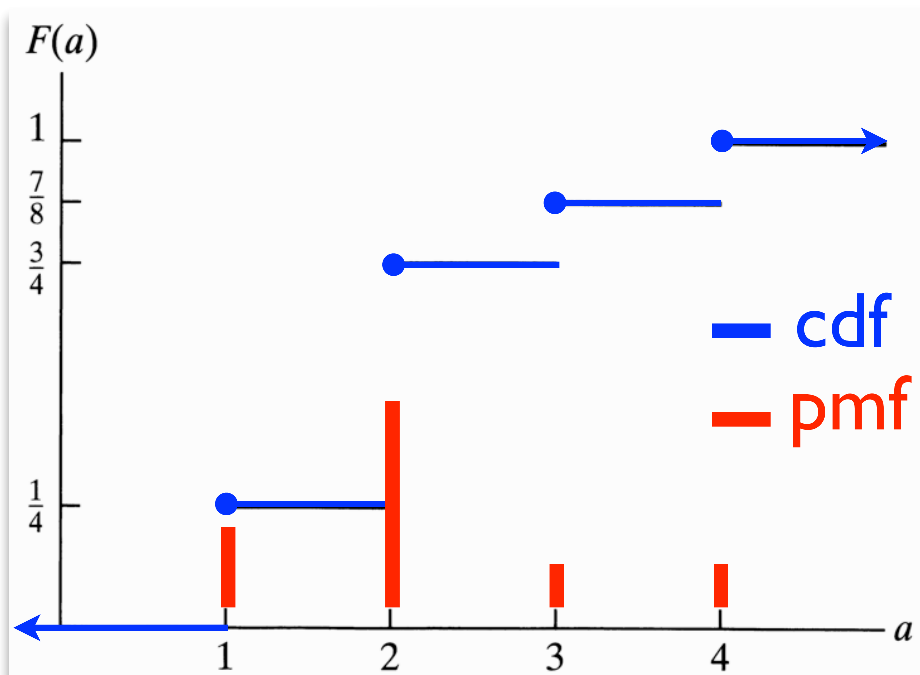
The *cumulative distribution function* for a random variable X is the function $F: \mathcal{R} \rightarrow [0,1]$ defined by

$$\boxed{F(a) = P[X \leq a]}$$

Ex: if X has probability mass function given by:

$$p(1) = \frac{1}{4} \quad p(2) = \frac{1}{2} \quad p(3) = \frac{1}{8} \quad p(4) = \frac{1}{8}$$

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \leq a < 2 \\ \frac{3}{4} & 2 \leq a < 3 \\ \frac{7}{8} & 3 \leq a < 4 \\ 1 & 4 \leq a \end{cases}$$



NB: for discrete random variables, be careful about "≤" vs "<"

expectation

For a discrete r.v. X with p.m.f. p(•), the *expectation of X*, aka *expected value* or *mean*, is

$$\boxed{E[X] = \Sigma_x\, xp(x)}$$

average of random values, *weighted* by their respective probabilities

For the equally-likely outcomes case, this is just the average of the possible random values of X

For *un*equally-likely outcomes, it is again the average of the possible random values of X, weighted by their respective probabilities

Ex 1: Let X = value seen rolling a fair die $p(1), p(2), ..., p(6) = 1/6$

$$E[X] = \sum_{i=1}^{6} ip(i) = \frac{1}{6}(1+2+\cdots+6) = \frac{21}{6} = 3.5$$

Ex 2: Coin flip; X = +1 if H (win $1), -1 if T (lose $1)

$$E[X] = (+1)\bullet p(+1) + (-1)\bullet p(-1) = 1\bullet(1/2) + (-1)\bullet(1/2) = 0$$

Linearity of expectation, I

For any constants *a*, *b*: $\boxed{E[aX + b] = aE[X] + b}$

Proof:

$$
\begin{aligned}
E[aX + b] &= \sum_x (ax + b) \cdot p(x) \\
&= a \sum_x xp(x) + b \sum_x p(x) \\
&= aE[X] + b
\end{aligned}
$$

A & *B* each bet \$1, then flip 2 coins:

HH	*A* wins \$2
HT	Each takes back \$1
TH	
TT	*B* wins \$2

Let X = *A*'s net gain: +1, 0, -1, resp.:

$P(X = +1) = 1/4$
$P(X = 0) = 1/2$
$P(X = -1) = 1/4$

What is $E[X]$?

$$E[X] = 1 \cdot 1/4 + 0 \cdot 1/2 + (-1) \cdot 1/4 = 0$$

What is $E[X^2]$?

$$E[X^2] = 1^2 \cdot 1/4 + 0^2 \cdot 1/2 + (-1)^2 \cdot 1/4 = 1/2$$

What is $E[2X+1]$?

$$E[2X + 1] = 2E[X] + 1 = 2 \cdot 0 + 1 = 1$$

From slide 20

Note:

Linearity is special!

It is *not* true in general that

$$E[X \cdot Y] = E[X] \cdot E[Y]$$

$$E[X^2] = E[X]^2$$ ← counterexample above

$$E[X/Y] = E[X] / E[Y]$$

$$E[\operatorname{asinh}(X)] = \operatorname{asinh}(E[X])$$

•

•

•

variance

The *variance* of a random variable X with mean $E[X] = \mu$ is $Var[X] = E[(X-\mu)^2]$, often denoted σ^2.

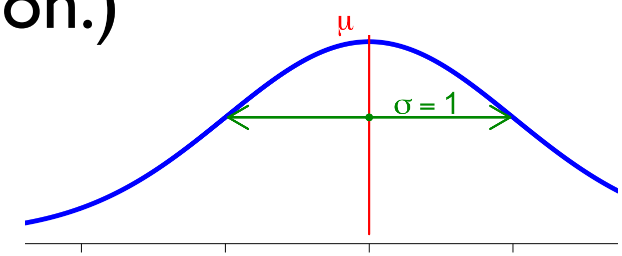
I: Square always ≥ 0, and exaggerated as X moves away from μ, so $Var[X]$ emphasizes *deviation* from the mean.

II: Numbers vary a lot depending on exact distribution of X, but it is common that X is

within $\mu \pm \sigma$ ~66% of the time, and

within $\mu \pm 2\sigma$ ~95% of the time.

(We'll see the reasons for this soon.)



$$\boxed{\mathrm{Var}[aX+b] = a^2\,\mathrm{Var}[X]}$$

NOT linear;
insensitive to location (b),
quadratic in scale (a)

$$\begin{aligned}\mathrm{Var}(aX+b) &= E[(aX+b-a\mu-b)^2]\\ &= E[a^2(X-\mu)^2]\\ &= a^2E[(X-\mu)^2]\\ &= a^2\mathrm{Var}(X)\end{aligned}$$

Ex:

$$X = \begin{cases} +1 & \text{if Heads} \\ -1 & \text{if Tails}\end{cases}$$

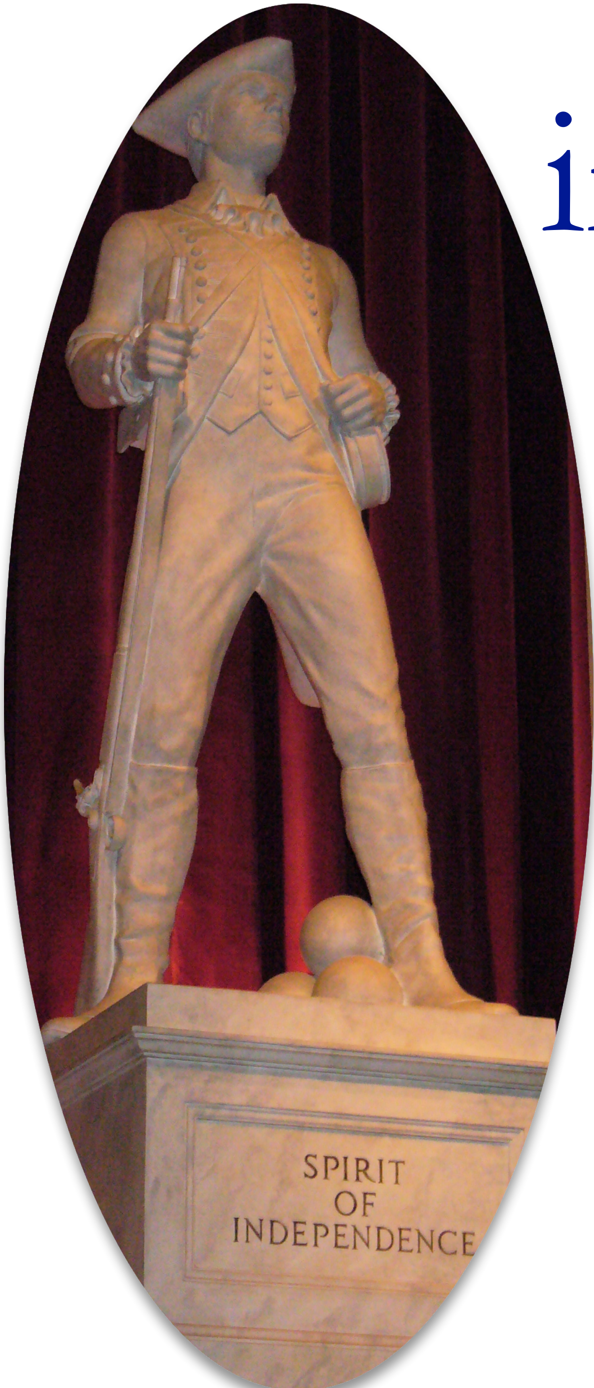
E[X] = 0
Var[X] = 1

$$Y = \begin{cases} +1000 & \text{if Heads} \\ -1000 & \text{if Tails}\end{cases}$$

Y = 1000 X

E[Y] = E[1000 X] = 1000 E[X] = 0

Var[Y] = Var[10^3 X]=10^6Var[X] = 10^6



independence

and

joint

distributions

(Bienaymé, 1853)

Theorem: If X & Y are *independent*, (any dist, not just binomial) then

$$Var[X+Y] = Var[X]+Var[Y]$$

Alternate Proof:

$$
\begin{aligned}
&Var[X+Y] \\
&= E[(X+Y)^2] - (E[X+Y])^2 \\
&= E[X^2+2XY+Y^2] - (E[X]+E[Y])^2 \\
&= E[X^2] + 2E[XY] + E[Y^2] - ((E[X])^2 + 2E[X]E[Y] + (E[Y])^2) \\
&= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 + 2(E[XY] - E[X]E[Y]) \\
&= Var[X] + Var[Y] + 2(E[X]E[Y] - E[X]E[Y]) \\
&= Var[X] + Var[Y]
\end{aligned}
$$

slide 60

FYI, the quantity E[XY]-E[X]E[Y] is called the *covariance* of X,Y. As shown, it is 0 if X,Y are independent; if not zero it is a useful measure of their degree of dependence.

Conditional Expectation:

$$E[X \mid A] = \sum_x x \cdot P(X=x \mid A)$$

Law of Total Expectation

$$E[X] = E[X \mid A] \cdot P(A) + E[X \mid \neg A] \cdot P(\neg A)$$

Variance:

$$Var[X] = E[\,(X-E[X])^2\,] = E[X^2] - (E[X])^2]$$

Standard deviation: $\sigma = \sqrt{\mathrm{Var}[X]}$

$\mathrm{Var}[aX+b] = a^2\,\mathrm{Var}[X]$ "Variance is insensitive to location, quadratic in scale"

If X & Y are *independent*, then

$$\left.\begin{array}{l} E[X \cdot Y] = E[X] \cdot E[Y] \\ \mathrm{Var}[X+Y] = \mathrm{Var}[X]+\mathrm{Var}[Y] \end{array}\right\}$$

(These two equalities hold for *indp* rv's; but not in general.)