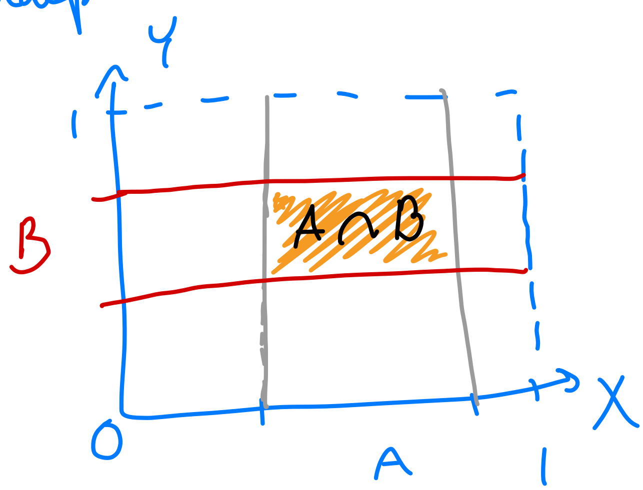
10/4/22

Independence



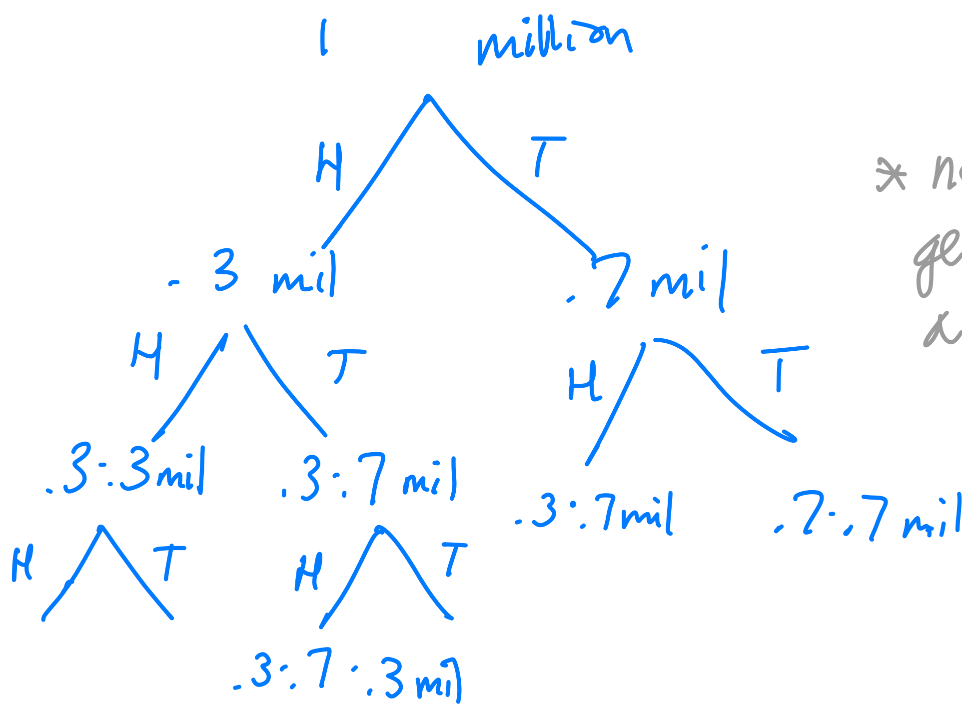
"independent"

$A \perp B \iff P(A \cap B)$

$= |A \cap B| = |A||B| = P(A)P(B)$

ex.) Flip a coin independently $P(\text{head}) = .3$

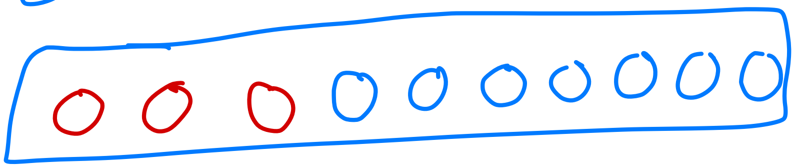
1 million people doing same experiments (repetitions)



* no dependence – getting a H, doesn't affect next flip

ex.) Sampling from a population

3 red + 7 blue



ways to get RRB

$$P(RRB) = .3 \cdot .3 \cdot .7$$

$$= \frac{3 \cdot 3 \cdot 7}{10^3} \quad \cdots \quad \frac{|A|}{|\Omega|}$$

combos of 3 balls

ex.) Flip a coin (p) independently

$$P(HHTTHT) = p \cdot p \cdot (1-p) \cdot (1-p) \cdot p \cdot (1-p)$$

$$= p^3 (1-p)^3$$

$$P(\text{a sequence w/ } k \text{ heads}, n-k \text{ tails}) = p^k (1-p)^{n-k}$$

$$X \sim Bin(n, p)$$

$$P(X=k) = \underbrace{\binom{n}{k}}_{\#\text{ of seq.s w/ } k \text{ heads}} \underbrace{p^k (1-p)^{n-k}}_{\text{prob of each sequence w/ } k \text{ heads}}$$

Binomial Formula

→ correspondance to binary tree

$q = 1 - p$

$$\underbrace{(p+q)}_{=1}^{n} = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k}$$

$$= \sum_{k=0}^{n} P(X=k) = 1$$

$$E(X) = \sum_{x} x\, p(x)$$

change notation

$$= \sum_{k=0}^{n} k\, P(X=k)$$

$$= \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

can skip $k=0$ →

$$= \sum_{k=1}^{n} \frac{n(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k}$$

$$= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} q^{n-k}$$

$n' = n-1$

$k' = k-1$

$$= np \underbrace{\sum_{k'=0}^{n'} \frac{n'!}{k'!(n'-k')!} p^{k'} q^{n'-k'}}_{=1}$$

so $E(X) = np$

$$\mathrm{Var}(X) = E(X^2) - \underbrace{E(X)^2}$$

could change k^2 but cancellation is not simple

$$\mathrm{Var}(X) = \;?\; - \underbrace{(np)^2}$$

prev. proof

$$E(X(X-1)) = \sum_{k=0}^{n} k(k-1)\frac{n!}{k!\,(n-k)!}p^k q^{n-k}$$

$$= \sum_{k=2}^{n} \frac{n(n-1)(n-2)!}{(k-2)!\,(n-k)!}p^2 p^{k-2} q^{n-k}$$

$$= n(n-1)p^2 \underbrace{\sum_{k=2}^{n} \frac{\overset{n''}{(n-2)}!}{(k-2)!\,(n-k)!}p^{\overset{k''}{k-2}} q^{n-k}}_{=1}$$

$$= n(n-1)p^2$$

$$E(x^2 - x) = E(x^2) - E(X) = n(n-1)p^2$$

$$E(X^2) = n(n-1)p^2 + np$$

$$\mathrm{Var}(X) = E(X^2) + E(X)^2$$

$$= n(n-1)p^2 + np - (np)^2$$

$$= n^2p^2 - np^2 + np - n^2p^2$$

$$\mathrm{Var}(X) = np(1-p)$$

$$X = \sum_{i=1}^{n} Z_i$$

$$Z_i \sim Ber(p)$$

$$E(Z_i) = p \qquad Var(Z_i) = p(1-p)$$

$$E(X) = \sum_{i=1}^{n} E(Z_i) = np$$

$$Var(X) \overset{ind}{=} \sum_{i=1}^{n} Var(Z_i) = np(1-p)$$

$\frac{X}{n}$: frequency / sample proportion

$$E\left(\frac{X}{n}\right) = \frac{E(X)}{n} = \frac{np}{n} = p$$

$$Var\left(\frac{X}{n}\right) = \frac{Var(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n} \xrightarrow{n \to \infty} = 0$$

$$P\left(\left|\frac{X}{n} - p\right| < \varepsilon\right) \longrightarrow 1$$

Sample space Ω

$p = \frac{1}{2}$ $\qquad \Omega = \{H, T\}^n$

$\omega \in \Omega$

sequence

$X(\omega) = \#$ of heads in ω

$$E(X) = \frac{1}{2^n} \sum_{\omega \in \Omega} X(\omega) = \frac{n}{2}$$

# of sequences

$$Var(X) = \frac{1}{2^n} \sum_{\omega \in \Omega} \left(X(\omega) - \frac{n}{2}\right)^2$$

$$P\left(\left|\frac{X}{n} - \frac{1}{2}\right| < \varepsilon\right) = \frac{|A_{n,\varepsilon}|}{2^n} \longrightarrow 1$$

$$A_{n,\varepsilon} = \left\{\omega : \left|\frac{X(\omega)}{n} - \frac{1}{2}\right| < \varepsilon\right\}$$

ex.) Sample from a population of N people

randomly sample n people

$\Omega = \{N^n \text{ sequences}\}$

\# of possible samples

$\omega \in \Omega \qquad X(\omega) = \#$ red balls

$$E(X) = \frac{1}{N^n} \sum_{w} X(w) = np$$

$$P\left(\left|\frac{X}{n} - p\right| < \varepsilon\right) = \frac{|A_{n,\varepsilon}|}{N^n}$$

ex.) Flip a coin (p) independently

T = # of flips until 1^{st} H $\sim$ Geometric (p)

$P(T = k)$

if $k=1$: head

if $k=2$: tail, head

if $k=3$: tail, tail, head

$$P(T=k) = P(\underbrace{T, \dots, T}_{k-1}, H) = (1-p)^{k-1} p$$

· prob. dist.

$$E(T) = \sum_{k=1}^{\infty} k P(T=k) = \frac{1}{p}$$

If $q = 1-p$

$$= \sum_{k=1}^{\infty} k q^{k-1} p$$

$$= p \sum_{k=1}^{\infty} k q^{k-1}$$

$$= p \sum_{k=1}^{\infty} \frac{d}{dq} q^k = p \frac{d}{dq} \sum_{k=1}^{\infty} q^k$$

$$S = 1 + a + a^2 + \cdots + a^m$$

$$aS = a + a^2 + a^3 + \cdots a^{m+1}$$

$$(1-a)S = 1 - a^{m+1}$$

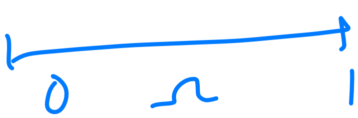
$$S = \frac{1 - a^{m+1}}{1-a} \qquad \overset{m \to \infty}{=} 0$$

$$|a| < 1 \qquad \sum_{k=0}^{\infty} a^k = \frac{1}{1-a}$$

$$E(T) = \sum_{k=1}^{\infty} k P(T = k) = \frac{1}{p}$$

Continuous RVs

Uniform $[0, 1]$



$$U \sim \mathrm{Unif}[0, 1]$$

$$P(U \in A) = \frac{|A|}{|\Omega|} = |A|$$

$$P(U = .3) = 0$$ length of a point $= 0$

$$P(U \in (u, u+\Delta u)) = \Delta u$$

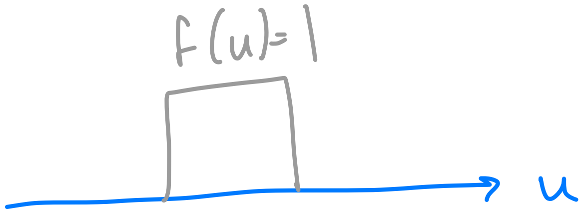
A

$$f(u) = \frac{P(U \in (u, u+\Delta u))}{\Delta u}$$

– probability mass

– length

$$= \frac{\Delta u}{\Delta u} = 1 \qquad u \in [0,1]$$



$$P(U \in (u, u+\Delta u)) = f(u)\,\Delta u$$

basic event statement

Idealization of discrete

large number

$$X \sim \text{Unif}\{0, 1, \dots, M-1\}$$

$$U = \frac{X}{M} \in [0,1]$$

linear congruential method $\longrightarrow$ computer

X_0 seed carefully chosen integers

iterate

$$X_{t+1} = (\underset{7^5}{a} X_t + \underset{0}{b}) \bmod \underset{2^{31}-1}{M}$$

$$X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \ldots$$

$$U_t = \frac{X_t}{M} \overset{ind}{\sim} \mathrm{Unif}[0,1]$$ pseudo-random

$$X \sim f(x)$$

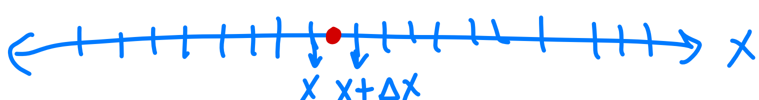
basic event

$$f(x) = \frac{P(X \in (x, x+\Delta x))}{\Delta x}$$

$$\Delta x \longrightarrow 0$$

$$P(X \in (x, x+\Delta x)) = f(x)\,\Delta x$$

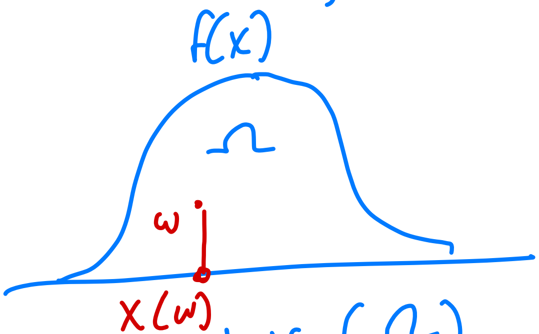
divide into small bins

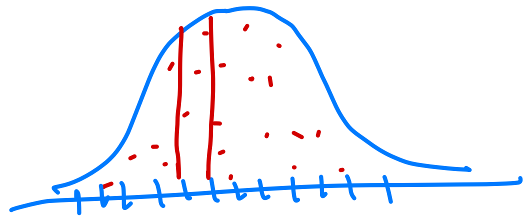


* recall discrete $P(X = x) = p(x)$

$$E(X) = \begin{cases} \sum_x x\, P(X=x) & \ldots \text{discrete} \\ \sum_{x \text{ bins}} x\, P(X \in (x, x+\Delta x)) & \\ = \sum_{\text{bins}} x f(x) \Delta x \xrightarrow[0]{\Delta x} \int x f(x)\, dx & \ldots \text{continuous} \end{cases}$$

$$E(h(X)) = \int h(x)\, f(x)\, dx$$





$\omega \sim \mathrm{Unif}(\Omega)$

$X(\omega)$ = horizontal coordinate $\sim f(x)$

n points

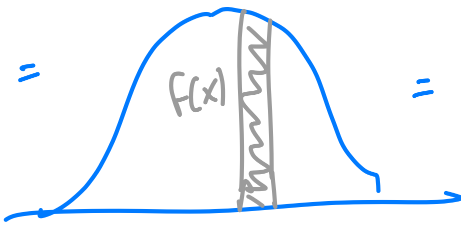
$n(x)$ = # of points in $(x, x+\Delta x)$

normalize

$$\frac{n(x)/n}{\Delta x} \to f(x)$$

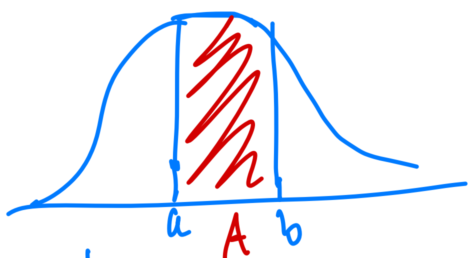
$$P(X \in (x, x+\Delta x))$$

$=$



$= f(x)\,\Delta x$

$P(X \in A) =$



$$= \int_a^b f(x)\, dx$$

$$\frac{1}{n} \sum_{\text{bins}} x \cdot n(x) = \sum_{\text{bins}} x \frac{n(x)}{n}$$

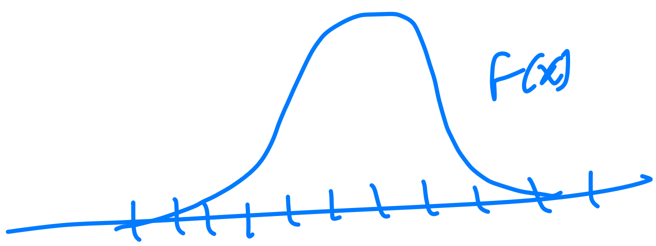
$$\to \int x\, f(x)\, \Delta x = E(X)$$

ex.)

Ω US population

ω = person

$X(\omega)$ = height of ω



6ft 6ft + 1 in

$N(x)$ = # of people in $(x, x+\Delta x)$

$$\frac{N(x)/N}{\Delta x} = f(x)$$

random sampling ω

under random sampling, prob becomes pop. proportion

$$P(X(\omega) \in (x, x+\Delta x)) = \frac{\frac{N(x)}{N}}{\Delta x} = f(x)$$

$$E(X) = \frac{1}{N} \sum_{\omega} X(\omega)$$