**Theory of Probability**

**Oct 12, 2020**

The expectation of $X$

$$E[X] = \sum_i x_i p(x_i) \quad \text{when} \quad P[X = x_i] = p(x_i).$$

Consider this example:

Roll 1, get 1
Roll 2, get 4

$Y = X^2$, when $X = \text{roll of a die}$.

$$P[Y = 1] = P[X = 1] = \frac{1}{6}$$
$$P[Y = 4] = P[X = 2] = \frac{1}{6}$$

$$E[Y] = \sum_i y_i p(y_i)$$

$$= \sum_i x_i^2 p(x_i)$$

expected value of $Y = X^2$.

In general, one can show that if $Y = g(X)$, then

$$E[Y] = E[g(X)] = \sum_i g(x_i) p(x_i).$$

Note: $E[X^2] \neq (E[X])^2$

$$(E[X])^2 = \left( \sum_i x_i p(x_i) \right)^2 = \sum_i \sum_j x_i x_j p(x_i) p(x_j) = \sum_i x_i^2 p(x_i)$$
**Corollary:** Expectation is a linear operator:

\[ E[aX + b] = \sum_i (ax_i + b) p(x_i) = a \sum_i x_i p(x_i) + b \sum_i p(x_i) = aE[X] + b \Rightarrow E \text{ is a linear transformation}. \]

**Remark:** \( E[X] \) is also known as the 1st moment of \( X \).

Furthermore, \( E[X^n] \) is known as the \( n \)th moment of \( X \):

\[ E[X^n] = \sum_i x_i^n p(x_i). \]

**Variance**

It's useful to talk about characteristics of random variables:

- Expected value \( \Rightarrow \) mean
- min value, max value
- "most likely value" \( \Rightarrow \) mode
- "spread" of a random variable.

Consider the following three R.V.s:

<table>
<thead>
<tr>
<th>R.V.</th>
<th>Probability</th>
<th>Expected Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( W )</td>
<td>( P[W = 0] = 1 )</td>
<td>( E[W] = 0 )</td>
</tr>
<tr>
<td>( Y )</td>
<td>( P[Y = -1] = \frac{1}{2} )</td>
<td>( E[Y] = 0 )</td>
</tr>
<tr>
<td>( Z )</td>
<td>( P[Z = 100] = \frac{1}{2} )</td>
<td>( E[Z] = 0 )</td>
</tr>
</tbody>
</table>

\( P[Z = -100] = \frac{1}{2} \)
Let \( \mu = E[X] \).

Want to characterize \( E[|X-\mu|] \), but it turns out to be better mathematically to examine \( E[(X-\mu)^2] \).

Variance of \( X \) = \( \text{Var}[X] = E[(X-\mu)^2] \)

\[
E[(X-\mu)^2] = \sum (x_i - \mu)^2 p(x_i)
\]

\[
= \sum (x_i^2 - 2\mu x_i + \mu^2) p(x_i)
\]

\[
= \sum x_i^2 p(x_i) - 2\mu \sum x_i p(x_i) + \mu^2 \sum p(x_i)
\]

\[
= \frac{E[X^2]}{E[X]} - \mu \frac{E[X]}{E[X]} + \mu^2 \frac{1}{E[X]} - \mu^2
\]

\[
= E[X^2] - 2\mu^2 + \mu^2
\]

\[
= E[X^2] - \mu^2
\]

\[
= E[X^2] - (E[X])^2
\]

\[
\text{Variance is not a linear transformation.}
\]

\[
\text{Var}[aX+b] = E[(aX+b)^2] - (E[aX+b])^2
\]

\[
= E[a^2X^2 + 2abX + b^2] - (a\mu + b)^2
\]

\[
= a^2 E[X^2] + 2ab \mu + b^2 - a^2 \mu^2 - 2ab \mu - b^2
\]

\[
= a^2 (E[X^2] - (E[X])^2) = a^2 \text{Var}[X]
\]
Another useful (common) quantity is the standard deviation:

\[
\text{Std}[X] = \sqrt{\text{Var}[X]}
\]

\[
\text{Std}[aX+b] = \sqrt{\text{Var}[aX+b]}
\]

\[
= \sqrt{a^2 \cdot \text{Var}[X]}
\]

\[
= a \cdot \text{Std}[X],
\]