{16}n}\Big)^2\Big)$.Then

$$\begin{aligned}\mathbb{P}(X \geq \frac{3}{4}n) &= \mathbb{P}(\frac{X - \frac{1}{4}n}{\sqrt{\frac{3}{16}n}} > \frac{\frac{3}{4}n - \frac{1}{4}n}{\sqrt{\frac{3}{16}n}}) \\ &= \mathbb{P}(X > \frac{2}{\sqrt{3}}\sqrt{n}) \\ &= 0 \text{ if } n \text{ large}\end{aligned}$$

Using Markov's:

$$\mathbb{P}(X \geq \frac{3}{4}n) \leq \frac{\frac{1}{4}n}{\frac{3}{4}n} = \frac{1}{3}$$

Using Chebychev's:

$$\begin{aligned}\mathbb{P}(X \geq \frac{3}{4}n) &= \mathbb{P}(X - \frac{1}{4}n \geq \frac{3}{4}n - \frac{1}{4}n) \\ &\leq \mathbb{P}(X - \frac{1}{4}n \geq \frac{1}{2}n) + \mathbb{P}(\frac{1}{4}n - X \geq \frac{1}{2}n) \\ &= \mathbb{P}(X - \frac{1}{4}n \geq \frac{1}{2}n \text{ or } \frac{1}{4}n - X \geq \frac{1}{2}n) \\ &= \mathbb{P}(|X - \frac{1}{4}n| \geq \frac{1}{2}n) \\ &\geq \frac{\frac{3}{16}n}{\frac{1}{4}n^2} \\ &= \frac{3}{4}n\end{aligned}$$

Let $X \sim \text{Binom}(n, p)$.

$$\begin{aligned}M_X(t) &= \text{E}[e^{tX}] \\ &= \sum_{x=0}^{n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} \\ &= \sum_{x=0}^{n}\binom{n}{x}(e^tp)^x(1-p)^{n-x} \\ &= (1 - p + pe^t)^n\end{aligned}$$

Therefore,

$$X \sim \text{Binom}\Big(n, \frac{1}{4}\Big) \to M_X(t) = \Big(\frac{3}{4} + \frac{1}{4}e^t\Big)^n$$

Using Chernoff's:

$$\begin{aligned}
\mathbb{P}(X \geq \frac{3}{4}n) \leq & t \overset{\min}{>} 0 \left\{ e^{-t\left(\frac{3}{4}n\right)} \left(\frac{3}{4} = \frac{1}{4}e^t\right)^t \right\} \\
= & t \overset{\min}{>} 0 \left\{ \left(\frac{3}{4}e^{-\frac{3}{4}t} + \frac{1}{4}e^{\frac{1}{4}t}\right)^n \right\}
\end{aligned}$$

To minimize, take the derivative of above and set it equal to 0

$$\begin{aligned}
e^{\frac{1}{4}t} &= 9e^{-\frac{3}{4}t} \\
\frac{1}{4}t &= \ln(9) - \frac{3}{4}t \\
t_{\min} &= \ln(9) \\
\mathbb{P}(X \geq \frac{3}{4}n) &= \left(\frac{3}{4}e^{-\frac{3}{4}\ln(9)} + \frac{1}{4}e^{\frac{1}{4}\ln(9)}\right)^n \\
&= \left(\frac{3}{4}9^{-\frac{3}{4}} + \frac{1}{4}9^{\frac{1}{4}}\right)^n \\
&= \frac{\sqrt[4]{9}}{4^n}\left(\frac{3}{9^3} + 1\right)^n \\
&= \sqrt[4]{9}\left(\frac{1.004}{4}\right)^n \\
&\to 0 \text{ exponentially fast}
\end{aligned}$$

Consider any two random variables X and Y with finite μ's and σ^2's. Let $W = (X - cY)^2$ such that $c \in \mathbb{R}$. Note that W is nonnegative.

$$\begin{aligned}
\mathrm{E}[W] &\geq 0 \\
\mathrm{E}[(X - cY)^2] &\geq 0 \\
\mathrm{E}[X^2 - 2cXY + c^2Y^2] &\geq 0 \\
\mathrm{E}[X^2] - 2c\mathrm{E}[XY] + c^2\mathrm{E}[Y^2] &\geq 0 \\
\text{Let } c &= \frac{\mathrm{E}[XY]}{\mathrm{E}[Y^2]} \\
\mathrm{E}[X^2] - 2\frac{\mathrm{E}[XY]}{\mathrm{E}[Y^2]}\mathrm{E}[XY] + \frac{\mathrm{E}[XY]}{\mathrm{E}[Y^2]}\mathrm{E}[Y^2] &\geq 0 \\
\mathrm{E}[X^2]\mathrm{E}[Y^2] - 2\mathrm{E}[XY]^2 + \mathrm{E}[XY]^2 &\geq 0 \\
\mathrm{E}[XY]^2 &\leq \mathrm{E}[X^2]\mathrm{E}[Y^2] \\
|\mathrm{E}[XY]| &\leq \sqrt{\mathrm{E}[X^2]\mathrm{E}[Y^2]}
\end{aligned}$$

This is Cauchy-Schwartz Inequality. It is equal when $X = cY$.

What is correlation, Let SE be standard error. Then

$$\begin{aligned}
\text{Corr}[X,Y] &= \text{Corr}[cY,Y] \\
&= \frac{\text{Cov}[cY,Y]}{\text{SE}[cY]\text{SE}[Y]} \\
&= \frac{c\text{Cov}[Y,Y]}{|c|\text{SE}[Y]^2} \\
&= \frac{c\text{Var}[Y]}{|c|\text{Var}[Y]} \\
&= \frac{c}{|c|} = \begin{cases} 1 & \text{if } c > 0 \\ -1 & \text{if } c < 0 \end{cases}
\end{aligned}$$

Can we prove that $\text{Corr}[X,Y] \in [-1,1]$ for all random variables X, Y?
Let $Z_X = \frac{X-\mu_X}{\sigma_X}$ and $Z_Y = \frac{Y-\mu_Y}{\sigma_Y}$. Then $\text{E}[Z_X] = \text{E}[Z_Y] = 0$ and $\text{SE}[Z_X] = \text{SE}[Z_y] = 1$ so $\text{E}[Z_X^2] = \text{E}[Z_Y^2] = 1$.
Note that

$$|\text{E}[Z_X Z_Y|] \leq \sqrt{\text{E}[Z_X^2]\text{E}[Z_Y^2]} = 1$$

Therefore $\text{E}[Z_X Z_Y] \in [-1,1]$.

$$\text{Corr}[Z_X,Z_Y] = \frac{\text{Cov}[Z_X,Z_Y]}{\text{SE}[Z_X]\text{SE}[Z_Y]} = \frac{\text{E}[Z_X Z_Y] - \text{E}[Z_X]\text{E}[Z_Y]}{\text{SE}[Z_X]\text{SE}[Z_Y]} = \text{E}[Z_X Z_Y] \in [-1,1]$$

Henceforth

$$\begin{aligned}
\text{Corr}[X,Y] &= \frac{\text{E}[XY] - \mu_X\mu_Y}{\sigma_X\sigma_Y} \\
&= \frac{\text{E}[(\sigma_X Z_X + \mu_X)(\sigma_Y Z_Y + \mu_Y)] - \mu_X\mu_Y}{\sigma_X\sigma_Y} \\
&= \frac{\sigma_X\sigma_Y\text{E}[Z_X Z_Y]}{\sigma_X\sigma_Y} \\
&= \text{E}[Z_X Z_Y]
\end{aligned}$$

Therefore

$$\text{Corr}[X,Y] \in [-1,1]$$

A function g is convex on an interval $I \in \mathbb{R}$ if for all $x_1,\dots,x_n \in I$ and for all $w_1,\dots,w_n$ such that for all $w_i > 0$ and $\sum w_i = 1$ (n weights),

$$g(w_1x_1 + \cdots + w_nx_n) \leq w_1g(x_1) + \cdots + w_ng(x_n)$$

OR

$$g(\sum w_ix_i) \leq \sum w_ig(x_i)$$

Note that $\sum w_ix_i \in I$.
Theorem: If g is twice differentiable, then g is convex if $g''(x) \geq 0$ for all $x \in I$.

Imagine a discrete random variable with $\text{Supp}[X] = \Big\{x_1, \dots, x_n\Big\}$ and pmf $p(x_i) = w_i$. Then $\sum w_i x_i = \sum_{x \in \text{Supp}[X]} x p(x) = \text{E}[X]$. Then $\sum w_i g(x_i) = \sum_{x \in \text{Supp}[X]} g(x)p(x) = \text{E}[g(x)]$. Then

$$g(\text{E}[X]) = \text{E}[g(X)]$$

This is Jensen's Inequality.

$$g(\text{E}[X]) \leq \text{E}[g(X)]$$

if g is convex.

If $g(X)$ is linear, then it is both convex and concave

$$g(\text{E}[X]) = \text{E}[g(X)]$$

Therefore

$$a\text{E}[X] + b = \text{E}[aX + b]$$

For example, $g(x) = x^2$ is convex.

$$\text{E}[X]^2 \leq \text{E}[X^2] \to \mu^2 \leq \sigma^2 + \mu^2 \to \sigma^2 \geq 0$$

Let $g(x) = -\ln(x)$ where $x > 0$. Is it convex? $g'(x) = -\frac{1}{x}$. $g''(x) = \frac{1}{x^2} \geq 0 \forall x > 0$. Therefore it is convex.
Let $X \sim \begin{cases} a^p & \text{with probability } \frac{1}{p} \\ b^q & \text{with probability } \frac{1}{q} \end{cases}$. Note that $\frac{1}{p} + \frac{1}{q} = 1$, $p, q > 0$ and $a, b > 0$. Therefore $X > 0$.

$$\begin{aligned}
\text{E}[X] &= \frac{a^p}{p} + \frac{b^q}{q} \\
g(X) &\sim \begin{cases} -p\ln(a) & \text{with probability } \frac{1}{p} \\ -q\ln(b) & \text{with probability} \frac{1}{q} \end{cases} \\
g(\text{E}[X]) &= -\ln(\frac{a^p}{p} + \frac{b^q}{q}) \\
\text{E}[g(x)] &= -\frac{p\ln(a)}{p} + -\frac{q\ln(b)}{q} = -\ln(ab) \\
g(\text{E}[X]) &\leq \text{E}[g(X)] \\
-\ln(\frac{a^p}{p} + \frac{b^q}{q}) &\leq -\ln(ab) \\
ab &\leq \frac{a^p}{p} + \frac{b^q}{q}
\end{aligned}$$

This is Young's inequality.
Now let $a = X$ and $b = Y$.

$$XY \leq \frac{X^p}{p} + \frac{Y^q}{q} \to \text{E}[XY] \leq \frac{\text{E}[X^p]}{p} + \frac{\text{E}[Y^q]}{q}$$

Let $a = \frac{X}{A}$ and $b = \frac{Y}{B}$.

$$\frac{XY}{AB} \leq \frac{X^p}{pA^p} + \frac{Y^q}{qB^q} \to \frac{\mathrm{E}[XY]}{AB} \leq \frac{\mathrm{E}[X^P]}{pA} + \frac{\mathrm{E}[Y^q]}{qB}$$

Let $A = \mathrm{E}[X^p]^{\frac{1}{p}}$ and $B = \mathrm{E}[Y^q]^{\frac{1}{q}}$.

$$\frac{\mathrm{E}[XY]}{\mathrm{E}[X^p]^{\frac{1}{p}}\mathrm{E}[Y^q]^{\frac{1}{q}}} \leq \frac{1}{p} + \frac{1}{q} = 1$$

Therefore

$$\mathrm{E}[XY] \leq \mathrm{E}[X^p]^{\frac{1}{p}}\mathrm{E}[Y^q]^{\frac{1}{q}}$$

This is Halden's Inequality.

Let $0 < r < s$, $p = \frac{s}{r}$, $q = \frac{p}{p-1} = \frac{\frac{s}{r}}{\frac{s}{r}-1} = \frac{s}{s-r}$. Let $X = V^r$, $Y = 1$.
Then

$$\mathrm{E}[V^r] \leq \mathrm{E}[(V^r)^{\frac{s}{r}}]^{\frac{1}{\frac{s}{r}}}$$

Furthermore

$$\mathrm{E}[V^r] \leq \mathrm{E}[V^s]^{\frac{r}{s}}$$

If $\mathrm{E}[V^s]$ is fininte, then $\mathrm{E}[V^r]$ is finite.
For any random variable X, if $\mathrm{E}[|X|^s]$ is finite, then any moment less than s is finite too. Also,

$$\mathrm{E}[X^s] \leq \mathrm{E}[|X|^s]$$

This is because

$$\int_{\mathbb{R}} x^s f(x)\,dx \leq \int_{\mathbb{R}} |x^s f(x)|\,dx = \int_{\mathbb{R}} |x^s| f(x)\,dx$$

Consider the sequence of random variables $X_1, X_2, \dots$. There are many types of convergences.

Convergence in Distribution: Let $X_n \sim \begin{cases} \frac{1}{n+1} & \text{with probability} \frac{1}{3} \\ 1 - \frac{1}{n+1} & \text{with probability } \frac{2}{3} \end{cases}$.

For example, $X_3 = \begin{cases} \frac{1}{4} & \text{with probability } \frac{1}{3} \\ \frac{3}{4} & \text{with probability } \frac{2}{3} \end{cases}$.

Another one, $X_{99} \sim \begin{cases} 0.01 & \text{with probability } \frac{1}{3} \\ 0.99 & \text{with probability } \frac{2}{3} \end{cases}$.

Another one, $X_n \to \begin{cases} 0 & \text{with probability } \frac{1}{3} \\ 1 & \text{with probability } \frac{2}{3} \end{cases}$.

We say that $X_n \xrightarrow{d} X$ if for all x, $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$.
Theorem: if $\mathrm{Supp}[X_n] \in \mathbb{N}$ and $\mathrm{Supp}[X] \in \mathbb{N}$, then

$$X_n \xrightarrow{d} X \iff \forall x \in \mathbb{N} \lim p_{X_n}(x) = p_X(x)$$

Proof of Forward: Note that $p_{X_n}(x) = F_{X_n}(x+\frac{1}{2}) - F_{X_n}(x-\frac{1}{2})$.

$$\begin{aligned}\lim_{n\to\infty} p_{X_n}(x) &= \lim_{n\to\infty} F_{X_n}(x+\frac{1}{2}) - \lim_{n\to\infty} F_{X_n}(x-\frac{1}{2}) \\ &= F_X(x+\frac{1}{2}) - F_X(x-\frac{1}{2}) \\ &= p_X(x)\end{aligned}$$

Proof of Reverse: For all $x \in \mathbb{N}$,

$$\begin{aligned}\lim_{n\to\infty} F_{X_n}(x) &= \lim_{n\to\infty} \mathbb{P}(X_n \leq x) \\ &= \lim_{n\to\infty} \sum_{i=1}^{x} p_{X_n}(i) \\ &= \sum_{i=1}^{x} \lim_{n\to\infty} p_{X_n}(i) \\ &= \sum_{i=1}^{n} p_X(i) \\ &= F_X(x)\end{aligned}$$

If $X_n \sim \begin{cases} \frac{1}{n+1} & \text{with probability } \frac{1}{3} \\ 1-\frac{1}{n+1} & \text{with probability } \frac{2}{3}\end{cases}$, prove $X_n \xrightarrow{d} \text{Bern}\Big(\frac{2}{3}\Big)$.

$$p_{X_n}(x) = \Big(\frac{1}{3}\Big)^{\mathbb{1}_{x=\frac{1}{n+1}}} \Big(\frac{2}{3}\Big)^{\mathbb{1}_{x=1-\frac{1}{n+1}}} \mathbb{1}_{x\in\Big\{\frac{1}{n+1}, 1-\frac{1}{n+1}\Big\}}$$

$$\begin{aligned}\lim_{n\to\infty} p_{X_n}(x) &= \Big(\frac{1}{3}\Big)^{\lim_{n\to\infty}\mathbb{1}_{x=\frac{1}{n+1}}} \Big(\frac{2}{3}\Big)^{\lim_{n\to\infty}\mathbb{1}_{x=1-\frac{n}{n+1}}} \mathbb{1}_{x\in\Big\{\frac{1}{n+1}, 1-\frac{1}{n+1}\Big\}} \\ &= \Big(\frac{1}{3}\Big)^{\mathbb{1}_{x=0}} \Big(\frac{2}{3}\Big)^{\mathbb{1}_{x=1}} \mathbb{1}_{x\in\Big\{0,1\Big\}} \\ &= \text{Bern}\Big(\frac{2}{3}\Big)\end{aligned}$$

Notable Convergences:

$$X_n \sim \text{Binom}\Big(n, \frac{\lambda}{n}\Big) \xrightarrow{d} X \sim \text{Poisson}(\lambda)$$

Let $X_n \sim \text{Geom}(n\lambda)$ where $\text{Supp}[X_n] = \Big\{0, \frac{1}{n}, \frac{2}{n}, \dots\Big\}$. Then $X_n \xrightarrow{d} X \sim \text{Exp}(\lambda)$. Let $X_n \sim \text{Binom}(n,p)$ and $Y_n = \frac{X_n - np}{\sqrt{np(1-p)}}$, then

$$Y_n \xrightarrow{d} N(0,1)$$

$$X_n \xrightarrow{d} X \text{ means } \forall x \lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

The cdfs converges point wise. Note that

$$X_n \xrightarrow{d} X \iff \forall x \lim_{n\to\infty} p_{X_n}(x) = p_X(x)$$

This is true for discrete random variables with support $\mathbb{N}$ as well as for random variables with support $\mathbb{Z}$.
Consider $X_n \xrightarrow{d} c$ such that $c \in \mathbb{R}$. What is this? Recall that $c \sim \text{Deg}(c)$. That means that for all x, $\lim_{n\to\infty} F_{X_n}(x) = \begin{cases} 1 & \text{if } x \geq c \\ 0 & \text{if } x < c \end{cases}$.
Convergence in Probability: X_n converges in probability to a constant c, denoted $X_n \xrightarrow{p} c$ if $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|X_n - c| \geq \varepsilon) = 0$$

Let $X_n \sim U(-\frac{1}{n}, \frac{1}{n})$. Then $f_{X_n}(x) = \frac{n}{2}$. Prove that $X_n \xrightarrow{p} 0$.

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(|X_n - 0| \geq \varepsilon) &= 0 \\ \lim_{n\to\infty} \mathbb{P}(|X_n| \geq \varepsilon) &= 0 \\ \lim_{n\to\infty} \mathbb{P}(X_n < -\varepsilon) + \mathbb{P}(X_n > \varepsilon) &= 0 \\ \lim_{n\to\infty} \left(\frac{1}{n} - \varepsilon\right)\frac{n}{2}\mathbb{1}_{\varepsilon < \frac{1}{n}} + \left(\frac{1}{n} - \varepsilon\right)\frac{n}{2}\mathbb{1}_{\varepsilon < \frac{1}{n}} &= 0 \\ \lim_{n\to\infty} (1 - \varepsilon n)\mathbb{1}_{\varepsilon < \frac{1}{n}} &= 0 \end{aligned}$$

Consider $X_1, X_2, \dots \overset{iid}{\sim}$ with mean μ and variance σ^2. Define $\bar{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$. Consider $\bar{X}_1, \bar{X}_2, \dots$. They are all μ. But they are not iid since its variance is $\frac{\sigma^2}{n}$. Prove that $\bar{X}_n \xrightarrow{p} \mu$. This is the weak law of large numbers.

$$\begin{aligned} \lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) &= 0 \\ \text{Note that } \mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) &\leq \frac{\left(\frac{\sigma^2}{n}\right)}{\varepsilon^2} \\ \lim_{n\to\infty} \mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) &\leq \lim_{n\to\infty} \frac{\sigma^2}{n\varepsilon^2} = 0 \end{aligned}$$

This was easy because we assumed finite variance.

Convergence in L^r norm: For $r \geq 1$:

$$X_n \xrightarrow{L^r} c \text{ means } \lim_{n\to\infty} \text{E}[|X_n - c|^r] = 0$$

For example, $X_n \xrightarrow{L^1} c$ means $\lim_{n\to\infty} \text{E}[|X_n - c|] = 0$. We say this is convergence in mean. Also, $X_n \xrightarrow{L^2} c$ means $\lim_{n\to\infty} \text{E}[|X_n - c|^2] = 0$. We say this is mean square convergence.

If $X_n \to U\left(0, \frac{1}{n}\right)$, prove that $X_n \xrightarrow{L^r} 0 \forall r$.

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{E}[|X_n - 0|^r] &= 0 \\
\lim_{n\to\infty} \mathrm{E}[|X|^r] &= 0 \\
\lim_{n\to\infty} \mathrm{E}[X^r] &= 0 \\
\lim_{n\to\infty} \int_0^{\frac{1}{n}} |x|^r (n)\, dx &= \lim_{n\to\infty} \left[\frac{|x|^{r+1}}{r+1}\right]_0^{\frac{1}{n}} \\
&= \lim_{n\to\infty} \frac{1}{n^{r+1}(r+1)} n \\
&= \lim_{n\to\infty} \frac{1}{n^r(r+1)} = 0
\end{aligned}$$

Let $1 \leq r < s$. Prove that if $X_n \xrightarrow{L^s} c$ then $X_n \xrightarrow{L^r} c$. Recall that we used Halden's inequality to show that

$$\mathrm{E}[|X|^r] \geq (\mathrm{E}[|X|^s])^{\frac{r}{s}}$$

$$\lim_{n\to\infty} \mathrm{E}[|X_n - c|^r] \leq \lim_{n\to\infty} (\mathrm{E}[|X_n - c|^s])^{\frac{r}{s}} = \left(\lim_{n\to\infty} \mathrm{E}[|X_n - c|^s]\right)^{\frac{r}{s}} = 0^{\frac{r}{s}} = 0$$

Note that $\mathrm{E}[|X|] \geq 0$ since $|X|$ has positive support. Then

$$\lim_{n\to\infty} \mathrm{E}[|X_n - c|^r] \geq 0 \to \lim_{n\to\infty} \mathrm{E}[|X_n - c|^r] - 0 \to X_n \xrightarrow{L^r} c$$

Prove that if $X_n \xrightarrow{L^r} c$, then $X_n \xrightarrow{p} c$.

$$\lim_{n\to\infty} \mathbb{P}(|X_n - c| \geq \varepsilon) = \lim_{n\to\infty} \mathbb{P}(|X_n - c|^r \geq \varepsilon^r) \leq \lim_{n\to\infty} \frac{\mathrm{E}[|X_n - c|^r]}{\varepsilon^r} = 0$$

This works due to Markov's inequality.

Note that if $X_n \xrightarrow{p} c$ then it is not true that $X_n \xrightarrow{L^r} c$.
For example, let $X_n \sim \begin{cases} n^2 & \text{with probability } \frac{1}{n} \\ 0 & \text{with probability } 1 - \frac{1}{n} \end{cases}$. Here $X_n \xrightarrow{p} 0$.

$$\lim_{n\to\infty} \mathbb{P}(|X_n - 0| \geq \varepsilon) = \lim_{n\to\infty} \mathbb{P}(X_n \geq \varepsilon) = \lim_{n\to\infty} \mathbb{P}(X_n = n^2) = \lim_{n\to\infty} \frac{1}{n} = 0$$

But $X_n \overset{L^r}{\not\to} 0$.

$$\lim_{n\to\infty} \mathrm{E}[|X_n - 0|^r] = \lim_{n\to\infty} \mathrm{E}[X_n^r] = \lim_{n\to\infty} \sum_{\mathrm{Supp}[X_n]} x_n^r p_{X_n}(x_n) = \lim_{n\to\infty} (n^2)^r \frac{1}{n} = \lim_{n\to\infty} n^{2r-1} = 0$$

This shows that convergence in mean is stronger than convergence in probability because probabilities can variate but expectation will not.

Let $X_n \sim N\Big(0, \Big(\frac{1}{n}\Big)^2\Big)$. Prove that $X_n \xrightarrow{p} 0$.

$$\lim_{n\to\infty} \mathbb{P}(|X_n - 0| \geq \varepsilon) \leq \lim_{n\to\infty} \frac{\frac{\sigma^2}{n}}{\varepsilon^2} = 0$$

Prove that $X_n \xrightarrow{L^4} 0$.

$$\lim_{n\to\infty} \mathrm{E}[|X_n - 0|^4] = \lim_{n\to\infty} \mathrm{E}[X_n^4] = \lim_{n\to\infty} \frac{3}{n^2} = 0$$

This is because if we recall the characteristic function for X_n, $\phi_{X_n}(t) = e^{-\frac{1}{2}\sigma^2 t^2} = e^{-\frac{t^2}{2n}}$, then

$$\phi_{X_n}^{(4)}(t) = e^{-\frac{t^2}{2n}}\Big(\frac{3n^2 - 6nt^2 + t^4}{n^4}\Big)$$

and so

$$\phi_{X_n}^{(4)}(0) = \frac{3n^2}{n^4} = \frac{3}{n^2} = \mathrm{E}[X_n^4]$$

END OF FINAL MATERIAL

Imagine two random variables creating a joint density $f_{X,Y}(x,y)$. If $\mathrm{E}[X]$ and $\mathrm{E}[Y]$ was graphed, a horizontal slice would represent $\mathrm{E}[Y|X=x]$.

$$\begin{aligned}
\mathrm{E}[Y] &= \int_{\mathrm{Supp}[Y]} y f_Y(y)\,dy \\
&= \int_{\mathrm{Supp}[Y]} y \int_{\mathrm{Supp}[X]} f_{X,Y}(x,y)\,dx\,dy \\
&= \int_{\mathrm{Supp}[Y]} \int_{\mathrm{Supp}[X]} f_{Y|X}(x,y) f_X(x)\,dx\,dy \\
&= \int_{\mathrm{Supp}[Y]} \int_{\mathrm{Supp}[X]} y f_{Y|X}(x,y) f_X(x)\,dxdy \\
&= \int_{\mathrm{Supp}[X]} \left(\int_{\mathrm{Supp}[Y]} y f_{Y|X}(x,y)\,dy \right) f_X(x)\,dx \\
&= \int_{\mathrm{Supp}[X]} \mathrm{E}[Y|X] f_X(x)\,dx \\
&= \mathrm{E}[g(x)]
\end{aligned}$$

This is the Law of Total Expectation.

$$\mathrm{E}[Y] = \mathrm{E}_X[\mathrm{E}_Y[Y|X]]$$

Now consider the variance.

$$\begin{aligned}
\text{Var}_Y[Y] &= \text{E}[Y^2] - \text{E}^2[Y] \\
&= \text{E}_X[\text{E}_Y[Y^2|X]] - \text{E}_X^2[\text{E}_Y[Y|X]] \\
\text{Note that } \text{Var}_Y[Y|X] &= \text{E}[Y^2|X] - \text{E}^2[Y|X] \\
&= \text{E}_X[\text{Var}[Y|X] + \text{E}^2[Y|X]] - \text{E}_X^2[\text{E}_Y[Y|X]] \\
&= \text{E}_X[\text{Var}_Y[Y|X]] + \underbrace{\text{E}_X[\text{E}^2[Y|X]] - \text{E}_X^2[\text{E}_Y[Y|X]]}_{\text{E}[Q^2]-\text{E}^2Q=\text{Var}[Q]}
\end{aligned}$$

This is the Law of Total Variance.

$$\text{Var}_Y[Y] = \text{E}_X[\text{Var}_Y[Y|X]] + \text{Var}_X[\text{E}_Y[Y|X]]$$