Review of probability concepts

Random experiment/process : deterministic prediction is hard.

Example: coin toss with high initial velocity

Random experiments generate simple events

Example: coin tossing twice

events: HH, HT, TH, TT

$S = \{HH, HT, TH, TT\}$ - sample space

An event is any subset space:

$A =$ " at least one H " $= \{HH, HT, TH\}$

Def: If events do not overlap, they are called mutually exclusive (disjoint in terms of set theory)

Note: do not confuse mutually exclusive with independent events

Event arithmetic:

$A \overset{\text{or}}{\cup} B$ = { simple events in A or B, or both }

$A \overset{\text{and}}{\cap} B$ = { simple events in A and B }

complement

A^c = { simple events not in A }

$A = \{HH, HT, TH\}$ $\quad$ $A \cup B = \{HH, HT, TH, TT\}$

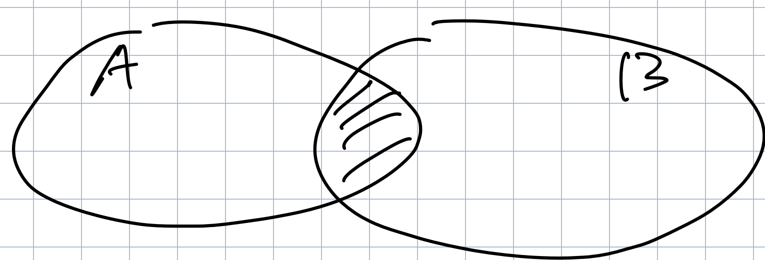
$B = \{HT, TH, TT\}$ $\quad$ $A \cap B = \{HT, TH\}$

$A^c = \{TT\}$

Probability is a function that maps events (subsets of the sample space) into $[0, 1]$.

1) For any $A \subset S$, $0 \le P(A) \le 1$
2) $P(S) = 1$
3) $A \cap B = \underset{\text{empty set}}{\emptyset} \Rightarrow P(A \cup B) = P(A) + Pr(B)$

Addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



Where do probabilities come from?

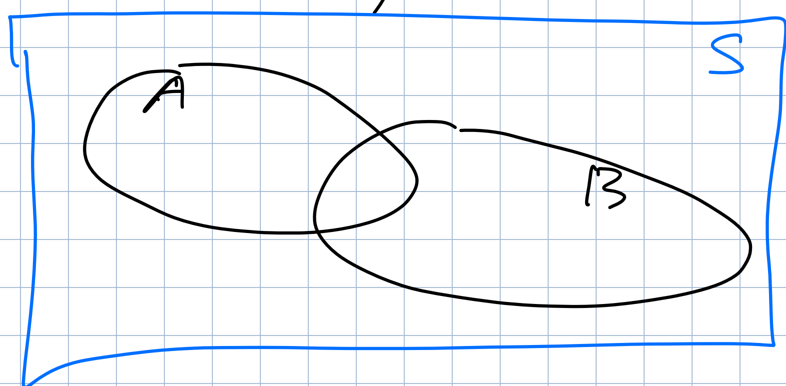
1) long term frequencies $P(A) = \dfrac{\#\text{ of times } A \text{ occurs}}{\#\text{ of r.e. repetitions}}$
2) $P(A)$ - come from some assumed prob mass function or density
3) $P(A)$ - subjective probabilities

Conditional probability

Sometimes we want to think about of A and how it depends on whether or not event B had occured

$Pr($"being struck by lightning"$) <$ $Pr($"being struck by lightning" $|$ "caught in a storm"$)$

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \text{ where } P(B) > 0$$



Example (medical testing)

Suppose we are evaluating results of a medical diagnostic test for a disease D. No test is perfect. This particular test has false negative rate/prob of 1% and fals positive rate/prob of 2%.

$T+$ = "test is positive" $D+$ = "disease is present"

$T-$ = "test is negative" $D-$ = "disease is absent"

$P(T+ \mid D+) = 0.09$ $\qquad$ $P(T+ \mid D-) = 0.02$

$P(T- \mid D+) = 0.01$ $\qquad$ $P(T- \mid D-) = 0.98$

We also assume that $P(D+) = 0.001$ – frequency of the disease in the population (prevalence)

In practice, every patient should be interested in:

$$P(D+ \mid T+) = \frac{P(D+ \cap T+)}{P(T+)}$$

$$P(T+ \mid D+) = \frac{P(T+ \cap D+)}{P(D+)} \Rightarrow$$

$$P(T+ \cap D+) = P(T+ \mid D+) \cdot P(D+)$$

$$P(T+) = P\Big((T+ \overset{\text{and}}{\cap} D+) \overset{\text{or}}{\cup} (T+ \overset{\text{and}}{\cap} D-)\Big) =$$

mutually exclusive

$$= P(T+ \cap D+) + P(T+ \cap D-) =$$

conditional on

$$= \underline{P(T+ \mid D+)} \cdot P(D+) + P(T+ \mid D-) \cdot P(D-) =$$

$$= 0.99 \cdot 0.001 + 0.02 \cdot (1 - 0.001) = 0.02097$$

conditional on

$$P(D+ \mid T+) = \frac{P(T+ \mid D+) \cdot P(D+)}{P(T+)} = \frac{0.99 \cdot 0.001}{0.02097} \approx 0.047$$

Note: if we redo these calculations

with $P(D+) = 0.01$ we get $P(D+|T+) \approx 0.33$

Note: if the patient takes the test twice:

$$P(D+ \mid T+ \cap T+)$$

Note: More practically, this patient would be interested in

$$P(D+ \mid T+ \cap \text{symptoms})$$

Random variables

A random variable is a function mapping sample spase $S \to \mathbb{R}^n$ - n-dimensional space of real numbers

Example: coin flipping

Random variable: # of heads $(g : S \to \mathbb{R})$

Simple events	g (simple event)
HH	2
HT	1
TH	1
TT	0

In practice, we start with a distributional assumption, which defines a sample space and probabilities for all events

Example. X is a discrete random variable with values $S = \{1, 2, 3\}$ with prob mass function : $p(1) = 0.1$

$p(2) = 0.4$

$p(3) = 0.5$

Bernoulli random variable :

$$X = \begin{cases} 1 & \text{if heads} \\ 0 & \text{if tails} \end{cases}$$

$P(X=1) = p$, $P(X=0) = 1-p$

Binomial random variable

$X_1, X_2, \ldots, X_n$ – independent and identically distributed Bernoulli random variables (so p is the same for all Xs)

$$y = \sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$$

$\uparrow$

# of successes in n Bernoulli (coin toss) experiments

prob mass function: $p(k) = \binom{n}{k} p^k (1-p)^{n-k}$

$k = 0, 1, \ldots, \ldots, n$

$010100 \to 2$ successes

$110000 \to 2$ succeses

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

Def $F(x) = P(X \le x)$ — cummulative distribution function (cdf) of X

random variable (X); fixed scalar (x)

Properties:

1) $0 \le F(x) \le 1$

2) $F(x) \le F(y)$ for $x \le y$

Discrete uniform random variable over $\{1, \dots, n\}$ $\quad p(k) = \frac{1}{n}, \; k = 1, \dots, n$

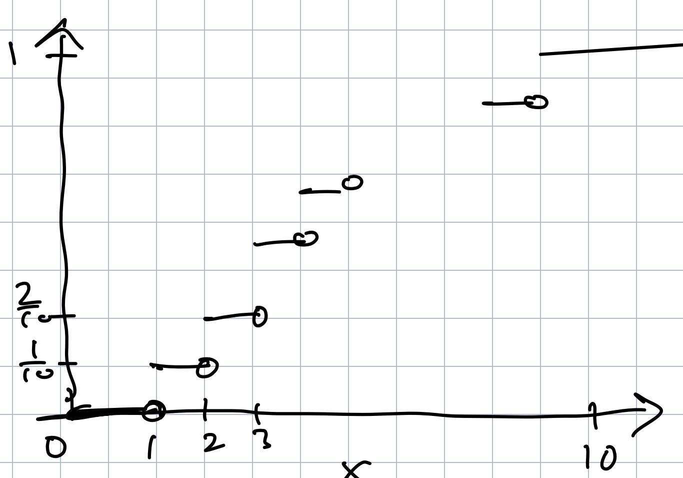
$F(x) = P(U \le x)$

if $x < 1$: $P(U \le x) = 0$

if $1 \le x < 2$ $\quad P(U \le x) = P(U = 1) = \frac{1}{n}$

$2 \le x < 3$ $\quad P(U \le x) = P(U=1) + P(U=2) = \frac{1}{n} + \frac{1}{n}$

$$F(x) = \begin{cases} 0, & \text{if } x < 1 \\ \frac{1}{n}, & \text{if } 1 \le x < 2 \\ \frac{2}{n}, & \text{if } 2 \le x < 3 \\ \vdots & \\ \frac{n-1}{n} & \text{if } n-1 \le x < n \\ 1 & \text{if } x \ge n \end{cases}$$



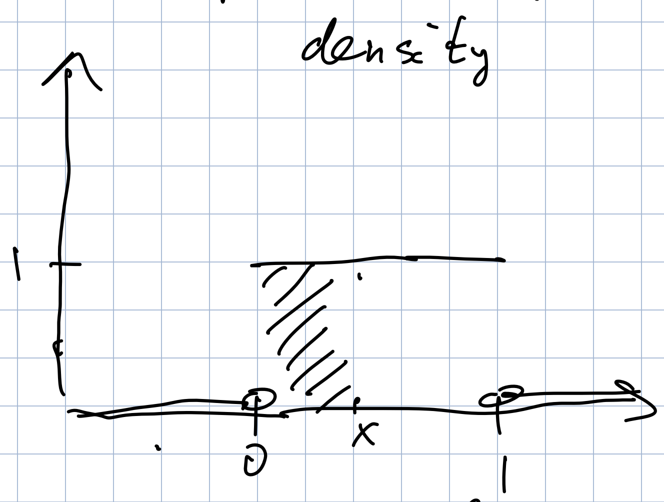
Def If $F(x) = \int_{-\infty}^{x} f(y)\,dy$ for some $f(x) \ge 0$

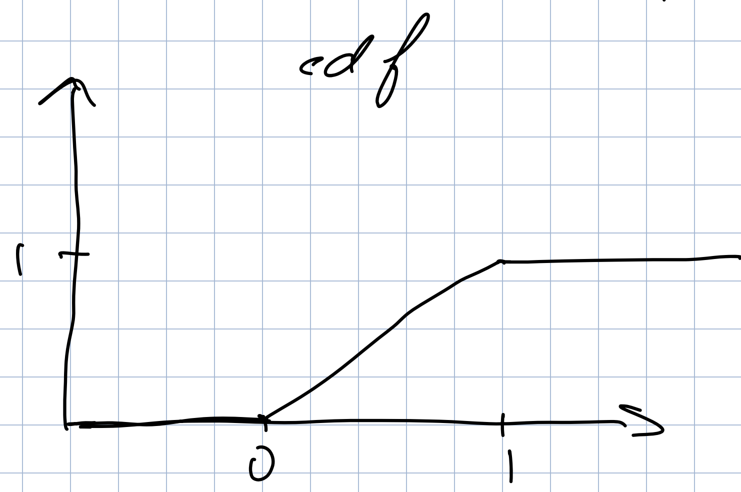
then $f(x)$ — density

Note: $\int_a^b f(x)\,dx = F(b) - F(a) = P(a \le X < b)$

$$\frac{dF(x)}{dx} = f(x)$$

Example: Uniform random variable on $[0,1]$





$$f(x) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

The cdf of U

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$

Inverse CDF method for generating realizations of random variables

Suppose we have a random variable with cdf $F(x)$. We would like to generate realizations of this random variable, but we only have access to a stream of independent Unif $[0,1]$ random variables.

<u>Theorem</u> Let X be a continuous random variable with cdf $F(x)$ such that $F^{-1}(u)$ exists for all $u \in (0,1)$. If $U \sim \text{Unif}[0,1]$, then random variable $F^{-1}(U)$ has the same distribution as X

Proof: since cdf uniquely characterizes the distribution of a random variable, it is enough to show the the cdf of $F^{-1}(U)$ is $F(x)$

recall: $x \le y \Rightarrow F(x) \le F(y)$

$$P(F^{-1}(U) \le x) = P(\cancel{F}(\cancel{F^{-1}}(U)) \le F(x)) =$$

$$= P(U \le \underbrace{F(x)}^{\in [0,1]}) = F(x) \checkmark$$

<u>Example</u>: simulating exponential random variable

Suppose $X \sim \text{Exp}(\lambda)$ with density

$$f(x) = \lambda e^{-\lambda x} \cdot 1_{\{x \ge 0\}}$$

cdf of X is $\int_{-\infty}^{x} f(x)dx = \int_{0}^{x} \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x}$

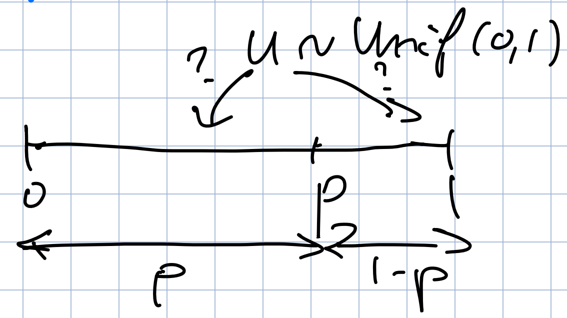
Let $U \sim \text{Unif}[0,1]$. Let's invert the cdf:

$$F(x) = U \iff 1 - e^{-\lambda x} = U \iff 1 - U = e^{-\lambda x} \iff$$

$$\iff \ln(1-U) = -\lambda x \iff x = -\frac{1}{\lambda}\ln(1-U) > 0$$

Inverse CDF for Discrete Random Variables

Bernoulli (p)



$X \sim \text{Discrete}(p_1, \ldots, p_m)$ $\quad \sum_{i=1}^{m} p_i = 1$