Expectation of a product when the $X_i$ are independent

Let $X_1, \ldots, X_n$ be independent r.v. each with finite $E(\mathbf{X}_i)$.

\[
E\left(\prod_{i=1}^{n} X_i\right) = \prod_{i=1}^{n} E(X_i)
\]

Contrast this with a sum: $E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$

Whether the $X_i$ are independent or not;

\[
E\left(\prod_{i=1}^{n} x_i\right) = \prod_{i=1}^{n} E(x_i) \text{ only when the } x_i \text{ are independent}
\]

Ex: You have a (Brita) water filter that you use to improve the taste of S.C. water. How much better would the filter do its job if you filtered the water twice instead of once?

Let $\Sigma_1 = \text{proportion of bad stuff removed in the 1st filtering}$

$\Sigma_2 = \text{proportion removed in 2nd filtering}$
Reasonable to assume that $X_1, X_2$ are independent

Suppose they're IID with common PDF

$$f_{X_i}(x_i) = \begin{cases} 4x_i^3 & 0 < x_i < 1 \\ 0 & \text{else} \end{cases}$$

Set $\bar{Y} =$ proportion of original bad stuff remaining after 2 filtrations = $(1-\bar{Y}_1)(1-\bar{Y}_2)$

Then $E(\bar{Y}) = E[(1-\bar{X}_1)(1-\bar{X}_2)] = E[(1-\bar{X}_1)] \cdot E[(1-\bar{X}_2)]$

$\bar{X}_1, \bar{X}_2$ independent $\Rightarrow (1-\bar{X}_1)(1-\bar{X}_2)$ independent too

$$E(1-\bar{X}_1) = E(1-\bar{X}_2) \equiv \mu \quad \text{then} \quad E(\bar{Y}) = \mu^2$$

Identical distribution

$$\mu = E(1-\bar{X}_i) = \int_0^1 (1-x_i) 4x_i^3 \, dx_i = 0.2$$

So 80% of bad stuff expected to be removed in 1st filtering

$$E(\bar{Y}) = \mu^2 = 0.04$$ so expect only 40% of bad stuff to remain after 2 filtrations

6.a) Suppose $X$ is a discrete rv with possible values $0, 1, 2$... then $E(X) = \sum_{x=0}^{\infty} \rho(x \geq x)$

6.b) If $X$ is a continuous rv with possible values $(0, \infty)$ then

$$E(X) = \int_0^\infty [1 - F_X(x)] \, dx$$
Ex of 6a
I throw a dart at a dartboard repeatedly, trying to get a bullseye (success)
\( \bar{X} = \# \text{ of throws on which I 1st succeed} \)

(Ex throws FFS \( \rightarrow \bar{X} = 3 \) F = failure S = success)

Suppose that my success probability is constant across the throws and equals \( p \), and throws are independent.

Then \( E(\bar{X}) \) should be inversely related to \( p \): The worse I am, the longer I expect the 1st bullseye to take. \( E(\bar{X}) = ? \)

Geometric distribution
At least 1 throw always required so \( P(\bar{X} > 1) = 1 \) for \( n > 1 \)

(at least \( n \) tosses required) \( \Rightarrow \) (none of the 1st \( n-1 \) throws succeeded)

so \( P(\bar{X} > n) = (1-p)^n-1 \) and

\[
E(\bar{X}) = \sum_{n=1}^{\infty} (1-p)^{n-1} = 1 + (1-p) + (1-p)^2 + \ldots = \frac{1}{1-(1-p)} = \frac{1}{p}
\]

\( \begin{align*}
\text{geometric series} & \quad \text{(inverse relation)}
\end{align*} \)

If I'm terrible (e.g. \( p = .01 \)) I expect to succeed on the \( \frac{1}{.01} = 100 \text{th} \) throw

Variance and Standard Deviation

\[
\begin{bmatrix}
1 \\
2 \\
3 \\
4 \\
5 \\
6 \\
7 \\
8 \\
9 \\
\mu \ (\text{mean})
\end{bmatrix}
\begin{bmatrix}
\$1 \\
\$2 \\
\$3 \\
\$4 \\
\$5 \\
\$6 \\
\$7 \\
\$8 \\
\$9
\end{bmatrix}
\]

mean \( \$4 = \mu \)
\( \bar{X} \) discrete rv, uniform \( \{1, 2, 3\} \)

\[ E(\bar{X}) = \bar{X} = \mu \]

Q: How spread out is the dist. of \( \bar{X} \) around its mean \( \mu \)?

\( (X - \mu) \sim \text{uniform} \{ -3, -2, +5 \} \)

\( \uparrow \) deviation from \( \mu \)

Could try calculating \( E(X - \mu) \), but this is 0 for any rv \( X \) because of cancellation of \( \pm \) and \( \mp \) deviations.

Two different easy fixes:

\[ E|X - \mu| = \text{mean} \]

or

\[ E(\bar{X} - \mu)^2 = \text{variance of rv} \ X \]

\( \text{AAD} \): not used much

\( \text{Variance} \): used constantly

Def: \( \bar{X} \) rv with finite mean \( E(\bar{X}) = \mu \)

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

If \( E(\bar{X}) = \pm \infty \) or \( E(\bar{X}) \) doesn't exist,

\( V(\bar{X}) \) doesn't exist.

One problem with variance:

The units are wrong: if \( X \) is in \$\text{ }, \( V(X) \) is in \$^2\text{.}

Easy fix: standard deviation of \( \bar{X} \) = \( \sqrt{V(\bar{X})} = SD(\bar{X}) \text{.} \)
Consequences of these definitions

1) \( V(X) = E[(X-\mu)^2] = E(X^2 - 2\mu X + \mu^2) \)

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

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

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

So \( V(X) = (\text{expectation of } X^2) - (\text{square of expectation of } X) \)

Toy example

\[ \begin{bmatrix} 1 \\ 2 \\ 9 \end{bmatrix} \]

mean \( \mu = 4 \)

\[ X \sim \text{Uniform} \{1, 2, 9\} \]

\[ E(X) = \frac{1}{3} (1+2+9) = \frac{12}{3} = 4 \]

\[ E(X^2) = \frac{1}{3} (1^2 + 2^2 + 9^2) = \frac{1}{3} (1 + 4 + 81) = \frac{86}{3} \]

\[ E(X-M)^2 = \frac{1}{3} (1-4)^2 + \frac{1}{3} (2-4)^2 + \frac{1}{3} (9-4)^2 = 12.7 \]

So SD(X) = \( \sqrt{12.7} \approx 3.6 \)

This is a reasonable summary of the lengths of the arrows.

2) For any rv \( X \), \( V(X) \geq 0 \)

If \( X \) is bounded, \( V(X) \) exists and is finite.

This is a consequence of Jensen's inequality:

\[ g(x) = x^2 \text{ is convex } \Rightarrow E(X^2) \geq [E(X)]^2 \]

i.e. \( V(X) = E(X^2) - [E(X)]^2 \geq 0 \)
3) \( \textsf{V}(X) = 0 \iff P(\bar{X} = c) = 1 \) for some constant \( c \) (this is a trivial rv)

**Notation**

In the same way that by convention \( \textsf{E}(X) = \mu_X \)

\( \textsf{V}(X) = \sigma_X^2 \) and \( \textsf{SD}(X) = \sigma_X \)

4) \( \bar{X} = aX + b \Rightarrow \textsf{V}(\bar{X}) = a^2 \textsf{V}(X) = a^2 \sigma_X^2 \)

and \( \textsf{SD}(\bar{X}) = |a| \sigma_X \) (for any constants \( a, b \))

\[
\begin{align*}
\textsf{V}(\bar{X}) &= \textsf{V}(aX + b) \\
&= a^2 \textsf{V}(X) \\
\textsf{SD}(aX + b) &= |a| \textsf{SD}(X)
\end{align*}
\]

**Special cases**

\( a = 1 \):

\( \textsf{V}(\bar{X} + C) = \textsf{V}(\bar{X}) \)

\( \textsf{SD}(\bar{X} + C) = \textsf{SD}(\bar{X}) \)

\( a = 0 \): \( \textsf{V}(aX) = a^2 \textsf{V}(X) \)

\( b = 0 \): \( \textsf{SD}(aX) = |a| \textsf{SD}(X) \)

5) If \( X_1, \ldots, X_n \) are independent rv with finite means,

\[
\textsf{V}\left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \textsf{V}(X_i)
\]

This is why the concept of variance has endured even though the units of the variance are wrong: for independent rv's, variance is additive whereas \( \textsf{SD} \) is not.
\(X_1, X_2\) independent \(\Rightarrow \quad \text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2)\)

\[
\sqrt{\text{Var}(X_1 + X_2)} = \sqrt{\text{Var}(X_1)} + \sqrt{\text{Var}(X_2)}
\]

\[
\text{SD}(X_1 + X_2) = \sqrt{\text{SD}(X_1)^2 + \text{SD}(X_2)^2}
\]

SD grows like the hypotenuse of a right triangle.

Immediately, \(\max\left\{ \frac{\text{SD}(X_1)}{\text{SD}(X_2)} \right\} < \frac{\text{SD}(X_1 + X_2)}{\text{SD}(X_2)} \)

\[
\text{Consequence of } S
\]

\(X_1, \ldots, X_n\) independent \(r.v.\), \(a_1, \ldots, a_n, b\) constants \(\Rightarrow \)

\[
\text{Var}\left(\sum_{i=1}^{n} a_i X_i + b\right) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i)
\]

\(\text{Ex: } X \sim \text{Binomial}(n, p)\)

We know that \(\text{E}(X) = np\)

What about \(\text{Var}(X)\) and \(\text{SD}(X)\)?

Let \(S_i = \begin{cases} 1 & \text{if success on } i\text{th success/failure trial} \\ 0 & \text{else} \end{cases}\)

for \((i = 1, \ldots, n)\) and suppose as usual that \(S_1, \ldots, S_n\) are IID Bernoulli\((p)\) \(\Rightarrow \) then \(X = \sum_{i=1}^{n} S_i\) and we can work out its variance without difficulty

\[
\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{n} S_i\right) = \sum_{i=1}^{n} \text{Var}(S_i)
\]

independence

So we need to work out the variance of a Bernoulli\((p)\) rv. We already know that \(\text{E}(S_i) = p\), so if we use the formula \(\text{Var}(S_i) = \text{E}(S_i^2) - [\text{E}(S_i)]^2\)

we're halfway there.
Bernoulli rv's are funny:

\[ S_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } (1-p) \end{cases} \]

so \[ S_i^2 = \begin{cases} 1^2 = 1 & \text{with probability } p \\ 0^2 = 0 & \text{with probability } (1-p) \end{cases} \]

so \[ E(S_i^2) = E(S_i) = p \]

\[ V(S_i) = E(S_i^2) - [E(S_i)]^2 = p - p^2 = p(1-p) \]

\[ V(S) = \sum_{i=1}^{n} V(S_i) = \sum_{i=1}^{n} p(1-p) = np(1-p) \]

\[ \text{SD}(S) = \sqrt{np(1-p)} \]

Ex: T-S disease

\[ E = \text{(# of babies in family of } n = 5 \text{ both parents carriers so } p = p(\text{T-S baby}) = \frac{1}{2}) \]

\[ \sim \text{Binomial (n, p) = Binomial (5, } \frac{1}{2}) \]

We already worked out that \( E(E) = np = 1.25 \)

Now \[ \text{SD}(E) = \sqrt{np(1-p)} = \sqrt{5 \left( \frac{1}{2} \right) \left( \frac{3}{2} \right)} = 0.97 \approx 1 \]

It's useful to summarize this by saying "The # of T-S babies this couple will have will be around 1.25, give or take about 1." \( \sim 0.97 \)
How do you measure the spread of a distribution if the variance doesn't exist?

**Ex:** Standard Cauchy 
\[ f_{\mathcal{C}}(x) = \frac{1}{\pi(1+x^2)} \text{ for all } -\infty < x < \infty \]

We saw that \( E(\mathcal{C}) \) doesn't exist, so \( V(\mathcal{C}) \) doesn't exist either.

But we can use the idea of quantiles on any dist. whether its variance exists or not.

We defined the interquartile range (IQR) as 
\[ \text{IQR} = F_{\mathcal{C}}^{-1}(0.75) - F_{\mathcal{C}}^{-1}(0.25) \]

For the standard Cauchy, \( CDF \) is 
\[ F_{\mathcal{C}}(x) = \int_{-\infty}^{x} \frac{1}{\pi(1+t^2)} \, dt = \frac{1}{2} + \frac{\tan^{-1}(t)}{\pi} \]

Need to solve \( F_{\mathcal{C}}(x) = \frac{1}{2} + \frac{\tan^{-1}(x)}{\pi} = \rho \) for \( x \)

Result is \( x = F_{\mathcal{C}}^{-1}(\rho) = \tan\left(\frac{\rho - \frac{1}{2}}{\pi}\right) = -\cot\left(\frac{\rho \pi}{2}\right) \)

So the IQR for the standard Cauchy dist. is
\[ \text{IQR} = F_{\mathcal{C}}^{-1}(\frac{3}{4}) - F_{\mathcal{C}}^{-1}(\frac{1}{4}) = +\tan\left(\frac{\pi}{4}\right) - -\tan\left(-\frac{\pi}{4}\right) = 2 \]

Standard Cauchy PDF
Moments of a rv

\[ E(X) = E(X') \]

\[ V(X) = E(X^2) - [E(X')]^2 = E(X - \mu)^2 \]

With the usual mathematical imprise to generalize

Def: \( X \) rv, \( k \) integer \( \geq 1 \) \( \Rightarrow \) the \( k \)th moment of \( X \)

Of course \( E(X^k) \) may not exist, and if it does it may be infinite, but the idea is still useful.

You can show that \( \text{(kth moment)} \) \( \Rightarrow \) \( E(|X|^k) < \infty \)

Consequences of the moment definition

1) If \( E(|X|^k) < \infty \) for some integer \( k \geq 1 \), then \( E(|X|^j) < \infty \) for all integers \( j \leq k \).

2) If the \( k \)th moment of \( X \) exists, so do the \( (k-1) \), \( (k-2) \), ..., \( \mu \) moments.

Def: \( X \) rv with expectation \( E(X) = \mu \), \( k \) integer \( \geq 1 \)

\( E[(X-\mu)^k] \) is called the \( k \)th central moment of \( X \) or the \( k \)th moment of \( X \) around its mean.

This generalizes the variance of \( X = E((X - \mu)^2) \)

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

\( \Rightarrow \) every rv has 1st central moment 0.
If the dist. of $X$ is symmetric around $\mu$, then $E[(X-\mu)^k] = 0$ for all odd integers $k$ for which $E[(X-\mu)^k]$ exists.

**Def:** $X$ rv with mean $\mu$ so $\mu = \bar{X}$

If the third moment of $X$ exists and is finite, then skewness $(X) = E\left(\frac{X-\mu}{\sigma_X}\right)^3$ converting $\bar{X}$ to standard units.

All symmetric distributions with finite 3rd moment have skewness 0.

---

Moment generalizing functions

**Def:** $X$ rv, $t$: real 

$\psi_{X}(t) = E(e^{tX})$ is called the moment generating function of $X$ (MGF).

**Thm:** $X$ rv with MGF $\psi_{X}(t)$, finite for all values of $t$ in an open interval $(-a, b)$ around 0 ($a > 0, b > 0$), then for all integers $n \geq 0$ $E(X^n) = \frac{d^n}{dt^n} \psi_{X}(t)\bigg|_{t=0}$ is the $n$th derivative of $\psi_{X}$ evaluated at $t = 0$. 
If its premise is satisfied and the calculations are manageable, you get all the moments of $\mathbf{X}$ just by computing $\psi_{\mathbf{X}}(z)$ and differentiating it over and over.