P2 - Discrete Distributions

STAT 401 (Engineering) - Iowa State University

January 10, 2018

Random variables

Definition

A random variable X is a function $X : \Omega \mapsto \mathbb{R}$.

Intuitive idea: If the value of a numerical variable depends on the outcome of an experiment, we call the variable a *random variable*.

Examples of random variables from rolling two 6-sided dice:

- Sum of the two dice
- Indicator of the sum being greater than 5

Generally, we will use an upper case Roman letter to indicate a random variable and a lower case Roman letters to indicate a realized value of the random variable.

8 bit example

Example

Suppose, 8 bits are sent through a communication channel. Each bit has a certain probability to be received incorrectly. We are interested in the number of bits that are received incorrectly.

- Let X be the number of incorrect bits received.
- The possible values for X are $\{0, 1, 2, 3, 4, 5, 6, 7, 8\}$.
- Example events:
  - No incorrect bits received: $\{X = 0\}$.
  - At least one incorrect bit received: $\{X \geq 1\}$.
  - Exactly two incorrect bits received: $\{X = 2\}$.
  - Between two and seven (inclusive) incorrect bits received: $\{2 \leq X \leq 7\}$.

Image of random variables

Definition

The image of a random variable X is defined as

$$Im(X) := \{x : x = X(\omega) \text{ for some } \omega \in \Omega\}$$

If the image is finite or countable, we have a discrete random variable. If the image is uncountably infinite, we have a continuous random variable.

Example

- Put a hard drive into service, measure Y = "time till the first major failure" and thus $Im(Y) = (0, \infty)$. Image of Y is an interval (uncountable image), so Y is a continuous random variable.
- Communication channel: X = "# of incorrectly received bits" with $Im(X) = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$. Image of X is a finite set, so X is a discrete random variable.

Distribution

Definition

The collection of all the probabilities related to X is the distribution of X.

For a discrete random variable, the function

$$p_X(x) = P(X = x)$$

is the probability mass function (pmf) and the cumulative distribution function (cdf) is

$$F(x) = P(X \leq x) = \sum_{y \leq x} P(y).$$

The set of possible values of X is called the support of the distribution F and is the same as the image of X.

Examples

A probability mass function is valid if it defines a valid set of probabilities, i.e.

- the probabilities are non-negative,
- the probabilities sum to 1.

Example

Which of the following functions are a valid probability mass functions?

-

x	-3	-1	0	5	7
$P_X(x)$	0.1	0.45	0.15	0.25	0.05

-

y	-1	0	1.5	3	4.5
$P_Y(y)$	0.1	0.45	0.25	-0.05	0.25

-

z	0	1	3	5	7
$P_Z(z)$	0.22	0.18	0.24	0.17	0.18

Rolling a fair 6-sided die

Example

Let Y be the number of pips on the upturned face of a die. The image of Y is $\{1, 2, 3, 4, 5, 6\}$. If we believe the die has equal probability for each face, then the pmf and cdf for Y are

y	1	2	3	4	5	6
$p_Y(y) = P(Y = y)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$F_Y(y) = P(Y \leq y)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	$\frac{4}{6}$	$\frac{5}{6}$	$\frac{6}{6}$

Dragonwood

Example

Dragonwood has 6-sided dice with the following # on the 6 sides: $\{1, 2, 2, 3, 3, 4\}$.

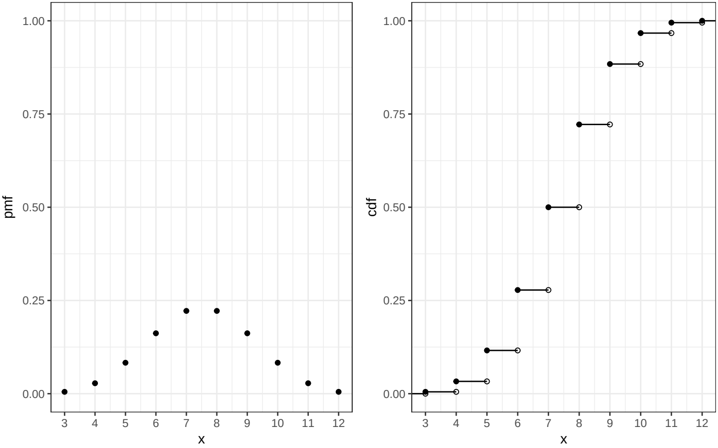
What is the image, pmf, and cdf for the sum of the upturned numbers when rolling 3 Dragonwood dice?

```
# Three dice
die   = c(1,2,2,3,3,4)
rolls = expand.grid(die1 = die, die2 = die, die3 = die)
sum   = rowSums(rolls); tsum = table(sum)
dragonwood3 = data.frame(x    = round(as.numeric(names(tsum)),0),
                         pmf = round(as.numeric(table(sum)/length(sum)),3)) %>%
        mutate(cdf = cumsum(pmf))

t(dragonwood3)
```

```
     [,1]  [,2]  [,3]  [,4]  [,5]  [,6]  [,7]   [,8]   [,9]  [,10]
x   3.000 4.000 5.000 6.000 7.000 8.000 9.000 10.000 11.000 12.000
pmf 0.005 0.028 0.083 0.162 0.222 0.222 0.162  0.083  0.028  0.005
cdf 0.005 0.033 0.116 0.278 0.500 0.722 0.884  0.967  0.995  1.000
```

Dragonwood - pmf and cdf



Properties of pmf and cdf

Properties of probability mass function $p_X(x) = P(X = x)$:

- $0 \leq p_X(x) \leq 1$ for all $x \in \mathbb{R}$.
- $\sum_{x \in S} p_X(x) = 1$ where S is the support.

Properties of cumulative distribution function $F_X(x)$:

- $0 \leq F_X(x) \leq 1$ for all $x \in \mathbb{R}$
- F_X is nondecreasing, (i.e. if $x_1 \leq x_2$ then $F_X(x_1) \leq F_X(x_2)$.)
- $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$.
- $F_X(x)$ is right continuous with respect to x

Dragonwood (cont.)

In Dragonwood, you capture monsters by rolling a sum equal to or greater than its defense. Suppose you have 3 dice and the following monsters are available to be captured:

- A monster worth 1 victory point with a defense of 3.
- A monster worth 3 victory points with a defense of 7.
- A monster worth 4 victory points with a defense of 8.

Which monster should your attack?

We can calculate the probability of defeating each monster by computing one minus the cdf evaluated at "defense minus 1". Let X be the sum of the number on 3 Dragonwood dice. Then

- $P(X \geq 3) = 1 - P(X \leq 2) = 1$
- $P(X \geq 7) = 1 - P(X \leq 6) = 0.722.$
- $P(X \geq 8) = 1 - P(X \leq 7) = 0.5.$

Expectation

Definition

Let X be a random variable and h be some function. The expected value of a function of a (discrete) random variable is

$$E[h(X)] = \sum_i h(x_i) \cdot p_X(x_i).$$

If $h(x) = x$, then

$$E[X] = \sum_i x_i \cdot p_X(x_i)$$

and we call this the expectation of X. We commonly use the symbol μ for the expectation.

Expected values are *weighted averages* of the possible values weighted by their probability.

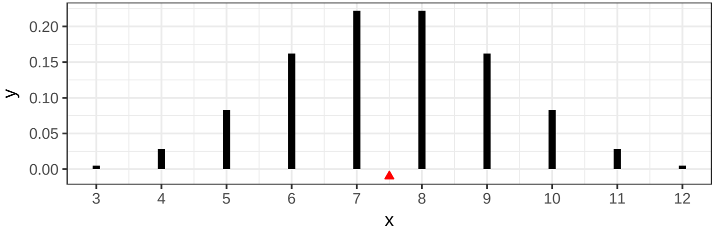
Dragonwood (cont.)

What is the expectation of the sum of 3 Dragonwood dice?

```
expectation = with(dragonwood3, sum(x*pmf))
expectation

[1] 7.5
```

The expectation can be thought of as the center of mass if we place mass $p_X(x)$ at corresponding points x.



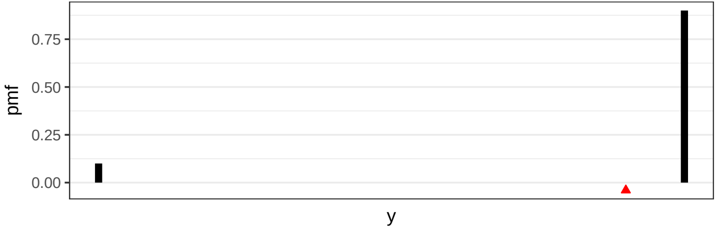
Biased coin

Suppose we have a biased coin represented by the following pmf:

y	0	1
$p_Y(y)$	$1-p$	p

What is the expected value?

If $p = 0.9$,



Properties of expectations

Theorem

Let X and Y be random variables and a, b, and c be constants. Then

$$E[aX + bY + c] \quad = aE[X] + bE[Y] + c.$$

In particular

Corollary

- $E[X + Y] = E[X] + E[Y]$,
- $E[aX] = aE[X]$, *and*
- $E[c] = c$.

Dragonwood (cont.)

Enhancement cards in Dragonwood allow you to improve your rolls. Here are two enhancement cards:

- *Cloak of Darkness* adds 2 points to all capture attempts and
- *Friendly Bunny* allows you (once) to roll an extra die.

What is the expected attack roll total if you had 3 Dragonwood dice, the Cloak of Darkness, and are using the Friendly Bunny?

Let

- X be the sum of 3 Dragonwood dice (we know $E[X] = 7.5$),
- Y be the sum of 1 Dragonwood die which has $E[X] = 2.5$.

Then the attack roll total is $X + Y + 2$ and the *expected* attack roll total is

$$E[X + Y + 2] = E[X] + E[Y] + 2 = 7.5 + 2.5 + 2 = 12,$$

or the attack roll is $4Y + 2$ and the *expected* attack roll total is

$$E[4Y + 2] = 4E[Y] + 2 = 12.$$

Variance

Definition

The variance of a random variable is defined as the expected squared deviation from the mean. For discrete random variables, variance is

$$Var[X] = E[(X-\mu)^2] = \sum_i (x_i - \mu)^2 \cdot p_X(x_i)$$

where $\mu = E[X]$. The symbol σ^2 is commonly used for the variance.

The variance is analogous to moment of intertia in classical mechanics.

Definition

The standard deviation is the positive square root of the variance

$$SD[X] = \sqrt{Var[X]}.$$

The symbol σ is commonly used for the standard deviation.

Properties of variance

If X and Y are independent and a, b, and c are constants, then

$$Var[aX + bY + c] = a^2 Var[X] + b^2 Var[Y].$$

Special cases:

- $Var[c] = 0$
- $Var[aX] = a^2 Var[X]$
- $Var[X + Y] = Var[X] + Var[Y]$ (if X and Y are independent)

Dragonwood (cont.)

What is the variance for the sum of the 3 Dragonwood dice?

```
variance = with(dragonwood3, sum((x-expectation)^2*pmf))
variance

[1] 2.766
```

What is the standard deviation for the sum of the pips on 3 Dragonwood dice?

```
sqrt(variance)

[1] 1.66313
```

Biased coin

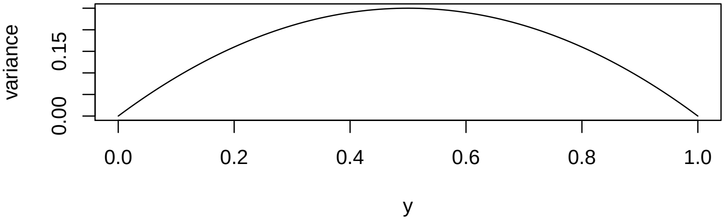
Suppose we have a biased coin represented by the following pmf:

y	0	1
$p_Y(y)$	$1-p$	p

What is the variance?

1. $E[Y] = p$
2. $V[y] = (0-p)^2(1-p) + (1-p)^2 \times p = p - p^2 = p(1-p)$

When is this variance maximized?



Special discrete distributions

- Bernoulli
- Binomial
- Poisson

Note: The image is always finite or countable.

Bernoulli distribution

A Bernoulli experiment has only two outcomes: success/failure.

Let

- $X = 1$ represent success and
- $X = 0$ represent failure.

The probability mass function $p_X(x)$ is

$$p_X(0) = 1 - p \quad p_X(1) = p.$$

We use the notation $X \sim Ber(p)$ to denote a random variable X that follows a Bernoulli distribution with success probability p, i.e. $P(X = 1) = p$.

Bernoulli experiment examples

Example

- Toss a coin: $\Omega = \{H, T\}$
- Throw a fair die and ask if the face value is a six: $\Omega = \{\text{face value is a six}, \text{face value is not a six}\}$
- Send a message through a network and record whether or not it is received: $\Omega = \{\text{successful transmission},\ \text{unsuccessful transmission}\}$
- Draw a part from an assembly line and record whether or not it is defective: $\Omega = \{\text{defective},\ \text{good}\}$
- Response to the question "Are you in favor of the above measure"? (in reference to a new tax levy to be imposed on city residents): $\Omega = \{\text{yes}, \text{no}\}$

Bernoulli distribution (cont.)

The cdf of the Bernoulli distribution is

$$F_X(x) = P(X \le x) = \begin{cases} 0 & x < 0 \\ 1-p & 0 \le x < 1 \\ 1 & 1 \le x \end{cases}$$

The expected value is

$$E[X] = \sum_x p_X(x) = 0(1-p) + 1 \cdot p = p.$$

The variance is

$$Var[X] = \sum_x (x - E[X])^2 p_X(x) = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = p(1-p).$$

Sequence of Bernoulli experiments

A compound experiment consisting of n independent and identically distributed Bernoulli experiments. E.g.

- Toss a coin n times.
- Send 23 identical messages through the network independently.
- Draw 5 cards from a standard deck with replacement (and reshuffling) and record whether or not the card is a king.

What does independent and identically distributed mean?

Independent and identically distributed

Let X_i represent the i^{th} Bernoulli experiment.

Independence means

$$P(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} P(X_i = x_i),$$

i.e. the joint probability is the product of the individual probabilities.

Identically distributed (for Bernoulli random variables) means

$$P(X_i = 1) = p \quad \forall\, i,$$

and more generally, the distribution is the same for all the random variables.

We will use iid as a shorthand for *independent and identically distributed*, although I will often use ind to indicate *independent* and let *identically distributed* be clear from context.

Sequences of Bernoulli experiments

Let X_i denote the outcome of the i^{th} Bernoulli experiment. We use the notation

$$X_i \stackrel{iid}{\sim} Ber(p), \quad \text{for } i = 1, \ldots, n$$

to indicate a sequence of n independent and identically distributed Bernoulli experiments.

We could write this equivalently as

$$X_i \stackrel{ind}{\sim} Ber(p), \quad \text{for } i = 1, \ldots, n$$

but this is different than

$$X_i \stackrel{ind}{\sim} Ber(p_i), \quad \text{for } i = 1, \ldots, n$$

as the latter has a different success probability for each experiment.

Binomial distribution

Suppose we perform a sequence of n iid Bernoulli experiments and only record the number of successes, i.e.

$$Y = \sum_{i=1}^{n} X_i.$$

Then we use the notation $Y \sim Bin(n, p)$ to indicate a binomial distribution with

- n attempts and
- probability of success p.

Binomial probability mass function

We need to obtain

$$p_Y(y) = P(Y = y) \quad \forall\, y \in \Omega = \{0, 1, 2, \ldots, n\}.$$

The probability of obtaining a particular sequence of y success and $n - y$ failures is

$$p^y(1-p)^{n-y}$$

since the experiments are iid with success probability p. But there are

$$\binom{n}{y} = \frac{n!}{y!(n-y)!}$$

ways of obtaining a sequence of y success and $n - y$ failures. Thus, the binomial pmf is

$$p_Y(y) = P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}.$$

Properties of the binomial distribution

The expected value is

$$E[Y] = E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np.$$

The variance is

$$Var[Y] = \sum_{i=1}^{n} Var[X_i] = np(1-p)$$

since the X_i are independent.

The cumulative distribution function is

$$F_Y(y) = P(Y \le y) = \sum_{x=0}^{\lfloor y \rfloor} \binom{n}{x} p^x (1-p)^{n-x}.$$

Component failure rate

Suppose a box contains 15 components that each have a failure rate of 5%.

What is the probability that

1. exactly two out of the fifteen components are defective?
2. at most two components are defective?
3. more than three components are defective?
4. more than 1 but less than 4 are defective?

Let Y be the number of defective components and assume $Y \sim Bin(15, 0.05)$.

$$\begin{array}{lll}
1. & P(Y = 2) & = P(Y = 2) = \binom{15}{2}(0.05)^2(1-0.05)^{15-2} \\
2. & P(Y \le 2) & = \sum_{x=0}^{2} \binom{15}{x}(0.05)^x(1-0.05)^{15-x} \\
3. & P(Y > 3) & = 1 - P(Y \le 3) = 1 - \sum_{x=0}^{3} \binom{15}{x}(0.05)^x(1-0.05)^{15-x} \\
4. & P(1 < Y < 4) & = \sum_{x=2}^{3} \binom{15}{x}(0.05)^x(1-0.05)^{15-x}
\end{array}$$

Component failure rate (solutions in R)

```
n <- 15
p <- 0.05
choose(15,2)
```

```
[1] 105
```

```
dbinom(2,n,p)          # P(Y=2)
```

```
[1] 0.1347523
```

```
pbinom(2,n,p)          # P(Y<=2)
```

```
[1] 0.9637998
```

```
1-pbinom(3,n,p)        # P(Y>3)
```

```
[1] 0.005467259
```

```
sum(dbinom(c(2,3),n,p)) # P(1<Y<4) = P(Y=2)+P(Y=3)
```

```
[1] 0.1654853
```

Poisson distribution

Many experiments can be thought of as "how many *rare* events will occur in a certain amount of time or space?" For example,

- # of alpha particles emitted from a polonium bar in an 8 minute period
- # of flaws on a standard size piece of manufactured product, e.g., 100m coaxial cable, 100 sq.meter plastic sheeting
- # of hits on a web page in a 24h period

These situations can be effectly modeled using a Poisson distribution. The Poisson distribution has pmf

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, 3, \ldots$$

where λ is called the rate parameter. We write $X \sim Po(\lambda)$ to represent this random variable. We can show that

$$E[X] = Var[X] = \lambda.$$

Poisson distribution - example

Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day. What is the probability that more than 8 new accounts will be initiated today?

Let X be the number of initiated accounts today. Assume $X \sim Po(10)$.

$$P(X > 8) = 1 - P(X \leq 8) = 1 - \sum_{x=0}^{8} \frac{\lambda^x e^{-\lambda}}{x!} \approx 1 - 0.333 = 0.667$$

In R,

```
# Using pmf
1-sum(dpois(0:8, lambda=10))
```

```
[1] 0.6671803
```

```
# Using cdf
1-ppois(8, lambda=10)
```

```
[1] 0.6671803
```

Sum of Poisson random variables

Theorem

Let $X_i \overset{ind}{\sim} Po(\lambda_i)$ *for* $i = 1, \ldots, n$. *Then*

$$Y = \sum_{i=1}^{n} X_i \sim Po\left(\sum_{i=1}^{n} \lambda_i\right).$$

Corollary

Let $X_i \overset{iid}{\sim} Po(\lambda)$ *for* $i = 1, \ldots, n$. *Then*

$$Y = \sum_{i=1}^{n} X_i \sim Po\left(n\lambda\right).$$

Poisson distribution - example

Customers of an internet service provider initiate new accounts at the average rate of 10 accounts per day. What is the probability that more than 16 new accounts will be initiated in the next two days?

Since the rate is 10/day, then for two days we expect, on average, to have 20. Let Y be he number initiated in a two-day period and assume $Y \sim Po(20)$. Then

$$P(Y > 16) = 1 - P(Y \leq 16) = 1 - \sum_{x=0}^{16} \frac{\lambda^x e^{-\lambda}}{x!} = 1 - 0.221 = 0.779.$$

In R,

```
# Using pmf
1-sum(dpois(0:16, lambda=20))


[1] 0.7789258


# Using cdf
1-ppois(16, lambda=20)
```

Manufacturing example

A manufacturer produces 100 chips per day and, on average, 1% of these chips are defective. What is the probability that no defectives are found in a particular day?

Let X represent the number of defectives and assume $X \sim Bin(100, 0.01)$. Then

$$P(X = 0) = \binom{100}{0}(0.01)^0(1 - 0.01)^{100} \approx 0.366.$$

Alternatively, let Y represent the number of defectives and assume $Y \sim Po(100 \times 0.01)$. Then

$$P(Y = 0) = \frac{1^0 e^{-1}}{0!} \approx 0.368.$$

Poisson approximation to the binomial

Theorem

Let $\{X_n\}$ be a sequence of random variables such that $X_n \sim Bin(N_n, p_n)$ with $N_n \to \infty$ and $N_n p_n \to \lambda \in (0, \infty)$, then

$$X_n \to X \sim Po(\lambda)$$

in distribution.

For large n, the binomial dsitribution, $Bin(n, p)$, can be approximated by a Poisson distribution, $Po(np)$, since

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-np} \frac{(np)^k}{k!}.$$

Rule of thumb: use Poisson approximation if $n \geq 20$ and $p \leq 0.05$.

Example

Imagine you are supposed to proofread a paper. Let us assume that there are on average 2 typos on a page and a page has 1000 words. This gives a probability of 0.002 for each word to contain a typo. What is the probability the page has no typos?

Let X represent the number of typos on the page and assume $X \sim Bin(1000, 0.002)$. $P(X = 0)$ using R is

```
n = 1000; p = 0.002
dbinom(0, size=n, prob=p)

[1] 0.1350645
```

Alternatively, let Y represent the number of defectives and assume $Y \sim Po(1000 \times 0.002)$. $P(Y = 0)$ using R is

```
dpois(0, lambda=n*p)

[1] 0.1353353
```