Given values \( y = (y_1, \ldots, y_m) \) of \((Y_1, \ldots, Y_m)\), let \( \mathcal{A} \) be the set of points \((x_1, \ldots, x_n)\) such that
\[
\left\{ \begin{array}{l}
y_1 = h_1(x_1, \ldots, x_n) \\
\vdots \\
y_m = h_m(x_1, \ldots, x_n)
\end{array} \right.
\]
the joint PDF \( f_{\mathcal{A}}(x) \) is given by
\[
f_{\mathcal{A}}(x) = \sum_{(x_1, \ldots, x_n) \in \mathcal{A}} f_{\Xi}(x)
\]

\underline{Case 2:} \( n \) vs. \( \Xi_1, \ldots, \Xi_n \), continuous continuous
\((m=1)\) joint dist with joint PDF \( f_{\Xi}(x) \);
\[
\mathcal{I} = h(\Xi) \quad \text{For each } y \text{ define }
\quad A_y = \{ x : h(x) = y \}
\]
Then PDF of \( \mathcal{I} \) is
\[
f_{\mathcal{I}}(y) = \int_{A_y} \cdots \int f_{\Xi}(x) \, dx.
\]
Simple example of this result

\[ (X_1, X_2) \text{ joint continuous PDF} \]

with \[ Y = g(X_1 + \sigma_2 X_2 + b) \]

with PDF \[ f_Y(y) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) \frac{d x_2}{\sigma_1 \sigma_2} \]

Important Special Case

The simplest thing you can do with two RVS is to add them.

This is also important in statistics, where the sample mean \[ \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \] plays a key role.

In the result above, take \[ (\alpha_1, \alpha_2, b) = (1, 1, 0) \] to get \[ Y = X_1 + X_2. \]

Dist. of \( Y \) is called the convolution of the dists. of \( X_1 \) and \( X_2 \).
By the above result

$$f_x(\gamma) = \int_{-\infty}^{\infty} f_{x_i}(\gamma - x) f_{x_2}(x) \, dx$$

A more complicated example is defined to be

$$X_i \sim \text{CDF } F_{x_i} \Rightarrow P\{F_{x_i} \leq x\} = x$$

$$i = 1, \ldots, n \quad \text{(Continuous)}$$

$$\bar{X}_n = \min (X_1, \ldots, X_n)$$

$$\bar{X}_n \sim \text{max } (X_1, \ldots, X_n)$$

Take the first two statistics of the order problem.

$$F_{\bar{X}_n}(t) = P(\bar{X}_n \leq t)$$

$$= P(\bar{X}_n \leq t) = P(\bar{X}_1 \leq t, \bar{X}_2 \leq t, \ldots, \bar{X}_n \leq t)$$

$$= P(\bar{X}_1 \leq t) \cdot \ldots \cdot P(\bar{X}_n \leq t)$$

$$= [F_{\bar{X}_i}(t)]^n$$
So $Y_{(n)}$ has PDF $f_{Y_{(n)}}(t) = \frac{d}{dt} \left[ F_{X_0}(t) \right]^n$.

Similarly,

$$F_{Y_{(n)}}(t) = P(\bar{Y}_{(n)} \leq t) = 1 - P(\bar{Y}_{(n)} > t)$$

$$= 1 - P(X_1 > t, \ldots, X_n > t)$$

$$= 1 - \prod_i P(X_i > t)$$

$$\overset{equality}{=} 1 - \left[ 1 - F_{X_0}(t) \right]^n$$

Therefore $\bar{Y}_{(n)}$ has PDF $f_{\bar{Y}_{(n)}}(t) = \frac{d}{dt} F_{\bar{Y}_{(n)}}(t) = n \left[ 1 - F_{X_0}(t) \right]^{n-1} f_{X_0}(t)$. 
Generalities

Consider the earlier

Suppose $\mathbb{P}(\mathbf{X}, \mathbf{Y}) = \delta = 1$. Define new rv.

$P(\mathbf{X}, \mathbf{Y})_{\mathbf{Z}} = 1$. Define new rv.

Suppose $\mathbb{P}$ is a subset of $\mathbb{R}^n$ with

$Z$ continuous joint PDF $f_Z(z)$

$Z$ supported on $\mathbb{R}^n$.

Multivariate transformation

Assume $h_2, h_3, \ldots, h_n$ are functions $h_1, \ldots, h_n$ define a 1-1 transformation $T$ of $\mathbb{R}^n$ onto $\mathbb{R}^n$.

Some subset $T$.

$x_k = h_k (y_1, \ldots, y_n)$, $y_k = x_k$.
The joint PDF $f_{\mathbf{Z}}(\mathbf{Z})$ is

$$f_{\mathbf{Z}}(\mathbf{Z}) = \begin{cases} \sum_{\mathbf{Y} \in T} f_{\mathbf{X}} \left[ h_1'(\mathbf{Y}), \ldots, h_n'(\mathbf{Y}) \right] | \mathbf{J} | & \text{for } (\mathbf{Y}_1, \ldots, \mathbf{Y}_n) \in T \\ 0 & \text{else} \end{cases}$$

in which

$J$ is the determinant of the matrix

$$\begin{bmatrix} \frac{\partial h_1}{\partial \mathbf{Y}_1} & \cdots & \frac{\partial h_1}{\partial \mathbf{Y}_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial \mathbf{Y}_1} & \cdots & \frac{\partial h_n}{\partial \mathbf{Y}_n} \end{bmatrix}$$

and $| \cdot |$ is absolute value.

This is a chain rule generalization. $J$ is called the Jacobian of the transformation from $\mathbf{X}$ to $\mathbf{Z}$, named after the German mathematician Carl Gustav Jacob Jacobi (1804 - 1851).
Serves like a generalization of the derivative of the inverse in the earlier result. Example \((X_1, X_2)\) joint continuous PDF

\[
f_{X_1, X_2}(x_1, x_2) = \begin{cases} 4x_1x_2 & \text{for } 0 < x_1, x_2 < 1 \\ 0 & \text{else} \end{cases}
\]

(check: \(\int_0^1 \int_0^1 4x_1x_2 \, dx_1 \, dx_2 = 4\int_0^1 x_2 \left(\frac{x_1^2}{2}\right) \, dx_2 = 2\int_0^1 x_2^2 \, dx_2 = 2 \frac{x_2^3}{3} \bigg|_0^1 = 1\))

Let's work out the joint PDF of \((Y_1, Y_2) = (\frac{X_1}{X_2}, X_1 \cdot X_2)\) with

\[
y_1 = h_1(x_1, x_2) = \frac{x_1}{x_2}
\]

\[
y_2 = h_2(x_1, x_2) = x_1x_2
\]

Inverse transform:
solve \[ \begin{cases} \frac{x_1}{x_2} = y_1 \\ x_1 x_2 = y_2 \end{cases} \] for \((x_1, x_2)\): \[ x_1 = h^{-1}_1(y_1, y_2) = \sqrt{y_2/y_1} \]

image: how does \(0 < x_1 < 1\), \(0 < x_2 < 1\) transform? \[ \begin{bmatrix} 1 & -\frac{1}{y_1} \\ 0 & 1 \end{bmatrix} \]

\[ \begin{align*}
(a) & \quad \sqrt{y_1 y_2} > 0, \quad \sqrt{y_1 y_2} < 1 \\
(b) & \quad \sqrt{y_2/y_1} > 0, \quad \sqrt{y_2/y_1} < 1 \\
(c) & \quad \sqrt{y_2/y_1} > 0, \quad \sqrt{y_2/y_1} < 1 \\
(d) & \quad \sqrt{y_2/y_1} > 0, \quad \sqrt{y_2/y_1} < 1
\end{align*} \]

So \(\sqrt{y_1 y_2} > 0\) \(\Rightarrow\) \(\sqrt{y_1 y_2} < 1\) (9) equivalent to \((y_1 > 0) \land (y_2 > 0)\) or \((y_1 < 0) \land (y_2 < 0)\)

but \(y_1 = \frac{x_1}{x_2} > 0\) so it must be \((y_1 > 0) \land (y_2 > 0)\) (c) leads to the same thing

(b) says \(y_2 < \frac{1}{y_1}\) (d) says \(y_2 < y_1\)
\[ h_1^{-1}(y_1, y_2) = \sqrt{y_1 y_2} \]
\[ h_2^{-1}(y_1, y_2) = \sqrt{\frac{y_2}{y_1}} \]
\[ \frac{d}{dy_1} h_2^{-1} = -\frac{1}{2} \left( \frac{y_2}{y_1^2} \right)^{\frac{1}{2}} \]
\[ \frac{d}{dy_2} h_2^{-1} = \frac{1}{2} \sqrt{\frac{1}{y_1 y_2}} \]

So
\[
\frac{\partial}{\partial y_1} h_1^{-1} = \frac{1}{2} \sqrt{\frac{y_2}{y_1}}
\]
\[
\frac{\partial}{\partial y_2} h_1^{-1} = \frac{1}{2} \sqrt{\frac{y_1}{y_2}}
\]

So
\[
J = \det \begin{bmatrix}
\frac{1}{2} (\frac{y_2}{y_1})^{\frac{1}{2}} & \frac{1}{2} (\frac{y_1}{y_2})^{\frac{1}{2}} \\
-\frac{1}{2} \left( \frac{y_2}{y_1} \right)^{\frac{1}{2}} & \frac{1}{2} \left( \frac{1}{y_1 y_2} \right)^{\frac{1}{2}}
\end{bmatrix}
\]

Recall
\[
\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc
\]

\[
J = \frac{1}{2y_1}
\]

and (since \( y_1 > 0 \))
\[
|J| = \frac{1}{2y_1}
\]
To finish the calculation, in the PDF of $T$, \( f_T (t) = \frac{4x_1 x_2}{\left( \frac{y_1}{y_2} \right)^{\frac{1}{2}}} \) for \( 0 < t < 1 \) and \( 0 < y_2 < y_1 \).

Substitute \( x_1 = \sqrt{y_1 y_2} \), \( x_2 = \sqrt{\frac{y_2}{y_1}} \) and bring in the Jacobian:

\[
f_T (t) = f_{X_1 X_2} \left[ h_1^{-1}(t), h_2^{-1}(t) \right] |J|
\]

\[
= 4 \left( \sqrt{y_1 y_2} \right) \left( \sqrt{\frac{y_2}{y_1}} \right)^{\frac{1}{2}} \left( \frac{y_2}{y_1} \right)^{-\frac{1}{2}}
\]

\[
= \begin{cases} 2 \frac{y_2}{y_1} & \text{for } (y_1, y_2) \in T \\ 0 & \text{else} \end{cases}
\]
A useful trick: start with \((X_1, X_2)\) joint dist., suppose you're interested only in the dist. of \(X_1 = h_1(X_1, X_2)\). Then one way to compute this dist. is with the following 3 steps.

**Step 1:** Find another \(X_2 = h_2(X_1, X_2)\) such that the transformation \((X_1, X_2) \rightarrow (X_1, X_2)\) is 1-1 with a differentiable inverse transformation & the calculations are straightforward.

**Step 2:** Work out the joint dist. of 
\((X_1, X_2)\).

**Step 3:** Integrate \(X_2\) out of the joint dist. (i.e., marginalize over \(X_2\)) to get the marginal dist. of \(X_1\).
Example of 
4 $x_2$ not 
wouldn't work

$I_1 = 2 \mathbf{x}_1$
$I_2 = 3 \mathbf{x}_1 = \frac{3}{2} \mathbf{x}_1$

Here $I_2$ is linearly dependent on $I_1$, so the rank of the Jacobian matrix is only 1 and its determinant is therefore 0.

Earlier

Example: $(\mathbf{x}_1, \mathbf{x}_2)$ have joint (continuous) PDF $f_{\mathbf{x}_1, \mathbf{x}_2}(x_1, x_2) = \begin{cases} 4x_1x_2 & 0 < x_1, x_2 < 1 \\ 0 & \text{else} \end{cases}$

Earlier we found that with $(\mathbf{x}_1, \mathbf{x}_2) = (\frac{x_1}{x_2}, x_1, x_2)$ the transformed PDF was

$f_{T_1, T_2}(y_1, y_2) = \begin{cases} \frac{2y_2}{y_1} & \text{for } (y_1, y_2) \in T \\ 0 & \text{else} \end{cases}$

where $T = \{(y_1, y_2) : y_1 > 0, y_2 < \min(y_1, \frac{1}{y_1})\}$. 
Suppose you were only really interested in the dist. of $Z_i = \frac{X_i}{X_2}$; then all you have to do is integrate $Z_2$ out of the joint dist.

For $\gamma > 0$, the allowable region for $\gamma_2$ is in two parts: for $0 < \gamma_1 < 1$, $0 < \gamma_2 < 1$, and for $\gamma_1 > 1$, $0 < \gamma_2 < \frac{1}{\gamma_1}$.

So, $f_{\gamma_1}(\gamma_1) = \int_0^{\gamma_1} 2\left(\frac{\gamma_2}{\gamma_1}\right)^{-3}\,d\gamma_2 = \gamma_1^{-2}$ for $0 < \gamma_1 < 1$

\[
\begin{align*}
\int_0^{1/\gamma_1} 2\left(\frac{\gamma_2}{\gamma_1}\right)^{-3}\,d\gamma_2 &= \gamma_1^{-2} \quad \text{for } \gamma_1 > 1
\end{align*}
\]

$f_{\gamma_1}(\gamma_1)$

Weird pdf: not differentiable at $\gamma_1 = 1$
\[ A^{-1} = \frac{1}{2} \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2} A \]

\[ A = \begin{pmatrix} c & d \\ e & f \end{pmatrix} \]

\[ \det A = cf - de \]

\[ \det A + 1 = 0 \]

\[ \rightarrow \quad A^{-1} \]

---

Example

where \( A \) is an invertible matrix. Also, the PDF of \( \mathbf{Z} \) is

\[ f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{|A|} f_{\mathbf{X}}(A^{-1} \mathbf{z}) \]

for the transformation of \( \mathbf{X} = (X_1, \ldots, X_n) \) to \( \mathbf{Z} = (Z_1, \ldots, Z_n) \) where \( A \) is linear.
Example: Toy sickness

disease (continued)

Earlier we worked out the
discrete dist. of the rv

\( I = \# \text{ of } F S \text{ babies in family}
\)
of 5, both parents carriers.

We showed that \( I \sim \text{ binomial } (n, p) \) with

\[ n = 5, \quad p = \frac{1}{4} \]

<table>
<thead>
<tr>
<th>( y )</th>
<th>( P(I = y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>((5)\left(\frac{1}{4}\right)^0\left(\frac{3}{4}\right)^5 = 0.2373)</td>
</tr>
<tr>
<td>1</td>
<td>((\frac{5}{1})(\frac{1}{4})\left(\frac{3}{4}\right)^4 = 0.3955)</td>
</tr>
<tr>
<td>2</td>
<td>((\frac{5}{2})(\frac{1}{4})^2\left(\frac{3}{4}\right)^3 = 0.2637)</td>
</tr>
<tr>
<td>3</td>
<td>((\frac{5}{3})(\frac{1}{4})^3\left(\frac{3}{4}\right)^2 = 0.0879)</td>
</tr>
<tr>
<td>4</td>
<td>((\frac{5}{4})(\frac{1}{4})^4\left(\frac{3}{4}\right) = 0.0146)</td>
</tr>
<tr>
<td>5</td>
<td>((\frac{5}{5})(\frac{1}{4})^5\left(\frac{3}{4}\right)^0 = 0.0010)</td>
</tr>
</tbody>
</table>

\[ P(I = y) = \begin{cases} \binom{n}{y} p^y (1-p)^{n-y} & \text{if } y = 0, \ldots, n \\ 0 & \text{else} \end{cases} \]

\[ \begin{array}{c}
\text{About how many FS babies should these parents expect to have?}
\end{array} \]

(center of dist. of \( I \))
Most likely outcome is 1 T-S day (mode of the dist. of $Y$)

A2 (physics idea)

Let's work out the center of mass of the distribution (balance point)

Toy example

\[
\begin{pmatrix}
1 \\
2 \\
3 \\
4 \\
5 \\
6 \\
7 \\
8 \\
9 \\
\end{pmatrix}
\]

Outlier

Let's find the place $c$ where the histogram balances: where (the sum of forces exerted by the histogram bars to the left of $c$)

equals (the sum of forces to the right):

\[
\begin{pmatrix}
y_1 \\
y_2 \\
y_3 \\
y_4 \\
y_5 \\
y_6 \\
y_7 \\
y_8 \\
y_9 \\
\end{pmatrix} \rightarrow \begin{pmatrix}
y_1 - c \\
y_2 - c \\
y_3 - c \\
y_4 - c \\
y_5 - c \\
y_6 - c \\
y_7 - c \\
y_8 - c \\
y_9 - c \\
\end{pmatrix}
\]

want sum $= 0$

\[
\sum_{i=1}^{n} (y_i - c) = 0 = 
\sum_{i=1}^{n} y_i - nc = 0
\]
A3: median of $\frac{\sum y}{n} - nc = 0 \to T$

The A3 is the hep peti

Here $\{\frac{1}{\sqrt{2}} \}$

$(\sum y) = n \text{ like a map}$

If some values are more probable than others, the

Generalization of (5) weight on each

Let $Z$ be a bounded

Bound of $\text{it possible values}$

Def: Am $\exists \text{ some values}$

$\frac{1}{\sqrt{2}}$ of the sample of $\forall$
\[
E(Y) = \sum_{y=0}^{\infty} y f_Y(y) = \sum_{y=0}^{\infty} y P(Y=y)
\]

T. S. example

\[E(Y) = (0)(.2373) + (1)(.3955) + \ldots + (5)(.0012) = 1.2500000\]

Symbolically if \(Y\)~Binomial\( (n, p) \)

then \(E(Y) = \sum_{y=0}^{n} y \binom{n}{y} p^y (1-p)^{n-y}\)

\[= \sum_{y=1}^{n} y \frac{n!}{y!(n-y)!} p^y (1-p)^{n-y}\]

\[= \sum_{y=1}^{n} \frac{n(n-1)!}{y(y-1)!(n-1-(y-1))!} p^y (1-p)^{n-y}\]

\[= np \sum_{y=1}^{n} \frac{(n-1)!}{(y-1)!(n-y)!} p^{y-1} (1-p)^{n-1-(y-1)}\]
\[
\sum_{y=1}^{n} \binom{n-1}{y-1} p^{y-1} (1-p)^{n-1-(y-1)}
\]

\[
= n \sum_{y=1}^{n} \binom{n-1}{y-1} p^{y-1} (1-p)^{n-1-(y-1)}
\]

\[
= n \binom{n-1}{y-1} p^{y-1} (1-p)^{n-1-(y-1)}
\]

This is = 1

So: if Binomial\((n-1, p)\) dist.

\(Z \sim Binomial\(n, p)\)

because Binomial probability add up to 1

for \(n \geq 1\), \(E(Z) = np\)

When \(n = 1\), \(Binomial\(1, p) = Bernoulli(p)\).

In this case \(E(Z) = 0 \cdot P(Z = 0) + 1 \cdot P(Z = 1)\)

so: for all \(n \geq 1\) (integer)

\[
E(Z) = np \text{ with } n = 1 \text{ and } 0 < p < 1, \quad Z \sim Binomial\(n, p) \implies E(Z) = np.
\]
If discrete $X$ is unbounded, the expectation of $X$ may not exist, because

$$\sum_{x<0} x f_X(x) = -\infty \quad \text{(and) } \sum_{x \geq 0} x f_X(x) = +\infty$$

or the distribution puts too much mass near $\pm \infty$.

**Def.** $X$ discrete RV with

$M f_X(x)$; consider $\sum_{x<0} x f_X(x)$ and $\sum_{x \geq 0} x f_X(x)$. If both sums are infinite, $E(X)$ is undefined (or does not exist);

if at least one sum is finite, then

$$E(X) = \sum_{x \in \mathbb{X}} x f_X(x) \quad \text{exists (it may still be infinite)}$$
To create a discrete rv whose mean doesn't exist, you have to play a careful game, because \( \sum_{x \in \mathbb{X}} f_\mathbb{X}(x) \) has to be finite (it has to equal 1) but \( \sum_{x \in \mathbb{X}} x f_\mathbb{X}(x) \) has some \( x \) to be infinite. \[ \text{Example} \]

The harmonic series \( \sum_{i=1}^{\infty} \left( \frac{1}{i} + \frac{1}{i+1} + \frac{1}{i+2} + \ldots \right) = \sum_{x=1}^{\infty} \frac{1}{x} \) was known to the ancient Greeks, because the wavelengths of the overtones of a vibrating string are \( \frac{1}{2}, \frac{1}{3}, \ldots \). of the fundamental wavelength of the string. The fact that \( \sum_{x=1}^{\infty} \frac{1}{x} = \infty \) (ie, the harmonic series diverges) was first shown in the 1300s (!) by the French philosopher Nicole Oresme (\~{}1320 - 1382).
It's clear from this divergence that (1): you can't create a rv \( X \) with \( p_X(n) \)
\[
p(X=x) = \frac{c}{x}, \quad x = 1, 2, \ldots
\]
because the probability sum to \( \infty \). But
\[
p(X=x) = \frac{c}{x^2}
\]
or \( p(X=x) = \frac{c}{x(x+1)} \) turns out to work; for example,
\[
\sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6} (1) \text{ and, more conveniently,}
\[
\sum_{x=1}^{\infty} \frac{1}{x(x+1)} = 1.
\]

If we try to construct two pathological discrete distributions, to show what can go wrong with the idea of expectation.

\[
\text{Example} \quad f_X(x) = \begin{cases} \frac{1}{x(x+1)} & x = 1, 2, \ldots \\ 0 & \text{else} \end{cases}
\]
\[ E(\mathcal{X}) = \sum_{x=1}^{\infty} \frac{1}{x(x+1)} = \sum_{x=1}^{\infty} \frac{1}{x+1} = +\infty \]

So \( E(\mathcal{X}) \) exists, it's just infinite.

**Example 2**

\[ f_{\mathcal{X}}(x) = \begin{cases} \frac{1}{2|x| (1|x|+1)} & x = \pm 1, \pm 2, \\ 0 & \text{else} \end{cases} \]

We already know that \( \sum f_{\mathcal{X}}(x) = 1 \), so \( \mathcal{X} \) is a well-defined rv, but

\[ \sum_{x=-\infty}^{\infty} x \cdot \frac{1}{2|x| (1|x|+1)} = -\infty \]

and

\[ \sum_{x=1}^{\infty} x \cdot \frac{1}{2x(x+1)} = +\infty \], so \( E(\mathcal{X}) \) does not exist.

---

Expectation for continuous rvs

**Def.** \( \mathcal{X} \) bounded continuous rv

We won't work with pathological rv's, mostly.
with PDF \( f_{\xi}(x) \rightarrow E(\xi) = \int_{-\infty}^{\infty} x f_{\xi}(x) \, dx \).

Example: \( \xi \sim \text{Exponential} \left( \lambda \right) \) \( (\lambda > 0) \):

recall that \( f_{\xi}(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \text{else} \end{cases} \)

\[
\int_{0}^{\infty} 2 x e^{-\lambda x} \, dx = \frac{1}{\lambda}
\]

So \( E(\xi) = \int_{0}^{\infty} x \cdot f_{\xi}(x) \, dx \)

For this reason, many people parameterize the exponential distribution differently:

Alternative definition: \( \xi \sim \text{Exponential} \left( \frac{1}{\eta} \right) \) \( (\eta > 0) \)

\[ f_{\xi}(x) = \begin{cases} \frac{1}{\eta} e^{-\frac{x}{\eta}} & \text{for } x > 0 \\ 0 & \text{else} \end{cases} \]

With this parameterization, you can see that \( E(\xi) = \frac{1}{\eta} \) (expectation).
Nevertheless, to avoid confusion with \( \text{de}^{-\lambda x} \), I'll stick with \( \text{de}^{-\lambda x} \). If continuous rv \( X \) is unbounded, a bit of care is once again required to define \( E(\mathbb{I}) \). 

Def.

If continuous rv with PDF \( f_X(x) \); consider \( \int_{-\infty}^{\infty} y f_X(y) \, dy \) and \( \int_{0}^{\infty} y f_X(y) \, dy \). If both integrals are infinite, \( E(\mathbb{I}) \) is undefined (or does not exist); if at least one of these integrals is finite, \( E(\mathbb{I}) = \int_{-\infty}^{\infty} y f_X(y) \, dy \) exists (but it may still be infinite).
Example: A list that does arise in practical statistical applications is the Cauchy distribution (attributed to Augustin-Louis Cauchy (1789–1857), a French mathematician who wrote 800 & 5 textbooks & 5 research articles in a 52-year period (15/yr), but actually first studied carefully by Poisson in 1824). \[ f(x) = \frac{1}{\pi (1+x^2)} \]
is the (standard) Cauchy distribution.

It does integrate to 1, but \[ \int_{-\infty}^{\infty} \frac{1}{\pi (1+x^2)} \, dx = \frac{1}{\pi} \ln(1+x^2) \bigg|_{-\infty}^{\infty} \]
and \[ \int_{-\infty}^{0} \frac{1}{\pi (1+x^2)} \, dx = -\infty, \text{ so } E(X) \text{ does not exist, because its tails go to 0 extremely slowly.} \]
This is because for large $\gamma$, \( \frac{1}{1+\gamma^2} \approx \frac{1}{\gamma} \)
and \( \int_0^\infty \frac{1}{\gamma} \, d\gamma = +\infty \), the continuous analogue of the harmonic series.

Expectation of a function of a random variable:

\[ f_\mathcal{X}(x), \quad \mathcal{X} = h(\mathcal{X}). \]

Method 1: Work out PDF \( f_\mathcal{Y}(\gamma) \);

Then \( E(\mathcal{Y}) = \int_\mathbb{R} \gamma f_\mathcal{Y}(\gamma) \, d\gamma \).

(If this exists)

Method 2 (faster):

\[ E(\mathcal{Y}) = \int_\mathbb{R} h(x) f_\mathcal{X}(x) \, dx. \]

Discrete version:

\[ E[h(\mathcal{X})] = \sum_{\text{all } x} h(x) f_\mathcal{X}(x). \]
DS (and some other people) call method \( \text{(17.4)} \)
the law of the unconscious statistician,
because method 2 looks like a definition
but is actually a theorem (in full generality)
(measure theory: push-forward measure,...)

Example \( X \sim \text{Exponential}(\lambda), \lambda > 0 \)

\[ E(X) = \frac{1}{\lambda} \]

\[ Y = X^2 \]

\[ E(Y) = \int_0^\infty x^2 \lambda e^{-\lambda x} \, dx = \frac{2}{\lambda^2} \]

Notice that

\[ E(X^2) \neq \left[ E(X) \right]^2 \]

\[ \frac{2}{\lambda^2} \neq \left( \frac{1}{\lambda} \right)^2 \]

The only functions \( Y = h(X) \) for which \( E[h(X)] = h(E(X)) \)
are linear: \( h(x) = a + bx \), as we'll see later.
Def. A function \( g: \mathbb{R} \to \mathbb{R} \) is convex if for all constants \( a, b \) and \( x, y \in \mathbb{R} \), we have:

\[
\frac{g(ax + by) - g(x) - g(y)}{a} \leq 0.
\]

Properties

1. If \( g \) is a constant, \( g(x) = c \) for all \( x \), then \( E(g) = c \).
2. If \( E(x) \geq 0 \), then \( E(x + b) \geq E(x) + b \).
3. If \( E(x) \) is finite, \( E(a_1 x_1 + \cdots + a_n x_n) \) is finite with \( E(x) = \sum_{i=1}^n a_i x_i \).
4. If \( E(x) \) is a constant, \( E(x + b) = E(x) + b \).
If for every $0 < d < 1$ and every $x$ and $y$, 

$$g \left[ d \frac{x}{z} + (1-d) \frac{y}{z} \right] \leq d g(x) + (1-d) g(y)$$

Graphical version of this: pick any two points on the function and connect them with a line segment; the function is convex if the line segment entirely lies above the function except at the endpoints.

$g$ is concave if

$$g \left[ d \frac{x}{z} + (1-d) \frac{y}{z} \right] \geq d g(x) + (1-d) g(y)$$
The expectation of a random vector \( \overline{\mathbf{X}} = (X_1, \ldots, X_n) \) is \( E(\overline{\mathbf{X}}) = \left[ E(X_1), \ldots, E(X_n) \right] \).

(5) If a convex function \( g \) maps the expectation of a random vector with finite expectation \( E(\overline{\mathbf{X}}) \) to \( E\left[g(\overline{\mathbf{X}})\right] \), then
\[
\mathbb{E}[g(\overline{\mathbf{X}})] \geq g\left[\mathbb{E}(\overline{\mathbf{X}})\right].
\]

(1) If \( X_1, \ldots, X_n \) are independent and identically distributed (i.i.d.), then
\[
\mathbb{E}\left[g(\overline{\mathbf{X}})\right] = g\left[\mathbb{E}(\overline{\mathbf{X}})\right] - \frac{1}{2}n \sigma^2,
\]

(Attributed to Johan Jensen (1859 - 1925), Danish mathematician and engineer)

Suppose that \( X_1, \ldots, X_n \sim \text{Bernoulli}(p) \).

Then \( \mathbb{E}(X_i) = 0 \cdot (1 - p) + 1 \cdot p = p \) and
\[
P(X = 0) = \frac{\binom{n}{0}}{2^n} \quad P(X = 1) = \frac{\binom{n}{1}}{2^n}
\]

\[
\mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{E}(X_i) = np = \text{mean of } \text{Binomial}(n, p).
\]