DEF: The *expectation* = *expected value* = *mean* = 1^{st} *moment* of rv x is $E[x]=\bar{x}=\int X f_x(X)dX$ (continuous rv); $\sum X p_x(X)$ (discrete rv).

Note: $f_x(X)$=mass density→ $E[x]$=center of mass. $\frac{2}{3}$ rule for linear $f_x(X)$.

---

1. $E[\cdot]$ is a *linear operator:* $E[ax+by]=aE[x]+bE[y]$ since $E[ax+by]=\int\int(aX+bY)f_{x,y}(X,Y)dX\,dY = a\int\int X f_{x,y}(X,Y)dX\,dY$ $+b\int\int Y f_{x,y}(X,Y)dX\,dY = a\int X f_x(X)dX + b\int Y f_y(Y)dY$. QED.
2. Can have $p_x(E[x])=0$. Consider: Flip a fair coin. x=#heads. E[x]=1/2, but $p_x(E[x])=Pr[x=E[x]]=Pr[1/2\,head]=0!$
3. $E[g(x)]=\int Y f_y(Y)dY=\int g(X)f_x(X)dX$ (note $f_y(Y)dY=f_x(X)dX$).

Note: $E[g(x)]\neq g(E[x])$, e.g., $E[\frac{1}{x}]\neq 1/E[x]$! *Very* common mistake!

4. $Pr[x>E[x]]\neq Pr[x<E[x]]$ unless $f_x(X)$ symmetric about $X=E[x]$.

DEF: *Median* m of rv x is m s.t. $F_x(m)=Pr[x\leq m]=Pr[x>m]=\frac{1}{2}$.

Conditional Expectations:

1. $E[x|A]=\int X f_{x|A}(X|A)dX=\int\int_A X\frac{f_{x,y}(X,Y)}{Pr[A]}dX\,dY$ is a *number.*
2. $E[x|y=Y]=\int X f_{x|y}(X|Y)dX$ is a *function of* Y.
3. $E[x]=\int E[x|y=Y]f_y(Y)dY$ since $=\int\int X f_{x|y}(X|Y)dX\,f_y(Y)dY$ $=\int\int X f_{x,y}(X,Y)dX\,dY=\int X f_x(X)dX$. *Iterated* expectation.

EX: $f_{x,y}(X,Y)=2$ if $1<X<Y<2$; else 0. $A=\{(X,Y):X+Y<3\}$.

E[x]: $f_x(X)=\int f_{x,y}(X,Y)dY=\int_X^2 2dY=2(2-X)$ for $1<X<2$; else 0.
$E[x]=\int_1^2 dX\,X\int_X^2 dY\,2=\int_1^2 X2(2-X)dX=\frac{4}{3}$. (Use $\frac{2}{3}$ rule.)

$E[x|A]$**:** $Pr[A]=\int_1^{1.5}dX\int_X^{3-X}dY\,2=\int_1^{1.5}2(3-2X)dX=\frac{1}{2}$. (Symmetry.)
$f_{x,y|A}(X,Y|A)=2/\frac{1}{2}=4$ if $1<X<1.5\&X<Y<3-X$; else 0.
$f_{x|A}(X|A)=\int_X^{3-X}dY\,4=4(3-2X)$ if $1<X<1.5$; 0 otherwise.
$E[x|A]=\int X f_{x|A}(X|A)=\int_1^{1.5}X4(3-2X)dX=\frac{7}{6}$. (Use $\frac{2}{3}$ rule.)

$E[y|x=X]$**:** $f_{y|x}(Y|X)=\frac{f_{x,y}(X,Y)}{f_x(X)}=\frac{2}{2(2-X)}=\frac{1}{2-X}$ for $1<X<Y<2$; else 0.
$E[y|x=X]=\frac{1}{2}(X+2)$ for $1<X<2$ since $f_{y|x}(Y|X)$ is *uniform*!

E[y]= $E_x[E[y|x=X]]=E_x[\frac{1}{2}(X+2)]=\frac{1}{2}E[x]+1=\frac{1}{2}\frac{4}{3}+1=\frac{5}{3}$. ($\frac{2}{3}$ rule.)

DEF: *variance* = $(standard\, deviation)^2 = 2^{nd}$ *central moment* of rv x is
$\sigma_x^2 = Var[x] = E[(x - E[x])^2] = E[x^2] - (E[x])^2$.

Note: $f_x(X)$=mass density→ σ_x^2=moment of inertia.
σ_x=standard deviation→radius of gyration.

---

1. Translation-invariant: $\sigma_{x+b}^2 = \sigma_x^2$ since $f_{x+b}(X) = f_x(X - b)$ (shift).
2. Scaling: $\sigma_{ax}^2 = a^2\sigma_x^2$ since $E[(ax)^2] - (E[ax])^2 = a^2(E[x^2] - (E[x])^2)$.
3. $\sigma_x^2 \geq 0$; $\sigma_x^2 = 0 \to Pr[x = constant] = 1$. (The constant=E[x].)

---

4. Variance is NOT linear: $\sigma_{x+y}^2 \neq \sigma_x^2 + \sigma_y^2$ in general:
$\sigma_{x+y}^2 = E[(x + y)^2] - (E[x + y])^2 = E[x^2] + E[y^2] + 2E[xy]$
$-(E[x])^2 - (E[y])^2 - 2E[x]E[y] = \sigma_x^2 + \sigma_y^2 + 2(E[xy] - E[x]E[y])$.

   Necessary and sufficient: $x, y\, uncorrelated \leftrightarrow E[xy] = E[x]E[y]$.

Note: $x, y\, independent \to x, y\, uncorrelated$; converse NOT true:

Indpt: $E[xy] = \int\int XY f_{x,y}(X, Y)dX\, dY = \int X f_x(X)dX \int Y f_y(Y)dY$.

EX1: $f_x(X) = \frac{1}{b-a}, a < X < b$; 0 otherwise. Then $\sigma_x^2 = (b - a)^2/12$.
$f_y(Y) = 1, |Y| < \frac{1}{2}$. $E[y^2] = \int_{-1/2}^{1/2} Y^2 dY = \frac{1}{12}; E[y] = 0$. Shift & scale

EX2: x, y uncorrelated→ $\sigma_{3x-2y}^2 = 9\sigma_x^2 + 4\sigma_y^2$, NOT $9\sigma_x^2 - 4\sigma_y^2$!

EX3: Continue example from "Expectation" (overleaf):
$E[x^2|A] = \int_1^{1.5} X^2 4(3 - 2X)dX = \frac{11}{8}$. $\sigma_{x|A}^2 = \frac{11}{8} - (\frac{7}{6})^2 = \frac{1}{72}$.
$\sigma_{y|x=X}^2 = (2 - X)^2/12$ since $f_{y|x}(Y|X)$ uniform. Note $\sigma_{y|X=2}^2 = 0$!

Flip fair coin. If heads, use $f_{x_1}(X)$; if tails, use $f_{x_2}(X)$ for $f_x(X)$.

1. Let x be resulting rv. Then $f_x(X) = \frac{1}{2}f_{x_1}(X) + \frac{1}{2}f_{x_2}(X)$.
2. $E[x^n] = \frac{1}{2}E[x_1^n] + \frac{1}{2}E[x_2^n]$: n^{th} *moments* →probabilistic choice OK.
3. $\sigma_x^2 \neq \frac{1}{2}\sigma_{x_1}^2 + \frac{1}{2}\sigma_{x_2}^2$: *central* moments→ probabilistic choice NOT OK.

EX: Heads→ $f_{x_1}(X) = 1, 0 < X < 1$; tails→ $f_{x_2}(X) = 1, 1 < X < 2$.

Easy to compute $E[x_1] = \frac{1}{2}; E[x_2] = \frac{3}{2}; \sigma_{x_1}^2 = \sigma_{x_2}^2 = \frac{1}{12}$ (uniform)
$f_x(X) = \frac{1}{2}, 0 < X < 2$. $E[x] = 1 = \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{3}{2}$ but $\sigma_x^2 = \frac{1}{3} \neq \frac{1}{2}\frac{1}{12} + \frac{1}{2}\frac{1}{12}$.

Correct way: $E[x^2] = \frac{1}{2}\int_0^1 X^2 dX + \frac{1}{2}\int_1^2 X^2 dX = \frac{4}{3}$. $\sigma_x^2 = \frac{4}{3} - 1^2 = \frac{1}{3}$.

DEF: The $covariance = cross - covariance$ of two rvs x, y is $\lambda_{xy} = Cov[x,y] = E[(x-E[x])(y-E[y])] = E[xy] - E[x]E[y]$.

Note: $f_x(X)$=mass density$\rightarrow \lambda_{xy}$=cross or mixed moment of inertia.

---

1. $\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y + 2\lambda_{xy}$. Also, $\lambda_{xx} = Cov[x,x] = \sigma^2_x \geq 0$.
2. x, y independent$\rightarrow x, y$ uncorrelated $\leftrightarrow \lambda_{xy} = 0 \leftrightarrow \sigma^2_{x+y} = \sigma^2_x + \sigma^2_y$.
3. **Schwarz inequality:** $\lambda^2_{xy} \leq \lambda_{xx}\lambda_{yy} = \sigma^2_x \sigma^2_y$.

Proof: $E[(x - \frac{E[xy]}{E[y^2]}y)^2] = E[x^2] - \frac{E[xy]^2}{E[y^2]} \geq 0$ (with equality iff $x = \frac{E[xy]}{E[y^2]}y$) $\rightarrow E[xy]^2 \leq E[x^2]E[y^2]$. Set $x \rightarrow x - E[x]$ and $y \rightarrow y - E[y]$. QED.

DEF: *Normalized* rv $\tilde{x}$ associated with rv x is $\tilde{x} = \frac{x-E[x]}{\sigma_x}$.

Props: Normalized $\tilde{x}$ is dimensionless with $E[\tilde{x}] = 0$ and $\sigma_{\tilde{x}} = 1$.

DEF: *Correlation coefficient*$=\rho_{xy} = \lambda_{\tilde{x}\tilde{y}} = \frac{\lambda_{xy}}{\sigma_x \sigma_y} = E\left[\left(\frac{x-E[x]}{\sigma_x}\right)\left(\frac{y-E[y]}{\sigma_y}\right)\right]$.

Props: $|\rho_{xy}| \leq 1$ from Schwarz inequality. x, y uncorrelated$\leftrightarrow \rho_{xy} = 0$. x, y perfectly (anti)correlated$\leftrightarrow \rho_{xy} = \pm 1 \leftrightarrow \frac{x-E[x]}{\sigma_x} = \pm\frac{y-E[y]}{\sigma_y}$.

DEF: The *correlation* between rvs x, y is $E[xy]$. x, y 0-mean$\rightarrow E[xy] = \lambda_{xy}$.

DEF: x, y are *orthogonal* if $E[xy] = 0$. Unscramble the nomenclature:

1. x, y are *uncorrelated* iff their *covariance*=0. (I didn't name these!)
2. x, y are NOT *uncorrelated* if their *correlation*=0, unless 0-mean.
3. x, y are *orthogonal* if their *correlation*=0 and they are 0-mean.

PROBABILISTIC INEQUALITIES

1. **Union bound:** $Pr[\cup_{i=1}^N A_i] \leq \sum_{i=1}^N Pr[A_i]$ (see text p. 11).

Proof: By induction. $Pr[\cup_{i=1}^{N+1} A_i] = Pr[\cup_{i=1}^N A_i] + Pr[A_{N+1}] - Pr[(\cup_{i=1}^N A_i) \cap A_{N+1}] \leq Pr[\cup_{i=1}^N A_i] + Pr[A_{N+1}] \leq \sum_{i=1}^N A_i + Pr[A_{N+1}] = \sum_{i=1}^{N+1} Pr[A_i]$.

2. **Markov inequality:** $Pr[x > a] \leq E[x]/a$ if $x \geq 0, a > 0, E[x] < \infty$.

Proof: $E[x] = \int_0^\infty X f_x(X) dX \geq \int_a^\infty X f_x(X) dX \geq \int_a^\infty a f_x(X) = aPr[x > a]$.

3. **Chebyschev:** $Pr[|x - E[x]| > \epsilon] \leq \frac{\sigma^2_x}{\epsilon^2}$ if $\epsilon > 0$ and $E[x], \sigma^2_x < \infty$.

Proof: Apply Markov inequality to $(x - E[x])^2$ with $a = \epsilon^2$. Very loose bound.

4. **Chernoff:** $Pr[x > a] \leq E[e^{s(x-a)}]$ for any a and any $s > 0$.

Proof: $Pr[x > a] = Pr[e^{sx} > e^{sa}] \leq E[e^{sx}]/e^{sa} = E[e^{s(x-a)}]$ using Markov.