Probability 1

Lecture 3-4: Discrete Random Variables

Dr. Daniel Flores-Agreda,

(based on the notes of Prof. Davide La Vecchia)

Spring Semester 2021

Outline

Random variables - what are they?

Up to now we have considered probabilities associated with random experiments characterized by different types of events, e.g.

- events for card flip (e.g. the card may be 'hearts or diamonds')
- events associated with coin tosses (e.g. the coins may show two heads 'HH')
- events defined as combinations of sets (an event in '$A \cup B^c$')
- ...

Definition

A **random variable** is a variable that takes on different **numerical** values (different outcomes) with various probabilities of occurrence associated with each different outcome.

To define a random variable, we need

1. to list all possible numerical outcomes, and
2. the corresponding probability for each numerical outcome

Random variables - what are they?

Example

- Roll a single die, and record the number of dots on the top side
- The list of all possible outcomes of the random process is the number shown on the die
  - i.e. the possible outcomes are 1, 2, 3, 4, 5 and 6
- If we say each outcome is equally likely, then the probability of each outcome must be $1/6$

Random variables - what are they?

Example

- Flip a coin 10 times, and record the number of times T (tail) occurs
- The possible outcomes of the random process are

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9 \text{ and } 10$$

- For each number we associate a probability
- The probabilities are determined by the assumptions made about the coin flips, e.g.
  - what is the probability of a ‘tail’ (or ‘head’) appearing on a single coin flip
  - whether this probability is the same for every coin flip
  - whether the 10 coin flips are ‘independent’ of each other

Random variables - what are they?

Example

- Suppose we want to study the time taken by school students to complete a test. Suppose that no student is given more than 2 hours to finish the test.
- If X = completion time (in minutes), the possible values of the random variable X are contained in the interval

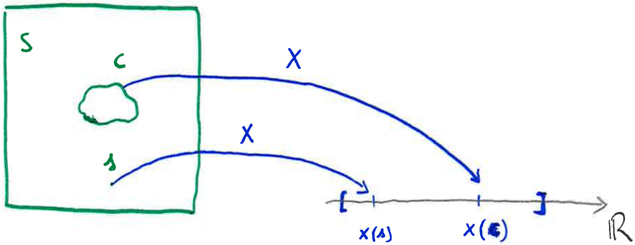
$$(0, 120] = \{x : 0 < x \leq 120\}.$$

- We then need to associate probabilities with all events we may wish to consider, such as

$$P(\{X \leq 15\}) \quad \text{or} \quad P(\{X > 60\}).$$

Formal definition of a random variable (I)

- Suppose we have:
  a. A sample space S
  b. A probability measure (Pr) "defined using the events" of S
- Let $X(s)$ be a function that takes an element $s \in S$ to a number x



Example: from S to D, via $X(\cdot)$

Example (Roll the die)

Experiment: We roll two dice and we consider the number of points in the first die, and the number of points in the second die. We already know that the sample space S is given by:

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

For the elements related to S we have a probability Pr

From S to D, via $X(\cdot)$

Example (cont'd)

Now define $X(s_{ij})$ as the sum of the outcome of the outcome i of the first die and the outcome j of the second die. Thus:

$$X(s_{ij}) = X(i,j) = i + j, \quad for \quad i = 1, ..., 6, and\ j = 1, ..., 6$$

In this notation $s_{ij} = (i,j)$ and $s_{ij} \in S$, each having probability $1/36$.

Facts:

- $X(\cdot)$ maps S into D. The (new) sample space D is given by

  $$D = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

  where, e.g., 2 is related to the pair $(1,1)$, 3 is related to the pairs $(1,2)$ and $(2,1)$, etc etc. To D is related the new P

Example (cont'd)

- To each element (event) in D we can attach a probability, using the probability of the corresponding event(s) in S. For instance,

$$P(2) = Pr(1,1) = 1/36, \quad \text{or} \quad P(3) = Pr(1,2) + Pr(2,1) = 2/36.$$

- How about the $P(7)$?

$$P(7) = Pr(3,4) + Pr(2,5) + Pr(1,6) + Pr(4,3) + Pr(5,2) + Pr(6,1) = 6/36.$$

- The latter equality can also be re-written as

$$P(7) = 2(Pr(3,4) + Pr(2,5) + Pr(1,6)) = 6\ Pr(3,4),$$

- What is $P(9)$? What is $P(13)$? [Hint: does 13 belong to D?]

Formal definition of a random variable (II)

Let us formalize all these ideas:

- Let D be the set of all values x that can be obtained by $X(s)$, for all $s \in S$:

$$D = \{x : x = X(s), \ s \in S\}$$

- D is a list of all possible numbers x that can be obtained, and thus is a **sample space** for X. *Remark that the random variable is X while x represents its realization (non random).*
- D can be either an **uncountable interval** (then X is a **continuous** random variable), or
- D can be **discrete** or **countable** (the X is a **discrete** random variable)

Formal definition of a random variable (III)

For X to be a random variable it is required that for each A (as "made" by elements in D)

$$P(A) = Pr(\{s \in S : X(s) \in A\})$$

where P and Pr stand for "probability" on D and on S, respectively. So

1. $P(A) \geq 0$
2. $P(D) = Pr(\{s \in S : X(s) \in D\}) = Pr(S) = 1$
3. If $A_1, A_2, A_3 ...$ is a sequence of events such that

   $$A_i \cap A_j = \varnothing$$

   for all $i \neq j$ then

   $$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

From now on, I drop the colors.

An Example from gambling

Example (Geometric random variable)

Consider the problem of rolling a die until a 6 appears.

- Let X denote the number of rolls required for the process to end
- The possible values of X are: $1, 2, 3, \ldots, n, \ldots$ $(\equiv \mathbb{N})$.
- $P(\{X = 1\}) = P($'6' appears on the 1st roll$) = \frac{1}{6}$
- $P(\{X = 2\}) = P\,($no '6' on the 1st roll and '6' on the 2nd roll$) = \frac{5}{6} \cdot \frac{1}{6} = \frac{5}{36}$
- $P(\{X = 3\}) =$
  $P\,($no '6' on either the 1st or 2nd roll and '6' on the third roll$)$
  $= \frac{5}{6} \cdot \frac{5}{6} \cdot \frac{1}{6} = \frac{25}{216}$
- . . . and so on . . .
- $P(\{X = n\}) = P($no '6' on the first n-1 rolls and '6' on the last roll$)$
  $= \left(\frac{5}{6}\right)^{n-1} \cdot \frac{1}{6}$
- . . . and so on . . ..

An Example from gambling

Example (cont'd)

- Rather than list the possible values of X along with the associated probabilities in a table, we can provide a formula that gives the required probabilities.
- The probability distribution of the random variable X is given by

$$P(\{X = n\}) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} \quad \text{for} \quad n = 1, 2, \ldots$$

- Note that (using properties of geometric series)

$$\sum_{n=1}^{\infty} \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} = 1.$$

Discrete random variables

Discrete random variables are often associated with the process of counting (see previous example). More generally,

Definition

Suppose X can take the values $x_1, x_2, x_3, \ldots, x_n$. The probability of x_i is

$$p_i = P(\{X = x_i\})$$

and we must have $p_1 + p_2 + p_3 + \cdots + p_n = 1$ and all $p_i \geq 0$. These probabilities may be put in a table

x_i	$P(\{X = x_i\})$
x_1	p_1
x_2	p_2
x_3	p_3
$\vdots$	$\vdots$
x_n	p_n
Total	1

Discrete random variables

For a **discrete random variable** X, any table listing all possible nonzero probabilities provides the entire **probability distribution**. The **probability mass function** $p(a)$ of X is defined by

$$p_a = p(a) = P(\{X = a\}),$$

and this is positive for at most a countable number of values of a. For instance, $p_1 = P(\{X = x_1\})$, $p_2 = P(\{X = x_2\})$, and so on. That is, if X must assume one of the values $x_1, x_2, ...$, then

$$\begin{aligned} p(x_i) &\geq 0 \quad \text{for} \quad i = 1, 2, ... \\ p(x) &= 0 \qquad \text{otherwise.} \end{aligned} \tag{1}$$

Clearly, we must have

$$\sum_{i=1}^{\infty} p(x_i) = 1.$$

Cumulative Distribution Function (I)

The **cumulative distribution function** (**CDF**) is a table listing the values that X can take, along with

$$F_X(a) = P\left(\{X \leq a\}\right) = \sum_{\text{all} \quad x \leq a} p(x).$$

If the random variable X takes on values $x_1, x_2, x_3, \ldots, x_n$ *listed in increasing order* $x_1 < x_2 < x_3 < \cdots < x_n$, the CDF is a step function, that it its value is constant in the intervals $(x_{i-1}, x_i]$ and takes a step/jump of size p_i at each x_i:

x_i	$F_X(x_i) = P\left(\{X \leq x_i\}\right)$
x_1	p_1
x_2	$p_1 + p_2$
x_3	$p_1 + p_2 + p_3$
$\vdots$	$\vdots$
x_n	$p_1 + p_2 + \cdots + p_n = 1$

Cumulative Distribution Function (II)

Example

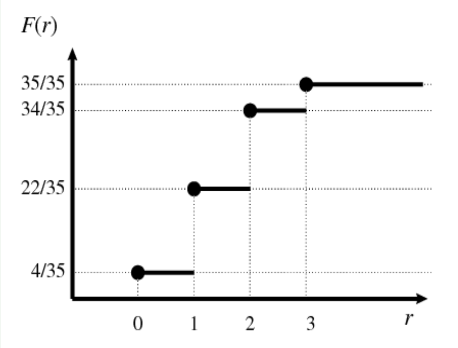
x_i	0	1	2	3	Tot
$P(\{X = x_i\})$	$4/35$	$18/35$	$12/35$	$1/35$	1
$P(\{X \leq x_i\})$	$4/35$	$22/35$	$34/35$	$35/35$	

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 4/35 & 0 \leq x < 1 \\ 22/35 & 1 \leq x < 2 \\ 34/35 & 2 \leq x < 3 \\ 1 & x \geq 3. \end{cases}$$

Cumulative Distribution Function (III)

Example (cont'd)

... or graphically, you get *a step function* ...



Cumulative Distribution Function (IV)

Remark

Suppose $a \leq b$. *Then, because the event* $\{X \leq a\}$ *is contained in the event* $\{X \leq b\}$, *namely*

$$\{X \leq a\} \subseteq \{X \leq b\},$$

it follows that

$$F_X(a) \leq F_X(b),$$

so, the probability of the former is less than or equal to the probability of the latter.

In other words, $F_X(x)$ is a nondecreasing function of x.

Cumulative Distribution Function (V)

Example (Quantiles)

The CDF can be inverted to define the value x of X that corresponds to a given probability α, namely $\alpha = P(X \leq x)$, for $\alpha \in [0, 1]$.

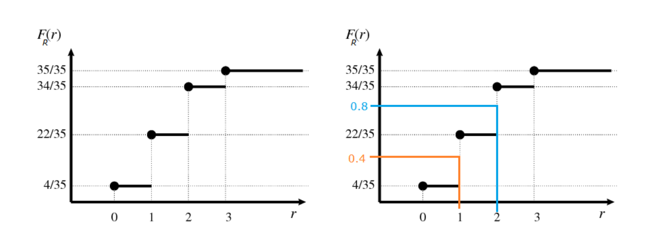
The inverse CDF $F_X^{-1}(\alpha)$ or quantile of order α, and labelled as $Q(\alpha)$, is the smallest realisation of X associated to a CDF greater or equal to α; in formula, the α-quantile $Q(\alpha)$ is the smallest number satisfying:

$$F_X[F_X^{-1}(\alpha)] = P[X \leq \underbrace{F_X^{-1}(\alpha)}_{Q(\alpha)}] \geq \alpha, \quad \text{for} \quad \alpha \in [0, 1].$$

By construction, a quantile of a discrete random variable is a realization of X. More to come in Lecture 5...

Cumulative Distribution Function (VI)

Example (cont'd, graphically)



...calling (only for this slide) R the rv, r its realizations and $F_R(r)$ its CDF at r...

Distributional summaries for discrete random variables

- For a discrete random variable, it is useful to describe some attributes or properties of the distribution
  - such as a measure of location and a measure of spread
- The **expected value**, or **mean value**, of the distribution

$$E[X] = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n = \sum_{i=1}^{n} p_i x_i$$

  is a measure of *location* — roughly speaking this is the *center* of the distribution;
- The **square root of the variance**, or **standard deviation,** of the distribution

$$\begin{aligned} s.d(X) &= \sqrt{Var(X)} \\ &= \sqrt{p_1 (x_1 - E[X])^2 + p_2 (x_2 - E[X])^2 + \cdots + p_n (x_n - E[X])^2} \end{aligned}$$

  is a measure of *spread (or 'variability' or 'dispersion').*

Important properties

If X is a discrete random variable and a is any real number, then

1. $E[\alpha X] = \alpha E[X]$
2. $E[\alpha + X] = \alpha + E[X]$
3. $Var(\alpha X) = \alpha^2 Var(X)$
4. $Var(\alpha + X) = Var(X)$

Exercise

Let us very the first property: $E[\alpha X] = \alpha E[X]$. From the Intro lecture we know that, for every $\alpha_i \in \mathbb{R}$,

$$\sum_{i=1}^{n} \alpha_i X_i = \alpha_1 X_1 + \alpha_2 X_2 + ... + \alpha_n X_n.$$

So, the required result follows as a special case, setting $\alpha_i = \alpha$, for every i, and applying the definition of expected value. Verify this and the other properties as an exercise. [Hint: set $\alpha_i = \alpha p_i$.]

Dependence/Independence

Definition

Consider two discrete random variables[a] X and Y. Then, X and Y are **independent** if

$$P(\{ X = x\} \cap \{Y = y\}) = P(\{X = x\}) \cdot P(\{Y = y\})$$

for all values x that X can take and all values y that Y can take.

[a]Technically speaking, X and Y should be defined on the same probability space, but we do not pursue this argument.

More important properties

- If X and Y are two discrete random variables, then

$$E[X+Y] = E[X] + E[Y]$$

- If X and Y are also *independent*, then

$$Var(X+Y) = Var(X) + Var(Y) \tag{2}$$

Remark

Note that Eq. (2) does not (typically) hold if X and Y are NOT independent—more to come on this later on...

More on expectations

Recall that the expectation of X was defined as

$$E[X] = \sum_{i=1}^{n} p_i x_i$$

Now, suppose we are interested in a function m of the random variable X, say $m(X)$. We define

$$E[m(X)] = p_1 m(x_1) + p_2 m(x_2) + \cdots p_n m(x_n).$$

Notice that the variance is a special case of expectation where,

$$m(X) = (X - E[X])^2.$$

Indeed,

$$Var(X) = E\left[(X - E[X])^2\right].$$

Exercise

Show that

$$Var(X) = E\left[X^2\right] - E[X]^2.$$

Some discrete distributions of interest

- Discrete Uniform
- Bernoulli
- Binomial
- Poisson
- Hypergeometric
- Negative binomial

Their main characteristic is that the probability $P(\{X = x_i\})$ is given by an appropriate mathematical formula: i.e.

$$p_i = P(\{X = x_i\}) = h(x_i)$$

for a suitably specified function $h(\cdot)$.

Discrete uniform distribution

Definition

We say X has a **discrete uniform distribution** when

- X can take the values $x = 0, 1, 2, ..., k$ (for some specified finite value $k \in \mathbb{N}$)
- The probability that $X = x$ is $1/(k+1)$, namely

$$P(\{X = x\}) = \frac{1}{(k+1)}.$$

The probability distribution is given by

x_i	$P(\{X = x_i\})$
0	$\frac{1}{(k+1)}$
1	$\frac{1}{(k+1)}$
$\vdots$	$\vdots$
k	$\frac{1}{(k+1)}$
Total	1

Discrete uniform: expected value

- The expected value of X is

$$\begin{aligned} E[X] &= x_1 p_1 + \ldots + x_k p_k \\ &= 0 \cdot \frac{1}{(k+1)} + 1 \cdot \frac{1}{(k+1)} + \cdots + k \cdot \frac{1}{(k+1)} \\ &= \frac{1}{(k+1)} \cdot (0 + 1 + \cdots + k) \\ &= \frac{1}{(k+1)} \cdot \frac{k(k+1)}{2} \\ &= \frac{k}{2}. \end{aligned}$$

E.g. when $k = 6$, then X can take on one of the seven distinct values $x = 0, 1, 2, 3, 4, 5, 6$, each with equal probability $\frac{1}{7}$, but the expected value of X is equal to 3, which is one of the possible outcomes!!!

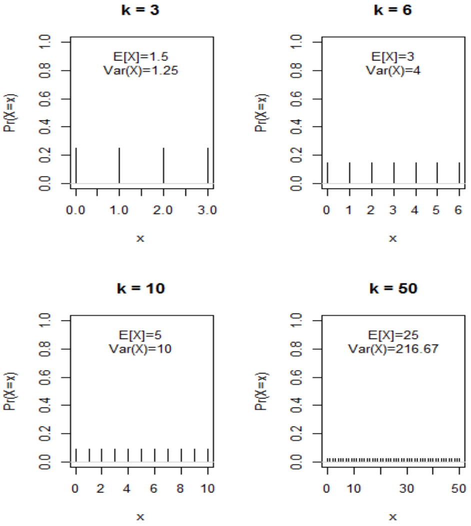
Discrete uniform: variance

- The variance of X – we will be denoting it as $Var(X)$ – is

$$
\begin{aligned}
Var\left(X\right) &= \left(0-\frac{k}{2}\right)^2\cdot\frac{1}{(k+1)}+\left(1-\frac{k}{2}\right)^2\cdot\frac{1}{(k+1)}+ \\
&\quad \cdots+\left(k-\frac{k}{2}\right)^2\cdot\frac{1}{(k+1)} \\
&= \frac{1}{(k+1)}\cdot\left\{\left(0-\frac{k}{2}\right)^2+\left(1-\frac{k}{2}\right)^2+\cdots+\left(k-\frac{k}{2}\right)^2\right\} \\
&= \frac{1}{(k+1)}\cdot\frac{k\left(k+1\right)\left(k+2\right)}{12} \\
&= \frac{k\left(k+2\right)}{12}
\end{aligned}
$$

E.g. when $k=6$, the variance of X is equal to 4, and the standard deviation of X is equal to $\sqrt{4}=2$.

Some illustrations of discrete uniform



Some illustrations of discrete uniform

Example

An example of discrete uniform is related to the experiment of rolling a die—remark, the outcome zero is not allowed in this specific example.

Let us call X the corresponding random variable and $\{x_1, x_2, ..., x_6\}$ its realizations.

The possible outcomes are

$$\{1, 2, 3, 4, 5, 6\}$$

each having probability $\frac{1}{6}$.

Moreover,

$$E(X) = (1 + 2 + 3 + 4 + 5 + 6) \cdot \frac{1}{6} = 3.5,$$

which is not one of the possible outcomes!!!

Bernoulli Trials

Definition

Bernoulli trial is the name given to the random variable X having probability distribution given by

x_i	$P(\{X = x_i\})$
1	p
0	$1 - p$

Often we write

$$P(\{X = x\}) = p^x \left(1 - p\right)^{1-x}, \qquad \text{for } x = 0, 1$$

Bernoulli Trials

A Bernoulli trial represents the most primitive form of all random variables. It derives from a random experiment having only two possible mutually exclusive outcomes. These are often labelled Success and Failure and

- Success occurs with probability p
- Failure occurs with probability $1 - p$.

Remark

Just for the sake of notation, let us set $X = 1$ if Success occurs, and $X = 0$ if Failure occurs

Example

Coin tossing: we can define a random variable

x_i	$P(\{X = x_i\})$
1	p
0	$1 - p$

and say that $X = 1$ if H and $X = 0$ if T.

Bernoulli Trials

- Mean:

$$\begin{aligned} E[X] &= 1 \cdot p + 0 \cdot (1-p) \\ &= p \end{aligned}$$

- Variance:

$$\begin{aligned} Var(X) &= (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) \\ &= p(1-p). \end{aligned}$$

Definition

Let us consider the random experiment consisting in a series of n trials having 3 characteristics

1. Only two mutually exclusive outcomes are possible in each trial: *success* (S) and *failure* (F)
2. The outcomes in the series of n trials constitute independent events
3. The probability of success p in each trial is constant from trial to trial

X is the *number of successes* occurring in n (Bernoulli) trials. Binomial probability distribution given by

$$\begin{aligned} P(\{X = x\}) &= \binom{n}{x} p^x (1-p)^{n-x} \\ &= \frac{n!}{x!\,(n-x)!} p^x (1-p)^{n-x}, \text{ for } x = 0, 1, 2, ..., n \end{aligned} \tag{3}$$

The Binomial Distribution

Recall (see Intro lecture) that combinations are defined as:

$$\left(\begin{array}{c} n \\ k \end{array} \right) = \frac{n!}{k!\,(n-k)!} = C_n^k$$

and, for $n \geq k$, we say "n choose k".

The binomial coefficient $\left(\begin{array}{c} n \\ k \end{array} \right)$ represents the number of possible combinations of n objects taken k at a time, without regard of the order.

Thus, C_n^k represents the number of different groups of size k that could be selected from a set of n objects when the order of selection is not relevant.

The Binomial Distribution

So, "What is the interpretation of the formula? "

1. The first factor
$$\left(\begin{array}{c} n \\ x \end{array} \right) = \frac{n!}{x!\,(n-x)!}$$
is the number of different combinations of individual 'successes' and 'failures' in n (Bernoulli) trials that result in a sequence containing a total of x 'successes' and $n-x$ 'failures'.
2. The second factor
$$p^x\,(1-p)^{n-x}$$
is the probability associated with any one sequence of x 'successes' and $(n-x)$ 'failures'.

Remark

Short-hand notation:
$$X \sim B(x,n,p)$$
or, occasionally, simply $X \sim B(n,p)$ *(no* x *in the formula).*

The Binomial Distribution: expectation and variance

- Mean:

$$
\begin{aligned}
E[X] &= \sum_{x=0}^{n} xP\{X = x\} \\
&= \sum_{x=0}^{n} x \binom{n}{x} p^{x} (1-p)^{n-x} = np
\end{aligned}
$$

- Variance:

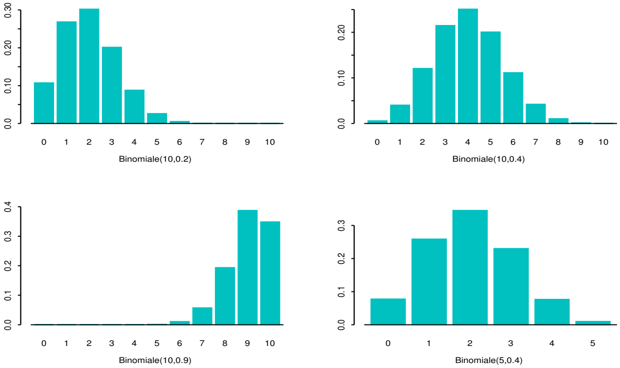
$$
\begin{aligned}
Var(X) &= \sum_{x=0}^{n} (x - np)^2 P(\{X = x\}) \\
&= np(1-p)
\end{aligned}
$$

Remark

Looking at (3), we remark that the Bernoulli distribution is a special case ($n = 1$) of the Binomial distribution. Roughly speaking, "a Binomial random variable arises when we sum n independent Bernoulli trails."

Some illustrations of Binomial

... same barplot as in slide 30, just a bit fancier...



Binomial Distribution

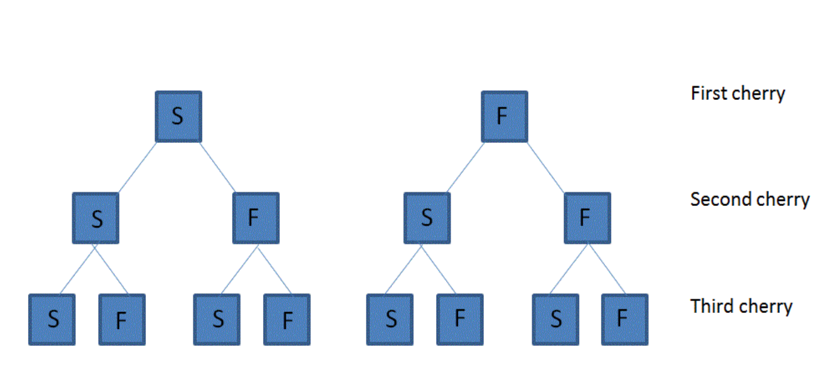
Example (cherry trees)

One night a storm washes three cherries ashore on an island. For each cherry, there is a probability $p = 0.8$ that its seed will produce a tree. What is the probability that these three cherries will produce two trees?

First, we notice that this can be determined using a **Bernoulli distribution**. To this end, consider whether each seed will produce a tree as a sequence of $n = 3$ trials. For each cherry:

- either the cherry produces a tree (Success) or it does not (Failure);
- the event that a cherry produces a tree is independent from the event that any of the other two cherries produces a tree.
- The probability that a cherry produces a tree is the same for all three cherries

Example (cont'd)



Example (cont'd)

- There are $2^3 = 8$ possible outcomes from the 3 individual trials
- It does not matter which of the three cherries produce a tree
- Consider all of the possible sequences of outcomes (S=success, F=failure)

SSS, SSF, SFS, SFF, FSS , FSF, FFS, FFF

- We are interested in SSF, SFS, FSS
- These possible events are *mutually exclusive*, so

$$P(\{SSF \cup SFS \cup FSS\}) = P(\{SSF\}) + P(\{SFS\}) + P(\{FSS\})$$

Binomial Distribution

Example (cont'd)

The three trials are assumed to be *independent*, so each of the three seed events corresponding to two trees growing has the same probability

$$\begin{aligned} P(\{SSF\}) &= P(\{S\}) \cdot P(\{S\}) \cdot (P\{F\}) \\ &= 0.8 \cdot 0.8 \cdot (1-0.8) \\ &= 0.8 \cdot (1-0.8) \cdot 0.8 = P(\{SFS\}) \\ &= (1-0.8) \cdot 0.8 \cdot 0.8 = P(\{FSS\}) \\ &= 0.128 \end{aligned}$$

So the probability of two trees resulting from the three seeds must be

$$\begin{aligned} P(\{SSF \cup SFS \cup FSS\}) &= 3 \cdot 0.128 \\ &= 0.384. \end{aligned}$$

Example

Finally, we notice that we can obtain the same result (in a more direct way), using the **binomial probability** for the random variable

$$X = \text{number of trees that grows from 3 seeds.}$$

Indeed

$$\begin{aligned} P(\{X = 2\}) &= \frac{3!}{2!\,(3-2)!} \cdot (0.8)^2 \cdot (1-0.8)^{3-2} \\ &= 3 \cdot (0.8)^2 \cdot (0.2) \\ &= 0.384. \end{aligned}$$

Poisson Distribution

Definition

Let us consider random variable X which takes values $0, 1, 2, \ldots$, namely the nonnegative integers in $\mathbb{N}$. X is said to be a Poisson random variable if its probability mass function, with $\lambda > 0$ fixed and providing info on the intensity, is

$$p(x) = P\left(\{X = x\}\right) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, \ldots \tag{4}$$

and we write $X \sim \text{Poisson}(\lambda)$.

The Eq. (4) defines a genuine probability mass function, since $p(x) \geq 0$ and

$$\begin{aligned} \sum_{x=0}^{\infty} p(x) &= \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= e^{-\lambda} e^{\lambda} = 1 \quad \text{(see Intro Lecture).} \end{aligned}$$

Poisson Distribution

Moreover, for a given value of λ also the CDF can be easily defined. E.g.

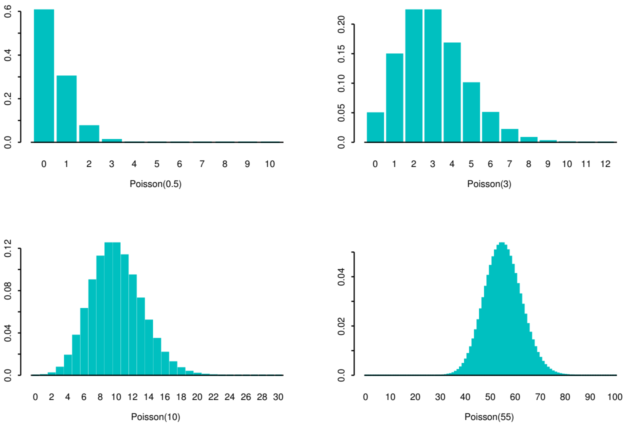
$$F_X(2) = P\left(\{X \leq 2\}\right) = e^{-\lambda} + \lambda e^{-\lambda} + \frac{\lambda^2 e^{-\lambda}}{2},$$

and the Expected value and Variance for Poisson distribution (see tutorial) can be obtained by "sum algebra" (and/or some algebra)

$$\begin{aligned} E[X] &= \lambda \\ Var(X) &= \lambda. \end{aligned}$$

Some illustrations of Poisson

... same barplot as in slide 30, just a bit fancier...



Poisson Distribution

Example

The average number of newspapers sold by Alfred is 5 per minute[a]. What is the probability that Alfred will sell at least 1 newspaper in a minute?

To answer, let X be the # of newspapers sold by Alfred in a minute. We have

$$X \sim \text{Poisson}(\lambda)$$

with $\lambda = 5$, so

$$\begin{aligned} P(X \geq 1) &= 1 - P(\{X = 0\}) \\ &= 1 - \exp^{-5} \frac{5^0}{0!} \\ &\approx 1 - 0.0067 \approx 99.33\%. \end{aligned}$$

How about $P(X \geq 2)$? Is it $P(X \geq 2) \geq P(X \geq 1)$ or not? Answer the question...

[a] This number provides info on the intensity at which a random phenomenon occurs.

Poisson Distribution

Example

A telephone switchboard handles 300 calls, on the average, during one hour. The board can make maximum 10 connections per minute. Use the Poisson distribution to evaluate the probability that the board will be overtaxed during a given minute.

To answer, let us set $\lambda = 300$ per hour, which is equivalent to 5 calls per minute. Noe let us define

$$X = \# \text{ of connections in a minute}$$

and by assumption we have $X \sim \text{Poisson}(\lambda)$. Thus,

$$\begin{aligned} P[\text{overtaxed}] &= P(\{X > 10\}) \\ &= 1 - \underbrace{P(\{X \leq 10\})}_{\text{using } \lambda = 5, \text{ minute base}} \\ &\approx 0.0137. \end{aligned}$$

Poisson Distribution (link to Binomial)

Let us consider $X \sim B(x, n, p)$, where n is large, p is small, and the product np is appreciable. Setting, $\lambda = np$, we then have that, for the Binomial probability as in Eq.(3), it is a good approximation to write:

$$p(k) = P(\{X = k\}) \approx \frac{\lambda^k}{k!} e^{-\lambda}.$$

To see this, remember that

$$\lim_{n \to \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

Then, let us consider that in our setting, we have $p = \lambda / n$. From the formula of the binomial probability mass function we have:

$$p(0) = (1 - p)^n = \left(1 - \frac{\lambda}{n}\right)^n \approx e^{-\lambda}, \quad \text{as} \quad n \to \infty.$$

Poisson Distribution (link to Binomial)

Moreover, it is easily found that

$$\frac{p(k)}{p(k-1)} \quad = \quad \frac{np-(k-1)p}{k(1-p)} \approx \frac{\lambda}{k}, \qquad \text{as} \quad n \to \infty.$$

Therefore, we have

$$\begin{array}{rcl}
p(1) & \approx & \dfrac{\lambda}{1!}p(0) \approx \lambda e^{-\lambda} \\
p(2) & \approx & \dfrac{\lambda}{2!}p(1) \approx \dfrac{\lambda^2}{2}e^{-\lambda} \\
\cdots & \cdots & \cdots \\
p(k) & \approx & \dfrac{\lambda}{k!}p(k-1) \approx \underbrace{\dfrac{\lambda^k}{k!}e^{-\lambda}}_{\text{see Eq. (4)}}
\end{array}$$

thus, we remark that $p(k)$ can be approximated by the probability mass function of a Poisson—which is easier to implement.

Poisson Distribution: example

Example (two-fold use of Poisson)

Suppose a certain high-speed printer makes errors at random on printed paper[a]. Assuming that the Poisson distribution with parameter $\lambda = 4$ is appropriate to model the number of errors per page (say, X), what is the probability that in a book containing 300 pages (produced by the printer) at least 7 will have no errors?

Let X denote the number of errors per page, so that

$$p(x) = \exp^{-4} \frac{4^x}{x!}, \quad \text{for} \quad x = 0, 1, 2, ...$$

The probability of any page to be error free is then

$$p(0) = \exp^{-4} \frac{4^0}{0!} = \exp^{-4} \approx 0.018.$$

[a] This exercise is related to Ex 2 of PS6. The calculation is similar but not identical: note the difference between the size of p in this example and in the tutorial.

Poisson Distribution: example

Example (cont'd)

Having no errors on a page is a success, and there are 300 independent pages. Hence, let us define

$$Y = \text{the number of pages without any errors.}$$

Y is binomially distributed with parameters $n = 300$ and $p = 0.018$, namely

$$Y \sim B(n, p).$$

But here we have

$$n \text{ large, } p \text{ small, and } np = 5.4$$

thus, we can compute $P(\{Y \geq 7\})$ using either the exact Binomial or its Poisson approximation. So

- using $B(300, 0.018)$, we get: $P(\{Y \geq 7\}) \approx 0.297$
- using Poisson(5.4), we get $P(\{Y \geq 7\}) \approx 0.298$.

The Hypergeometric Distribution

Definition

Let us consider a random experiment consisting of a series of n trials, having the following properties

1. Only two mutually exclusive outcomes are possible in each trials: success (S) and failure (F)
2. The population has N elements in which k are looked upon as S and the other $N - k$ are looked upon as F
3. Sampling from the population is done **without** replacement (so that the trials are not independent).

The random variable

$$X = \text{number of successes in } n \text{ such trials}$$

has an hypergeometric distribution and ....

The Hypergeometric Distribution

Definition (cont'd)

... the probability that $X = x$ is

$$P(\{X = x\}) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}.$$

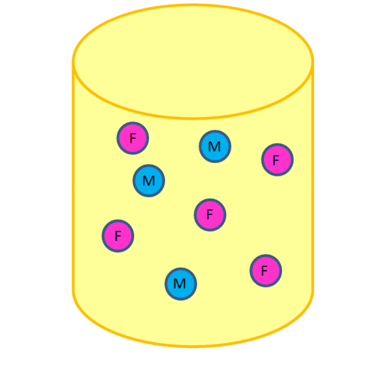
Moreover,

$$\begin{aligned} E\left[X\right] &= \frac{nk}{N} \\ Var\left(X\right) &= \frac{nk\left(N-k\right)\left(N-n\right)}{N^2\left(N-1\right)} \end{aligned}$$

Hypergeometric Distribution Example

Example (Psychological experiment)

A group of 8 students includes 5 women and 3 men: 3 students are randomly chosen to participate in a psychological experiment. What is the probability that *exactly* 2 women will be included in the sample?



Hypergeometric Distribution Example

Example (cont'd)

Consider each of the three participants being selected as a separate trial $\Rightarrow$ there are $n = 3$ trials. Consider a woman being selected in a trial as a 'success'
Then here $N = 8$, $k = 5$, $n = 3$, and $x = 2$, so that

$$
\begin{aligned}
P(\{X = 2\}) &= \frac{\begin{pmatrix} 5 \\ 2 \end{pmatrix}\begin{pmatrix} 8-5 \\ 3-2 \end{pmatrix}}{\begin{pmatrix} 8 \\ 3 \end{pmatrix}} \\
&= \frac{\frac{5!}{2!3!}\,\frac{3!}{1!2!}}{\frac{8!}{5!3!}} \\
&= 0.53571
\end{aligned}
$$

The Negative Binomial Distribution

Let us consider a random experiment consisting of a series of trials, having the following properties

1. Only two mutually exclusive outcomes are possible in each trial: 'success' (S) and 'failure' (F)
2. The outcomes in the series of trials constitute *independent events*
3. The probability of success p in each trial is *constant* from trial to trial

What is the probability of having exactly y F's before the r^{th} S?

Equivalently: What is the probability that in a sequence of $y + r$ (Bernoulli) trials the last trial yields the r^{th} S?

The Negative Binomial Distribution

Definition

Let

$X =$ the total number of trials required until a total of r successes is accumulated.

Then X is said to be a Negative Binomial random variable and its probability mass function $P(\{X = n\})$ equals the probability of $r - 1$ 'successes' in the first $n - 1$ trials, times the probability of a 'success' on the last trial. These probabilities are given by

$$P(\{X = n\}) = \begin{pmatrix} n-1 \\ r-1 \end{pmatrix} p^r (1-p)^{n-r} \quad \text{for } n = r, r+1, \ldots$$

The mean and variance for X are, respectively,

$$\begin{aligned} E[X] &= \frac{r}{p} \\ Var(X) &= \frac{r(1-p)}{p^2} \end{aligned}$$

Negative Binomial Distribution Example

Example (marketing research)

- A marketing researcher wants to find 5 people to join her focus group
- Let p denote the probability that a randomly selected individual agrees to participate in the focus group
- If $p = 0.2$, what is the probability that the researcher must ask 15 individuals before 5 are found who agree to participate?
- In this case, $p = 0.2$, $r = 5$, $n = 15$: we are looking for $P(\{X = 15\})$. By the negative binomial formula we have

$$
\begin{aligned}
P(\{X = 15\}) &= \binom{14}{4} (0.2)^5 (0.8)^{10} \\
&= 0.034
\end{aligned}
$$

The Geometric Distribution

Definition (a special case)

When $r = 1$, the negative binomial distribution is equivalent to the **Geometric distribution —** see Example in slide 13.

In this case, probabilities are given by

$$P(\{X = n\}) = p\,(1-p)^{n-1}\,,\ \text{for } n = 1, 2, \ldots$$

The corresponding mean and variance for X are, respectively,

$$\begin{aligned} E[X] &= \frac{1}{p} \\ Var(X) &= \frac{(1-p)}{p^2} \end{aligned}$$

Example (failure of a machine)

Items are produced by a machine having a 3% defective rate.

- What is the probability that the first defective occurs in the fifth item inspected?

$$\begin{aligned} P(\{X=5\}) &= P(\text{first 4 non-defective})P(\text{5th defective}) \\ &= (0.97)^4(0.03) \approx 0.026 \end{aligned}$$

- What is the probability that the first defective occurs in the first five inspections?

$$\begin{aligned} P(\{X \leq 5\}) = P(\{X < 6\}) &= P(\{X=1\}) + \ldots + P(\{X=5\}) \\ &= 1 - P(\text{first 5 non-defective}) = 0.1412. \end{aligned}$$

More generally, for a geometric random variable we have:

$$P(\{X \geq k\}) = (1-p)^{k-1}.$$

Thus, in the example we have $P(\{X \geq 6\}) = (1-0.03)^{6-1} \approx 0.8587$

$$P(\{X \leq 5\}) = 1 - P(\{X \geq 6\}) \approx 1 - 0.8587 \approx 0.1412.$$