If $A$ is a subset of the real line,

$$P[ X \in A ] = \int_A f_X(x) \, dx$$

Transformations of variables of the discrete type:

**Theorem:** Let $X$ be a discrete r.v. with p.m.f $p_X(x)$ and support $S_X$. If

$$T: Y = u(x)$$

is a one-to-one transformation which maps $S_X$ onto $S_Y$ (the support of $Y$), with inverse

$$T^{-1}: x = w(y)$$

then the p.m.f of $Y$ is given by

$$p_Y(y) = \begin{cases} p_X(w(y)) & \text{if } y \in S_Y \\ 0 & \text{elsewhere} \end{cases}$$

**Proof:**

$$p_Y(y) = P[Y = y] = P[u(X) = y] = P[X = w(y)] = p_X(w(y))$$

**Example:** Suppose the discrete r.v. $X$ has the p.m.f

$$p_X(x) = \begin{cases} \frac{e^{-7} 7^x}{x!} & \text{if } x = 0, 1, 2, \ldots \\ 0 & \text{elsewhere} \end{cases}$$
Find the p.m.f of \( Y = 5X \).

\[ T : \quad Y = 5X \]
\[ T' : \quad X = \frac{Y}{5} \]

\[ S_X = \{ x : \quad x = 0, 1, 2, \ldots \} \quad \rightarrow \quad S_Y = \{ y : \quad y = 0, 5, 10, \ldots \} \]

\[ p_Y(y) = p_X \left( \frac{y}{5} \right) = \begin{cases} \frac{e^{-\frac{y}{5}} \left( \frac{y}{5} \right)^y}{(\frac{y}{5})!} & \text{if } y = 0, 5, 10, \ldots \\ 0 & \text{elsewhere} \end{cases} \]

Transformations of variables of the continuous type.

**Theorem:** Let \( X \) be a continuous r.v. with p.d.f \( f_X(x) \) and support \( S_X \). If

\[ T : \quad Y = U(X) \]

with inverse

\[ T' : \quad X = W(Y) \]

is a one-to-one transformation that maps \( S_X \) onto \( S_Y \) (the support of \( Y \)) and \( \frac{dx}{dy} = \frac{dw(y)}{dy} \) is continuous and has non-zero value at all points in \( S_Y \) then the p.d.f \( f_Y(y) \) of \( Y \) is given by

\[ f_Y(y) = \int f_X(w(y)) \left| \frac{dx}{dy} \right| \quad \text{if } y \in S_Y \]
\[ = 0 \quad \text{elsewhere} \]
Example: Let the continuous r.v. $X$ have the p.d.f

$$f_X(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the p.d.f of $Y = -2\ln X$.

$T: \quad y = -2\ln x$

$T^{-1}: \quad x = e^{-\frac{y}{2}} \quad \frac{dx}{dy} = -\frac{1}{2} e^{-\frac{y}{2}}$

$S_X = \{ x : 0 < x < 1 \} \rightarrow S_Y = \{ y : y > 0 \}$

$$f_Y(y) = \int_{S_X} f_X(x) \left| \frac{dx}{dy} \right| dx = \frac{1}{2} e^{-\frac{y}{2}} \quad \text{if } y > 0$$

$$0 \quad \text{elsewhere}$$

Example: $f_X(x) = \begin{cases} 6x(1-x) & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$

Find the p.d.f of $Y = \frac{x}{1-x}$.

$T: \quad y = \frac{2x}{1-x}$

$T^{-1}: \quad x = \frac{y}{1+y} \quad \left| \frac{dx}{dy} \right| = \frac{1}{(1+y)^2}$

$S_X = \{ x : 0 < x < 1 \} \rightarrow S_Y = \{ y : y > 0 \}$

$$f_Y(y) = \int_{S_X} f_X(x) \left| \frac{dx}{dy} \right| dx = 6 \frac{y}{1+y} \left( 1 - \frac{y}{1+y} \right) \left( \frac{1}{1+y} \right)^2 = 6 - \frac{y}{1+y}$$

$$0 \quad \text{if } y > 0$$

$$\quad \text{elsewhere}$$
Expectation of a random variable.

Definition. Let \( U(X) \) be a function of a r.v. \( X \) which has p.d.f. \( f(x) \) and \( \int_{-\infty}^{\infty} |u(x)| f(x) \, dx < \infty \) then the expected value of \( U(X) \) is

\[
E[U(X)] = \int_{-\infty}^{\infty} u(x) f(x) \, dx
\]

If \( X \) is a discrete r.v. with p.m.f \( p(x) \) and \( \sum_{x} |u(x)| p(x) < \infty \), then \( E[U(X)] = \sum_{x} u(x) p(x) \)

Continuous case:

If \( Y = u(X) \) has the p.d.f \( f_Y(y) \) then

\[
\int_{-\infty}^{\infty} u(x) f(x) \, dx = E[U(X)] = E(Y) = \int_{-\infty}^{\infty} y \, f_Y(y) \, dy
\]

Discrete case:

If \( Y = u(X) \) has the p.m.f \( p_Y(y) \) then

\[
\sum_{x} u(x) p(x) = E[U(X)] = E(Y) = \sum_{y} y \, p_Y(y)
\]

Definition (Mean) If \( U(X) = X \) then \( E(X) \) is denoted by \( \mu \) and it is called the mean of \( X \) or the mean of the distribution of \( X \).
Definition. (Variance). If $U(X) = (X - \mu)^2$ then $E[(X - \mu)^2]$ is denoted by $\sigma^2$ and it is called the variance of $X$ or the variance of the distribution of $X$. $\sigma = \sqrt{\sigma^2}$ is called the standard deviation of $X$ or the standard deviation of the distribution of $X$.

Properties of Expectation:

Let $k_1, k_2$ be constants and let $U_1(X)$ and $U_2(X)$ be functions of a r.v $X$. Suppose the expectations of $U_1(X)$ and $U_2(X)$ exist.

1) $E(k_1) = k_1$
2) $E(k_1U_1(X)) = k_1E(U_1(X))$
3) $E[k_1U_1(X) + k_2U_2(X)] = k_1E[U_1(X)] + k_2E[U_2(X)]$

Note:

$\sigma^2 = E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2]$

$= E(X^2) - 2\mu E(X) + \mu^2$

$= E(X^2) - \mu^2$
Example. Let $A$ be some event, and let $P(A) = \theta$. Imagine that you are to receive $1$ if $A$ occurs and lose $1$ if $A$ does not occur. Let $X$ be the random variable defining your winning. The distribution of $X$ is

<table>
<thead>
<tr>
<th>$x$</th>
<th>$-1$</th>
<th>$1$</th>
<th>elsewhere</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(x)$</td>
<td>$1-\theta$</td>
<td>$\theta$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

and

$$E(X) = \sum_x x P(x) = -1 \cdot (1-\theta) + 1 \cdot \theta = 2\theta - 1$$

Thus $E(X) > 0$ if and only if $\theta > 0.5$, which agrees with our intuition.

Example. Suppose $P(x) = \begin{cases} (1-\theta)^{x-1} \theta & \text{for } x = 1, 2, \ldots \\ 0 & \text{elsewhere} \end{cases}$

where $0 < \theta < 1$. Find $E(X)$.

$$E(X) = \sum_{x=1}^{\infty} x (1-\theta)^{x-1} \theta$$

$$= \frac{d}{d\theta} \sum_{x=1}^{\infty} (1-\theta)^x$$

$$= \frac{d}{d\theta} \left( \frac{1}{1-(1-\theta)} \right)$$

$$= \theta \frac{d}{d\theta} \left( 1 - \frac{1}{\theta} \right) = \theta \left( \frac{1}{\theta^2} \right) = \frac{1}{\theta}$$
Example. Let $X$ be a continuous r.v. with p.d.f

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2} \quad \text{for} \quad -\infty < x < \infty.$$  

It is easy to check that the function $f(x)$ is indeed a p.d.f. 

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{1+x^2} \, dx = \frac{1}{\pi} \arctan x \bigg|_{-\infty}^{\infty} = 1$$

As regards expectation of $X$, the positive part of the defining integral equals

$$\frac{1}{\pi} \int_{0}^{\infty} \frac{x}{1+x^2} \, dx = \frac{1}{\pi} \frac{1}{2} \ln (1+x^2) \bigg|_{0}^{\infty} = \infty$$

In a similar fashion we check that

$$\int_{-\infty}^{0} x f(x) \, dx = -\infty$$

hence the expected value of $X$ does not exist.

Definition. Moment generating function (m.g.f.)

Suppose there is a $h > 0$ such that for $-h < t < h$ $E[e^{tX}]$ exists. The expectation is called the m.g.f. of $X$ and is denoted by $M(t)$.

Theorem. There is one-to-one correspondence between m.g.f.'s and the distribution functions
Now differentiating the expression for the m.g.f under the expectation sign (i.e. under the integral or summation sign) we obtain

\[ m'(t) = E(\exp(tX)), \quad m'(0) = E(X) \]
\[ m''(t) = E(X^2 \exp(tX)), \quad m''(0) = E(X^2) \]

and so on, so that

\[ \frac{d^k m(t)}{dt^k} = E\left[X^k \exp(tX)\right], \quad \frac{d^k m(t)}{dt^k}\bigg|_{t=0} = E(X^k) \]

Let \( M^{(k)}(t) \) denote the \( k \)th derivative of \( M(t) \).

Thus \( M^{(k)}(0) = E(X^k) \) for \( k = 1, 2, 3, \ldots \)

\( E[X^k] \) is called the \( k \)th moment of \( X \) about the origin.

Note that:

\[ \sigma^2 = E[(X - \mu)^2] = E(X^2) - \mu^2 = M''(0) - [M'(0)]^2 \]

Series representation of \( M(t) \):

\[ \exp(tX) = \sum_{i=0}^{\infty} \frac{(tx)^i}{i!} \quad \text{thus} \quad M(t) = \sum_{i=0}^{\infty} \frac{t^i}{i!} E(X^i) \]
Example. Suppose $E(X^i) = i!$, $i = 1, 2, 3, \ldots$

Find $M(t)$.

$M(t) = \sum_{i=0}^{\infty} \frac{t^i}{i!} E(X^i) = \sum_{i=0}^{\infty} \frac{t^i}{i!} i! = \sum_{i=0}^{\infty} t^i = \frac{1}{1-t}$

for $-1 < t < 1$.

Example. Let $X$ be a continuous r.v. with p.d.f $f(x) = x e^{-x}$ if $x > 0$, and zero elsewhere.

Find the m.g.f of $X$.

$M(t) = \int_0^\infty e^{tx} x e^{-x} dx = \int_0^\infty x e^{-x(1-t)} dx$

$= -\frac{x}{1-t} e^{-x(1-t)} \bigg|_0^\infty + \int_0^\infty \frac{1}{1-t} e^{-x(1-t)} dx$

$= -\left(\frac{1}{1-t}\right) e^{-x(1-t)} \bigg|_0^\infty - \frac{1}{(1-t)^2}$

for $t < 1$.

$\frac{d M(t)}{dt} \bigg|_{t=0} = 2 (1-t)^{-3} \bigg|_{t=0} = 2 = E(X)$

$\frac{d^2 M(t)}{dt^2} \bigg|_{t=0} = 6 (1-t)^{-4} \bigg|_{t=0} = 6 = E(X^2)$

Thus $\sigma^2 = E(X^2) - \mu^2 = 6 - 4 = 2$
Important inequalities.

**Theorem.** Let $k$ and $m$ be positive integers such that $k \leq m$. If $E[x^m]$ exists then $E[x^k]$ exists.

**Proof:** Without loss of generality assume that $X$ is a continuous r.v. with p.d.f $f(x)$ and let $A = \{ x : |x| \leq 1 \}$.

\[
E[x^k] = \int_{-\infty}^{\infty} x^k f(x) \, dx = \int_A x^k f(x) \, dx + \int_{A^c} x^k f(x) \, dx
\]

\[
\leq \int_A f(x) \, dx + \int_{A^c} 1 \, dx \leq \int_{-\infty}^{\infty} f(x) \, dx + \int_{-\infty}^{-1} 1 \, dx = 1 + E[|X|^m] < \infty
\]

**Theorem.** Let $Y$ be a r.v. such that $E(Y^{2m})$ exists, where $m > 0$. Then for any constant $\varepsilon > 0$,

\[
P[|Y| \geq \varepsilon] \leq \frac{E(Y^{2m})}{\varepsilon^{2m}}
\]

**Proof:** Without loss of generality assume that $Y$ is a continuous r.v. with p.d.f $f(y)$, and let $A = \{ y : |y| > \varepsilon \}$.

\[
E(Y^{2m}) = \int_A y^{2m} f(y) \, dy + \int_{A^c} y^{2m} f(y) \, dy \geq \int_A y^{2m} f(y) \, dy
\]

\[
\geq \varepsilon^{2m} \int_A f(y) \, dy = \varepsilon^{2m} P(A) = \varepsilon^{2m} P[|Y| \geq \varepsilon]
\]

In particular if we let $m = 1$, $Y = X - \mu$ where $\mu = E(X)$ then we have the following corollary.
Corollary: If $X$ is a r.v. with mean $\mu$ and variance $\sigma^2 < \infty$, then for every $\epsilon > 0$

$$P\left[ |X - \mu| \geq \epsilon \right] \leq \frac{\sigma^2}{\epsilon^2}$$

This corollary is a statement of the well-known Chebyshev's inequality.

If we let $\epsilon = k\sigma$, $k > 0$, we can write the Chebyshev's inequality in the following form

$$P\left[ |X - \mu| > k\sigma \right] \leq \frac{1}{k^2} \text{ for } k > 0$$

or equivalently

$$P\left[ |X - \mu| < k\sigma \right] \geq 1 - \frac{1}{k^2} \text{ for } k > 0.$$  

Theorem (Markov's Inequality): Let $u(X)$ be a non-negative function of the r.v. $X$. If $E[u(X)]$ exists then for every constant $c > 0$,

$$P\left[ u(X) \geq c \right] \leq \frac{E[u(X)]}{c}.$$  

Proof: Let $A = \{ x : u(x) > c \}$. WLOG assume $X$ to be continuous.

$$E[u(X)] = \int_A u(x) f(x) dx + \int_{A^c} u(x) f(x) dx \geq \int_A u(x) f(x) dx = \int_A \mathbf{1}_{u(x) > c} dx$$

$$= c P(A) = c P\left[ u(X) > c \right] . \text{ Thus } P\left[ u(X) > c \right] \leq \frac{E[u(X)]}{c}.$$
Definition. A function \( \phi(x) \) defined on \((a,b)\), \(a < b\), is said to be a \textbf{convex function} if for all \(x, y\) in \((a,b)\) and for all \(0 < \alpha < 1\),

\[
\phi \left( \alpha x + (1-\alpha)y \right) \leq \alpha \phi(x) + (1-\alpha) \phi(y)
\]

\(\phi\) is said to be strictly convex if the above inequality is strict.

Theorem. If \( \phi \) is differentiable on \((a,b)\) then

1) \( \phi \) is convex if and only if \( \phi'(x) \leq \phi'(y) \) for all \(a < x < y < b\).

2) \( \phi \) is strictly convex if and only if \( \phi'(x) < \phi'(y) \) for all \(a < x < y < b\).

Theorem. If \( \phi \) is twice differentiable on \((a,b)\) then

1) \( \phi \) is convex if and only if \( \phi''(x) \geq 0 \) for all \(a < x < b\).

2) \( \phi \) is strictly convex if \( \phi''(x) > 0 \) for all \(a < x < b\).

Informally, we can think of convex functions as functions that "hold water"—that is, they are bowl-shaped (\( \phi(x) = x^2 \) is convex). More formally, convex functions lie below lines connecting any two points. As \( \alpha \) goes from 0 to 1, \( \alpha \phi(x_1) + (1-\alpha) \phi(x_2) \) defines
a line connecting \((x_1, \phi(x_1))\) and \((x_2, \phi(x_2))\).

![Graph showing the relationship between \(x_1\), \(x_2\), and \(\phi(x)\).]

This line lies above \(\phi(x)\) if \(\phi(x)\) is convex. Furthermore, a convex function lies above all of its tangent lines, and that fact is the basis of Jensen's Inequality. Expand \(\phi(x)\) into Taylor series about \(c\), \(a < c < b\),

\[
\phi(x) = \phi(c) + \phi'(c)(x-c) + \frac{\phi''(c)(x-c)^2}{2!}
\]

where \(\phi\) is between \(x\) and \(c\). Since the last term is non-negative, we have

\[
\phi(x) \geq \phi(c) + \phi'(c)(x-c)
\]

**Theorem (Jensen's Inequality)**

For any r.v. \(X\), if \(\phi(x)\) is a convex function, then

\[
E[\phi(X)] \geq \phi[E(X)]
\]

**Proof:**

\[
\phi(x) \geq \phi(\mu) + \phi'(\mu)(x-\mu) \quad \text{where} \quad \mu = E(X)
\]

Taking expectation of both sides leads to the result.
One immediate application of Jensen's Inequality shows that $E(X^2) \geq [E(X)]^2$ since $\varphi(x) = x^2$ is convex.

Also if $x$ is positive then $\frac{1}{x}$ is convex, hence

$$E\left[\frac{1}{X}\right] \geq \frac{1}{E(X)},$$ another useful application.

**Example.** Jensen's Inequality can be used to prove an inequality between three kinds of means. If $a_1, a_2, \ldots, a_n$ are positive numbers, define

- **Arithmetic mean:** $a_A = \frac{2}{n} a_i$
- **Geometric mean:** $a_G = \left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}}$
- **Harmonic mean:** $a_H = \frac{1}{\frac{1}{a_i} - \frac{1}{2}}$

An inequality relating these means is

$$a_H \leq a_G \leq a_A$$

To apply Jensen's Inequality, let $X$ be a discrete r.v. with space $\{a_1, a_2, \ldots, a_n\}$ and $P(X = a_i) = \frac{1}{n}, i = 1, 2, \ldots, n$.

Since $-\ln x$ is a convex function, we have

$$E\left[-\ln X\right] \geq -\ln E(X) \quad \Rightarrow \quad E\left[-\ln X\right] \leq \ln E(X)$$
hence
\[ \ln q_G = \frac{1}{n} \sum_{i=1}^{n} \ln a_i = E[\ln X] \]
\[ \ln q_A = \ln \left( \frac{1}{n} \sum_{i=1}^{n} a_i \right) = \ln [E(X)] \]
Since \( E[\ln X] \leq \ln [E(X)] \) we have \( q_G \leq q_A \).

Now, use the fact that \( \frac{1}{x} \) is convex for \( x > 0 \), we get \( E \left[ \frac{1}{X} \right] \geq \frac{1}{E(X)} \).
\[ \frac{1}{a_H} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_i} = E \left[ \frac{1}{X} \right] \]
\[ \frac{1}{a_A} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_i} = \frac{1}{E(X)} \]

So, \( q_A \geq q_H \).

Since \( q_G \leq q_A \) we have \( \left( \prod_{i=1}^{n} a_i \right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{i=1}^{n} a_i \).

Now replace \( a_i \) by \( \frac{1}{a_i} \) (which is positive also) we have
\[ \left( \prod_{i=1}^{n} \frac{1}{a_i} \right)^{\frac{1}{n}} \leq \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_i} \]
or equivalently
\[ \frac{1}{\sqrt[n]{\prod_{i=1}^{n} \frac{1}{a_i}}} \leq \left( \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_i} \right)^{\frac{1}{n}} \]

That is \( q_H \leq q_G \). Thus \( q_H \leq q_G \leq q_A \).