Singapore Prep Course

Probability, Chapter 1

Roger Lee

Probability

Relevance of probability

- Probability is the branch of mathematics that deals with randomness.
- Financial modeling relies heavily on probability theory, because one of the central goals of finance is to manage risky payoffs.
- Interviewers for finance jobs like to ask probability questions. (including questions not directly applicable to the job)

Counting

Probability

Conditional Probability

Random Variables – Discrete

Random Variables – Continuous

Multiplication rule

- Multiplication rule: if an experiment consists of J stages, and stage j has n_j possible outcomes regardless of the outcomes of stages $1, \ldots, j-1$, then the total number of possible outcomes is

$$n_1 \times n_2 \times n_3 \times \cdots \times n_J$$

- Number of k-element sequences from an n-element set is

$$n \times n \times \cdots \times n = n^k$$

Permutations

- Number of k-element sequences with no repeated elements, from an n-element set is

$$n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$$

This is described as the number of *permutations* of n elements taken k at a time, sometimes written $P_{n,k}$ or $(n)_k$.

Combinations

Combinations: How many k-element subsets does an n-element set have, where order does not matter.

$$\begin{aligned}&(\text{\# of } k\text{-element subsets}) \times (\text{\# of ways to order a } k\text{-element subset})\\ &= (\text{\# of } k\text{-element repeat-free sequences drawn from an } n\text{-element set})\end{aligned}$$

Then

$$(\text{\# of } k\text{-element subsets}) = \frac{P_{n,k}}{k!} = \frac{n!}{(n-k)!k!}$$

Write this number as

$$C_{n,k} = \binom{n}{k} = \frac{n!}{(n-k)!k!}$$

and call it n *choose* k, or the number of *combinations* of n elements taken k at a time.

Interview question

(DRW Trading)

There are $\binom{52}{5}$ different 5-card poker hands (order does not matter).

How many of those hands are full houses?

Interview question

(Morgan Stanley)

You are at point $(0, 0, 0)$ in a 3-dimensional grid. In every step you may move 1 unit up, to the right or to the front. How many different paths are there for you to get to the point $(3, 3, 3)$?

Probability space

Model a situation of uncertainty using a probability space

A probability space consists of:

- A sample space.

  A sample space is a set Ω whose elements are called *outcomes*.

  Each outcome is a possible result of the situation.
- A collection of events
- A probability measure

Events

Each event is a set of outcomes

- Example: 3 coin flips.

  $\Omega = \{HHH, \ HHT, \ HTH, \ HTT, \ THH, \ THT, \ TTH, \ TTT\}$

  Let $A = \{HHH, \ TTT\}$.

  Then A is the event that all flips are the same.

  Let $B = \{THH, \ THT, \ TTH, \ TTT\}$.

  Then B is the event that the first coin flip is T.

Events

Events are sets. Can express concepts using set or event terminology.

Event language	Set language	Notation
Sample space	Universal set	Ω
Event	Subset of Ω	$A, B, C, \ldots$
Impossible event	Empty set	$\emptyset$
A or B occurs	A union B	$A \cup B$
A and B occurs	A intersect B	$A \cap B$
not A	complement of A	A^c or $\Omega \backslash A$
A, B mutually exclusive	A, B are disjoint	$A \cap B = \emptyset$
A and not B	set difference	$A \cap B^c$ or $A \backslash B$

Probability measure

Probability measure is a function $\mathbb{P}$ which assigns to each event A a real number $\mathbb{P}(A)$ such that

- $\mathbb{P}(A) \geq 0$
- $\mathbb{P}(\Omega) = 1$
- If $A_1, \ldots, A_n$ are pairwise disjoint, then

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \sum_{j=1}^{n} \mathbb{P}(A_j)$$

- For a countably infinite collection of pairwise disjoint events,

$$\mathbb{P}(\bigcup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$$

Probability measure

Interpretations of probability:

- $\mathbb{P}(A)$ is the frequency with which A would occur if the experiment is run repeatedly. Justified by the law of large numbers.
- $\mathbb{P}(A)$ is the degree of belief in the certainty of A (subjective)

Equally likely outcomes

Suppose that $\Omega = \{\omega_1, \omega_2, \omega_3, \ldots, \omega_N\}$ and all outcomes are equally likely, i.e.

$$\mathbb{P}(\omega_1) = \mathbb{P}(\omega_2) = \mathbb{P}(\omega_3) = \cdots = \mathbb{P}(\omega_N) = \frac{1}{N}$$

Then

$$\mathbb{P}(G) = \frac{|G|}{N}$$

because

$1 = \mathbb{P}(\Omega) = \mathbb{P}(\cup_{\omega\in\Omega}\{\omega\}) = \sum_{i=1}^{N} \mathbb{P}(\{\omega_i\}) \Rightarrow \mathbb{P}(\{\omega\}) = \frac{1}{N}$ for each ω,

hence $\mathbb{P}(G) = \mathbb{P}(\cup_{\omega\in G}\{\omega\}) = \sum_{\omega\in G} \mathbb{P}(\omega) = \frac{|G|}{N}$

- Example: If all 5-card subsets of a deck are equally likely, then the probability of a full house is $3744/\binom{52}{5}$.

Interview question

(Goldman Sachs)

Two people toss a fair coin n times. What is the probability that they both got same number of heads?

Interview question

(Goldman Sachs)

Can you write it without any summation notation ?

Complements

- $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

  because $1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c) = \mathbb{P}(A) + \mathbb{P}(A^c)$

- $\mathbb{P}(\emptyset) = 0$

  because $\mathbb{P}(\emptyset) = \mathbb{P}(\Omega^c) = 1 - \mathbb{P}(\Omega) = 1 - 1$

- $\mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A \cap B)$

  because $\mathbb{P}(B) = \mathbb{P}(B \cap A^c) + \mathbb{P}(B \cap A)$

Unions

- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

  because

  $RHS = \mathbb{P}(A \cap B^c) + \mathbb{P}(A \cap B) + \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B) - \mathbb{P}(A \cap B)$

- $\mathbb{P}(A \cup B \cup C) =$

  $\mathbb{P}(A)+\mathbb{P}(B)+\mathbb{P}(C)-\mathbb{P}(A\cap B)-\mathbb{P}(A\cap C)-\mathbb{P}(B\cap C)+\mathbb{P}(A\cap B\cap C)$

- "Inclusion-exclusion" principle: By induction,

$$\mathbb{P}\Big(\bigcup_{i=1}^{n} A_i\Big) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) + \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1}\mathbb{P}\Big(\bigcap_{i=1}^{n} A_i\Big)$$

  Also true when you replace $\mathbb{P}$ with "# of elements" (if finite).

Interview question

(EWT Trading)

How many 7-digit phone numbers don't start with a 0 or 1, and don't have the string 911 in them?

Counting

Probability

Conditional Probability

Random Variables – Discrete

Random Variables – Continuous

Conditional Probability

- Definition: For any events A, B where $\mathbb{P}(B) > 0$, the conditional probability of A given B is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

- The idea is that if we assume event B occurs, then the universe should be reduced from Ω to B. Within that smaller universe, the frequency with which A occurs is $\mathbb{P}(A \cap B)/\mathbb{P}(B)$.

Interview question

(DE Shaw)

- A family has exactly two children. At least one is a boy. What's the probability that both are boys?
- A family has exactly two children. The older one is a boy. What's the probability that both are boys?

Interview question

(Jane Street Capital – Trading Assistant)

If Blue Eyed people are more likely to be Blond Haired (than the general population), can you infer that Blond Haired people are more likely to be Blue Eyed (than the general population)?

Independence

Definition: Events A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

- If $\mathbb{P}(B) > 0$ then independence is equivalent to $\mathbb{P}(A|B) = \mathbb{P}(A)$.
- "Mutually exclusive" does not imply "independent". In fact, only in trivial cases are mutually exclusive events also independent.

Interview question

(JP Morgan)

There's an 84% chance of seeing a shooting star in any given hour.

What's the probability of seeing a shooting star in any given 1/2 hour?

Interview question

(Susquehanna International Group, Assistant Trader position)

- I want to have a party this weekend (either Saturday or Sunday). The only way I can have a party is if it is sunny (only two possibilities are either sunny or rainy). The probability it is sunny Saturday is 0.80, the probability it is sunny Sunday is 0.40. What is the probability I have the party?

Interview question

(Susquehanna International Group, Assistant Trader position)

- Fast forward to Monday. Now assume I had my party.

  What is the probability it rained on one of the days?

Independence for more than 2 events

- Events A, B, C are independent if every pair of them is independent, and

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C).$$

- Note that if B and C are independent, then

$$\mathbb{P}(A|B \cap C) = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)} = \frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B)\mathbb{P}(C)}$$

  So if $\mathbb{P}(A) = \mathbb{P}(A|B \cap C)$ then $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

- Events $A_1, A_2, \ldots, A_n$ are independent if for any indices $i_1 \leq i_2 \leq \cdots \leq i_n$ we have

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_n}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \cdots \mathbb{P}(A_{i_n}).$$

Example: Independent trials

Consider n independent trials of some experiment, each of which succeeds with probability p.

- $\mathbb{P}(\text{at least 1 success}) = 1 - \mathbb{P}(\text{no successes}) = 1 - \mathbb{P}(A_1^c \cap \cdots \cap A_n^c)$
  $= 1 - \mathbb{P}(A_1^c)\mathbb{P}(A_2^c) \cdots \mathbb{P}(A_n^c) = 1 - (1-p)^n$.
- $\mathbb{P}(\text{exactly } k \text{ successes}) = \binom{n}{k} p^k (1-p)^{n-k}$
  because there are $\binom{n}{k}$ ways to choose which k trials to succeed, and for a particular choice of k trials to succeed and $n-k$ to fail, the probability is $p^k(1-p)^{n-k}$.

Interview question

(Jane Street Capital)

What's the probability that the NBA finals go to game 7?

The East team has probability p of winning each game.

Multiplication rule

Multiplication rule: Regardless of independence

$$\begin{aligned}&\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)\\&\quad = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2)\cdots\mathbb{P}(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})\end{aligned}$$

provided that the relevant probabilities are nonzero. This is because

$$\text{RHS} = \mathbb{P}(A_1)\frac{\mathbb{P}(A_2 \cap A_1)}{\mathbb{P}(A_1)}\frac{\mathbb{P}(A_3 \cap A_2 \cap A_1)}{\mathbb{P}(A_2 \cap A_1)}\cdots\frac{\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)}{\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_{n-1})}$$

Rule of Total Probability

If $B_1, \ldots, B_n$ partition Ω (meaning $\Omega = \bigcup_i B_i$ and all B_i are disjoint) then

$$\mathbb{P}(A) = \mathbb{P}(B_1)\mathbb{P}(A|B_1) + \mathbb{P}(B_2)\mathbb{P}(A|B_2) + \cdots + \mathbb{P}(B_n)\mathbb{P}(A|B_n)$$

because

$$\mathbb{P}(A) = \mathbb{P}(A \cap B_1) + \mathbb{P}(A \cap B_2) + \cdots + \mathbb{P}(A \cap B_n).$$

Interview question

(Morgan Stanley and Goldman Sachs)

We start with 1 amoeba. It and every subsequent amoeba can equiprobably and indpendently split to either 0 amoebas (dies), to 1, or to 2. What is the probability of eventual extinction of the population?

Interview question

(Goldman Sachs)

You have 100 seats in an airplane and 100 passengers waiting outside. Originally, each seat is assigned to one passenger. However, the first passenger is drunk and sits in a random seat. Then, every other passenger sits in the seat that has been assigned to them, unless it is occupied, in which case he picks randomly from the empty seats. What is the probability that the last passenger will get the correct seat?

Bayes Rule

If $B_1, \ldots, B_n$ partition Ω then for any $i = 1, \ldots, n$,

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\mathbb{P}(A|B_1)\mathbb{P}(B_1) + \mathbb{P}(A|B_2)\mathbb{P}(B_2) + \cdots + \mathbb{P}(A|B_n)\mathbb{P}(B_n)}$$

because each side equals $\frac{\mathbb{P}(B_i \cap A)}{\mathbb{P}(A)}$.

- Intuition: You observe that event A has occurred.

  The possible "causes" of event A are $B_1, \ldots, B_n$.

  Bayes rule expresses $\mathbb{P}(B_i|A)$ as the ratio

  $\mathbb{P}(A$ was caused by $B_i)/\mathbb{P}(A$ was caused by any of $B_1, \ldots, B_n)$

Interview question

(Goldman Sachs)

1000 coins, one of them is fake, with 2 heads. Pick one at random, flip it 10 times, get 10 heads. What's the probability that the coin is fake?

Interview question

(Morgan Stanley)

There were 2 boys and 3 girls in a room. A new baby entered the room whose gender was unknown (but equally likely to be a boy/girl). A nurse came into the room and picked one baby randomly and it turned out to be a boy. What's the probability that the new baby was a boy?

Counting

Probability

Conditional Probability

Random Variables – Discrete

Random Variables – Continuous

Random variable

Definition: A [real-valued] *random variable* is a [measurable] function mapping Ω to $\mathbb{R}$.

- Example: Toss a coin 3 times. Let X be the number of heads. Then X is the function from Ω to $\mathbb{R}$ where

$$\begin{array}{ccc} \omega & \longmapsto & X(\omega) \\ \hline \mathrm{HHH} & \longmapsto & 3 \\ \mathrm{HHT} & \longmapsto & 2 \\ \mathrm{HTH} & \longmapsto & 2 \\ \mathrm{HTT} & \longmapsto & 1 \\ \mathrm{THH} & \longmapsto & 2 \\ & \vdots & \end{array}$$

Random variable

- Statements about random variables are interpreted as events.
  For example, $X = 2$ is, by definition, the event

$$\{\omega : X(\omega) = 2\} = \{\mathrm{HHT}, \mathrm{HTH}, \mathrm{THH}\}$$

- Create new random variables by taking functions of random variables. For $f : \mathbb{R} \to \mathbb{R}$, define $f(X)$ to be the random variable which maps ω to $f(X(\omega))$.
  Example: X^2 is the random variable which maps ω to $[X(\omega)]^2$.

Discrete random variable

- Range of a random variable is the set $\{X(\omega) : \omega \in \Omega\}$ of all possible values X can take.
- A discrete random variable is one that has a countable range
- Countable means finite or countably infinite
- A set is countably infinite if it can be put in to 1-to-1 correspondence with the positive integers.

Examples

- Any interval of positive length is uncountable, such as $[0, 1]$ or $\mathbb{R}$.
- All subsets of the integers are countable

Distribution and probability mass function

- The distribution of a random variable is the function that maps each interval $I \subseteq \mathbb{R}$ to $\mathbb{P}(X \in I)$.
- The probability mass function (PMF) of a random variable X is the function that maps each point a to $\mathbb{P}(X = a)$.
- For discrete random variables, the PMF completely determines the distribution, because if X has range $\{x_1, x_2, \dots x_N\}$ then $\mathbb{P}(X \in I) = \sum_{x_i \in I} \mathbb{P}(X = x_i)$.

Examples

- Let X be the number of H in 3 coin tosses.
- Let Y be the number of tails in 3 coin tosses. Then Y has the same PMF as X, but $Y = 3 - X$. Just because two variables have identical distributions does not mean they are equal.

Bernoulli and Binomial Distributions

Definitions

- X has Bernoulli distribution with parameter p if $\mathbb{P}(X=0)=1-p$ and $\mathbb{P}(X=1)=p$.
- Bernoulli(p) trials are independent experiments, each with probability p of success
- X has Binomial distribution with parameters (n,p) if

$$\mathbb{P}(X=k)=\binom{n}{k}p^k(1-p)^{n-k} \quad \text{for all } k=0,\dots,n$$

  Interpretation: X is the # of successes in n Bernoulli(p) trials.
- Bernoulli(p) distribution is the same as the Binomial($1,p$) distribution

Poisson Distribution

- Let's try to construct the distribution of the number of occurrences of rare "successes" that have many opportunities to occur.
- Intuitively, the Binomial$(300, 0.01)$ and Binomial$(150, 0.02)$ distributions should be similar.
- So let X_n have distribution Binomial(n, p_n) where $p_n = \lambda/n$ for some constant λ. Then

$$
\begin{aligned}
\mathbb{P}(X_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\
&= \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1)(n-2)\cdots(n-k+1)}{n \cdot n \cdot n \cdots n} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&\to \frac{e^{-\lambda}\lambda^k}{} \quad \text{as } n \to \infty.
\end{aligned}
$$

Poisson Distribution

Definition: X has Poisson(λ) distribution if

$$\mathbb{P}(X=k)=\frac{e^{-\lambda}\lambda^k}{k!} \qquad k=0,1,2,\ldots$$

This is a valid distribution because $\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$.

- Binomial(n,p) is approximately Poisson(np) for n big, p small.

  Example: Let X be the number of defaults this year in a portfolio of 300 companies, each with a 1% chance of default. Find the probability that $X=5$.

Uniform and Geometric Distributions

- X has Uniform distribution on $\{x_1, \ldots, x_n\}$ if

$$\mathbb{P}(X = x_k) = \frac{1}{n} \qquad \text{for } k = 1, 2, \ldots, n.$$

  Example: number on a die roll has uniform distibution on $\{1, 2, 3, 4, 5, 6\}$.

- N has Geometric distribution with parameter p if

$$\mathbb{P}(N = k) = (1 - p)^{k-1} p \qquad \text{for } k = 1, 2, 3, \ldots$$

  N is the # of Bernoulli(p) trials needed, to obtain the first success.

  Example: Repeatedly roll a die until you get a 5.

  Number of rolls needed has Geometric(1/6) distribution.

Joint Distributions

How does X relate to other random variables? Definitions:

- The joint probability mass function of discrete random variables (X, Y) is the function f where $f(x, y) = \mathbb{P}(X = x, Y = y)$.
- The joint distribution of (X, Y) is the function that maps sets $B \subseteq \mathbb{R}^2$ to $\mathbb{P}((X, Y) \in B)$. For discrete random variables, the joint PMF completely determines the distribution.

Example: Toss a coin 3 times, let X be # of heads, Y be # of heads in the first 2 tosses

Joint Distributions

Let (X, Y) have joint probability mass function f.

To determine the probability that X and Y satisfy some condition, just sum the f over all (x, y) satisfying that condition.

- $\mathbb{P}(X > Y) = \sum_{x>y} f(x, y)$
- $\mathbb{P}(X = Y) = \sum_{x=y} f(x, y) = \sum_x f(x, x)$
- $\mathbb{P}(X = x_0) = \sum_{(x,y):x=x_0} f(x, y) = \sum_y f(x_0, y)$
- $\mathbb{P}(Y = y_0) = \sum_{(x,y):y=y_0} f(x, y_0) = \sum_x f(x, y_0)$

Example: in the coin toss example, $\mathbb{P}(X = 1) = \frac{1}{8} + \frac{2}{8} + 0 = \frac{3}{8}$.

Marginal Distributions

With regard to the joint distribution of (X, Y), the individual distributions of X and Y are called marginal distributions.

Joint distribution determines marginal distributions but not vice versa

Independence of random variables

Definitions:

- X and Y are independent if
  $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$ for all sets A and B in $\mathbb{R}$,
- $X_1, X_2, \ldots X_n$ are independent if for all $A_1, A_2, \ldots, A_n \subseteq \mathbb{R}$, the events $\{X_1 \in A_1\}$, $\{X_2 \in A_2\}$, $\ldots$, $\{X_n \in A_n\}$ are independent

Properties:

- Discrete X, Y are independent if and only if
  $\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y)$ for all (x, y)
- Functions of independent random variables are independent:
  If $X_1, \ldots, X_n$ are independent and $r_j : \mathbb{R} \to \mathbb{R}$ then
  $r_1(X_1), \ldots, r_n(X_n)$ are independent

Interview question

(JP Morgan, Beijing office)

Player "A" tosses a fair coin 10 times, player "B" tosses the coin 11 times. What is the probability that B has more heads than A.

Continuous random variables

Definition: $X : \Omega \to \mathbb{R}$ is a continuous random variable if there exists a nonnegative function $f : \mathbb{R} \to \mathbb{R}$ such that for all intervals on the real line $I \subseteq \mathbb{R}$,

$$\mathbb{P}(X \in I) = \int_I f(x)\mathrm{d}x$$

We then say that X has probability *density* function f. Properties

- $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \mathbb{P}(X \in (-\infty, \infty)) = 1$
- f determines $\mathbb{P}(X \in I)$ for all I, hence the distribution of X.
- For any particular value a, probability is zero that $X = a$, because $\mathbb{P}(X = a) = \int_a^a f(x)\mathrm{d}x = 0$.
- Intuitively interpret $f(x)$ as a probability per unit length

$$\mathbb{P}(X \in (a - \varepsilon/2, a + \varepsilon/2)) \approx f(a) \times \varepsilon$$

Continuous random variables

- For any countable $A = \{a_1, a_2, \ldots\}$, probability is 0 that $X \in A$, because $\mathbb{P}(X \in A) = \sum_{j=1}^{\infty} \mathbb{P}(X = a_j) = 0$.
- Note that intervals are uncountable sets, so it is possible for $\mathbb{P}(X \in (2,3)) > 0$ even though $\mathbb{P}(X = x) = 0$ for all $x \in (2,3)$.
- Recall that the distribution of a discrete random variable N is determined by its probability mass function $P(N = k)$ for each k.
  But for a continuous random variable X, the probability "mass function" is useless because $\mathbb{P}(X = x) = 0$ for all x.
  Instead we can use a density function to specify its distribution.

Examples of continuous random variables

- X has uniform distribution on (a, b) if it has density

$$f(x) = \frac{1}{b-a} \qquad \text{for } x \in (a, b)$$

  and $f(x) = 0$ for all $x \notin (a, b)$.

- X has Exponential distribution with parameter λ if it has density

$$f(x) = \lambda e^{-\lambda x} \qquad \text{for } x \geq 0$$

  and $f(x) = 0$ for all $x < 0$.

  Example: Could use Exponential(0.2) distribution to model $X =$ the time, in years, until a company defaults.

Examples of continuous random variables

- X has standard Normal distribution, or Normal$(0,1)$ distribution if it has density

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Example: Could use a normal distribution to model the log return $X = \log(S_T/S_0)$ on a stock price from time 0 to T.

Cumulative distribution function (CDF)

The cumulative distribution function (CDF) of a random variable X (including both continuous and discrete cases) is the function

$$F(x) = \mathbb{P}(X \leq x)$$

Properties:

- F is increasing (non-decreasing): If $a < b$ then $F(a) = \mathbb{P}(X \leq a) \leq \mathbb{P}(X \leq b) = F(b)$.
- $\lim_{x\to\infty} F(x) = 1$
- $\lim_{x\to-\infty} F(x) = 0$
- $\lim_{x\to a-} F(x) = \mathbb{P}(X < a)$

Cumulative distribution function (CDF)

- If $a \leq b$ then

  $\mathbb{P}(a < X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X \leq a) = F(b) - F(a).$

  For a continuous random variable, whether you have $\leq$ or $<$ in the $a < X \leq b$ does not matter, the probability is the same.

  For general random variables, obtaining $\leq$ vs. $<$ can be done by taking limits, for example

  $\mathbb{P}(a \leq X \leq b) = \mathbb{P}(X \leq b) - \mathbb{P}(X < a) = F(b) - \lim_{x \to a-} F(x).$

- The CDF of X completely determines $\mathbb{P}(X \in I)$ for all intervals I, hence the distribution of X – for any random variable X.

  (PMF does for discrete, and the PDF does for continuous.

  The CDF does for all.)

Cumulative distribution function (CDF)

- Many ways to specify a distribution:

  Give its name and its parameters (if it has a name).

  Or give its CDF.

  Or give its probability density function or probability mass function.

- In the case where X is continuous with density f,

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(u)\mathrm{d}u$$

  so if f is continuous at x then

$$F'(x) = f(x)$$

  Integrate the density function to get the CDF.

  Differentiate the CDF to get the density function.

CDF Examples

- CDF of Exponential(λ) distribution is

$$F(x) = \int_0^x \lambda e^{-\lambda u} \mathrm{d}u = -e^{-\lambda u}|_0^x = 1 - e^{-\lambda x} \qquad x \geq 0$$

  and $F(x) = 0$ for $x < 0$.

- CDF of Normal$(0, 1)$ distribution is

$$F(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \mathrm{d}u$$

  Instead of F, this CDF is usually written using notation N or Φ.

Interview question

(BP)

Given $X_i \sim \text{Uniform}[0,1]$ for $i = 1, .., n$.

What is the distribution of $M := \min(X_1, ..., X_n)$?

Interview question

(Goldman Sachs)

Let $X \sim N(0,1)$. Flip a fair coin (independent of X). If heads, let $Y = X$. If tails, let $Y = -X$. What's the distribution of Y?

Change of variable

Suppose X has density f_X and $Y = r(X)$, what is the density of Y?

Consider case where $r(x) = ax + b$ where $a > 0$.

Let F_X and F_Y be the CDF of X and Y respectively. Then

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(aX + b \leq y) = \mathbb{P}\Big(X \leq \frac{y-b}{a}\Big) = F_X\Big(\frac{y-b}{a}\Big)$$

Differentiate wrt y, and use chain rule to conclude:

$$f_Y(y) = \frac{1}{a} f_X\Big(\frac{y-b}{a}\Big)$$

Example: Normal distribution

Let X be Normal$(0, 1)$. Then $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Let $r(x) = \mu + \sigma x$, where $\sigma > 0$, and $Y = r(X)$. Then

$$f_Y(y) = \frac{1}{\sigma} f_X\left(\frac{y-\mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$$

If Y has this density, we say Y has Normal(μ, σ^2) distribution.

Interview question

(ING Bank – online quant test)

Write down the PDF and CDF for a Normal(μ, σ^2) distribution

Interview question

(JP Morgan, paper-based math test)

$X \sim N(0, 1)$ and $Z = 3X^2 + 1$, what is the PDF of Z?

Joint density functions

We say that (X, Y) have a joint probability density function f if

$$\mathbb{P}((X,Y) \in C) = \iint_C f(x,y)\mathrm{d}x\mathrm{d}y$$

for all regions $C \subseteq \mathbb{R}^2$. In that case we say (X, Y) are jointly continuous random variables.

Relation to marginal densities

Relation to one-variable densities:

- If (X, Y) have joint density f, then densities of X and Y individually are known as *marginal* densities
- Given f, can obtain marginal densities f_X and f_Y of X and Y by:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}y \qquad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\mathrm{d}x$$

- Given marginal densities f_X and f_Y, cannot determine joint density in general. If, and only if, X and Y are independent, then (X, Y) have joint density

$$f(x, y) = f_X(x)f_Y(y).$$

- This extends to more than 2 random variables.

Interview question

(Susquehanna International Group)

Let $X \sim N(0, \sigma_x^2)$ and $Y \sim N(0, \sigma_y^2)$ be independent. If you observe $X + Y = c$, what is your best guess for the values of X and Y?