Review of probability concepts

Random experiment / process : an experiment or process, where deterministic prediction is hard.

<u>Example</u>: coin toss with high initial velocity

Random experiments generate <u>simple events</u>

<u>Example</u>: tossing a coin twice results in the following events:

HH, HT, TH, TT

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

An event is any subset of S:

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

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

<u>Note</u>: do not confuse mutually exclussive with independence of 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 $\}$

$\overset{\text{complement}}{A^c} = \{$ simple events not in A $\}$

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

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

$A^c = \{TT\}$

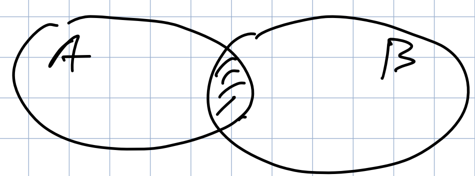
Probability is a function maps events (subsets of the sample space) into $[0,1]$.

1) For any $A \subset S$, $0 \leq P(A) \leq 1$

2) $P(S) =$ empty set

3) $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

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



Where do probabilities come from?

1) long term frequencies: $P(A) = \dfrac{\#\text{ of times } A \text{ occurs}}{\#\text{ of repeats of r.e.}}$

random experiment

2) $P(A)$ - come some assumed prob mass or density functions

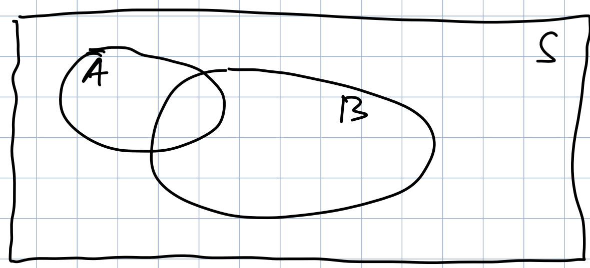
3) $P(A)$ = subjective probabilities

Conditional probability

Sometimes we want to think about prob of A and how it depends on whether or not event B has occurred

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

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



Example (medical testing)

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

In prob/math notation:

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

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

$P(T+ \mid D+) = 0.99 \qquad P(T+ \mid D-) = 0.02$

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

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

In practice, every patient is interested in the following probability:

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

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

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

mutually exclusive

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

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

Back to our calculation:

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

so about 5%

If we redo these calculations with $P(D+) = 0.01$

$$P(D+ \mid T+) \approx 0.33$$

Random variables:

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

Example: coin flipping

Sample space: $S = \{HH, HT, TH, TT\}$

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

Simple event	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

$X_1, X_2, \dots, X_n$ - independent and identically distributed Bernoulli random variable (so p is the same for all $X_1, \dots, X_n$)

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

$\uparrow$

# of successes in n Bernoulli (coin toss) experiments

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

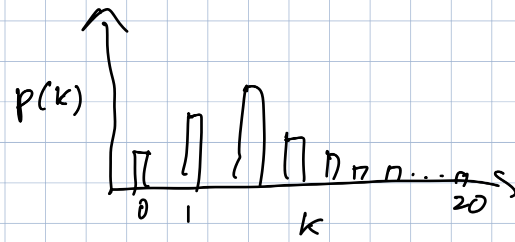
$k = 0, 1, \dots, n$

$010100 \to 2$ successes

$110000 \to 2$ successes

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

Note: prob mass functions are usually visualized via bar plots:



Def: $F(x) = P(X \le x)$ is called a cummulative distribution function (cdf) of X

Properties:

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

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

$\{-\infty, x\} \subset \{-\infty, y\}$

$P(\{-\infty, x\}) \le P(\{-\infty, y\})$

$P(X \le x) \le P(X \le y)$

Discrete uniform random variable

U uniformly distributed over $\{1, \dots, n\}$

$P(U = k) = \frac{1}{n}$ for all $k = 1, \dots, n$

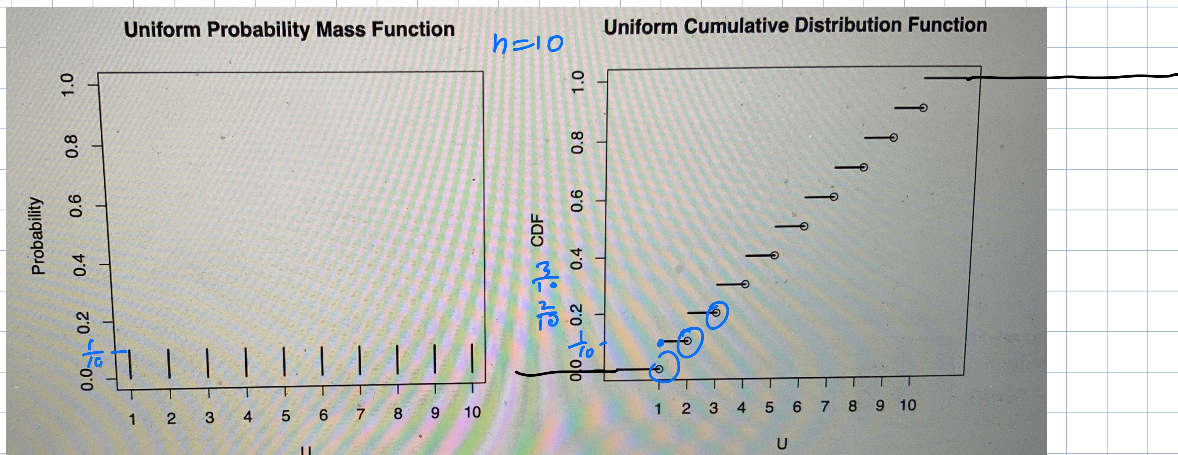
$F(x) = P(U \le x)$ if $x < 1$: $P(U \le x) = 0$

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

$2 \le x < 3$ $P(U \le x) = P(U = 1) +$

$+ P(U = 2) = \frac{1}{n} + \frac{1}{n} = \frac{2}{n}$

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



Def If $F(x) = \int_{-\infty}^{x} f(x)\,dx$ for some $f(x) \geq 0$, then $f(x)$ is called prob. density function of X – continuous random variable

Note: $\int_a^b f(x)\,dx = F(b) - F(a) = Pr(a \leq X < b)$.

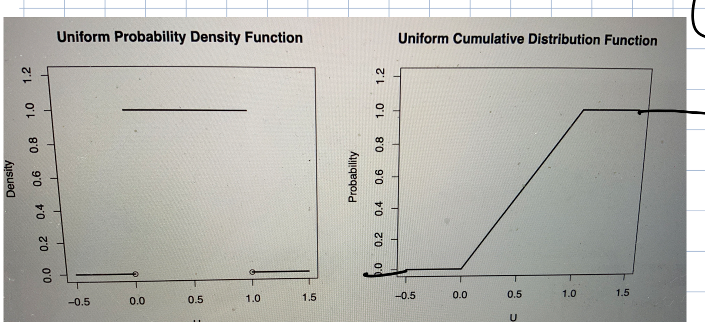
Moreover, $\frac{d}{dx} F(x) = f(x)$

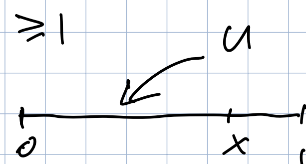
Example Uniform random variable on $[0,1]$

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

The cdf of U

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





$P(U \leq x) = x$

Expectation:

1) for disvrete random variable X

$$E(g(X)) = \sum_{k=1}^{\infty} g(x_k)\, P(X = x_k)$$

1st moment $\quad E(X) = \sum_{k=1}^{\infty} x_k\, P(X = x_k)$

2nd moment $\quad E(X^2) = \sum_{k=1}^{\infty} x_k^2\, P(X = x_k)$

2) for continuous random variable X

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

variance: $\quad g(x) = \boxed{[x - E(x)]^2}$

Example: exponential random variable

Exponential random variable has density

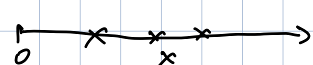
$$f(x) = \lambda e^{-\lambda x} \cdot 1_{\{x \geq 0\}}, \text{ wher } \lambda > 0 \text{ - rate parameter}$$

$X \sim Exp(\lambda)$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} = [\text{integration by parts}] =$$

$$= \ldots = \frac{1}{\lambda}$$

Expectations are linear operators:



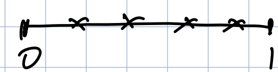
$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i)$$

This does <u>not</u> hold for variances in general, but if $X_1, \ldots, X_n$ are <u>independent</u>, then

$$Var\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 Var(X_i)$$

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.



Theorem Let X be a continuous random variable with cdf $F(x)$ such that $F^{-1}(u)$ exists for all $u \in (0,1)$. If $U \sim$ 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 that the cdf of random variable $F^{-1}(u)$ is $F(x)$.

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

$$P\left(F^{-1}(U) \le x\right) = [F \text{ is non-decreasing}] =$$
$$= P\left(F\left(F^{-1}(u)\right) \le F(x)\right) = P\left(U \le F(x)\right) =$$
$$= [\text{definition of Unif } [0,1] \text{ cdf}] = F(x) \checkmark$$

Example simulating exponential random variable

Suppose $X \sim Exp(\lambda)$ with density $f(x) = \lambda e^{-\lambda x} \cdot \mathbb{1}_{\{x \ge 0\}}$

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

Let $U \sim Unif[0,1]$. 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

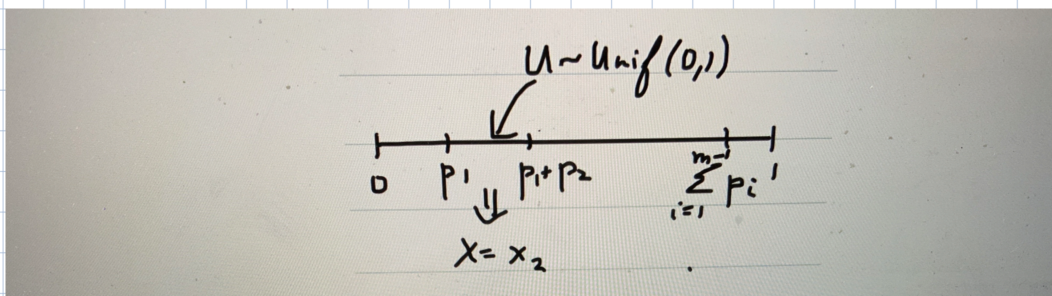


Figure 5: Illustration of the inverse cdf method for discrete random variables.

$$P(X = x_k) = p_k \qquad \sum_{k=1}^{n} p = 1$$

Bernoulli(p)