Discrete Distributions

Lecture 8

Termeh Shafie

random variables

- Many random experiments have outcomes that are numerical

  For example:
  - number of people on a train
  - the time a customer will spend in line at the post office
  - the number of people voting for a candidate in a political election

- In random experiments with outcomes not numerical, we map outcomes to numerical values

  For example:
  - in a coin toss experiment : H → 1 and T → 0
- A random variable associates a number with each outcome of a random experiment
- Parameters shape probability distributions of random variables

random variables

Given an experiment and the sample space Ω, a random variable is a function mapping an outcome ($\omega \in \Omega$) into a real number, i.e.

$$X : \omega \in \Omega \rightarrow X(\omega) \in (-\infty, \infty)$$

- We use a capital letter X to denote a random variable
- The values of a random variable will be denoted with a lower case letter x
- The range of a random variable is the set of values it can take
- a function of a random variable is another mapping from the sample space to real numbers, so another random variable

random variables

example

Toss a coin 3 times: the sample space is $\Omega : \{H,T\} \times \{H,T\} \times \{H,T\}$

Define the random variable: $X =$ the number of heads

What is the probability of each outcome of X?

Outcome (ω)	HHH	HTH THH HHT	HTT THT TTH	TTT
$X(\omega)$	3	2	1	0

$$P(X=3) = \frac{1}{8} \qquad P(X=1) = \frac{3}{8}$$

$$P(X=2) = \frac{3}{8} \qquad P(X=0) = \frac{1}{8}$$

random variables

exercise 1

Toss two dice, the sample space is given by $\Omega : \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\}$
Let the random variable X denote the sum of the two dice.
What is the probability of each outcome X?

6						
5						
4						
3						
2						
1						
	1	2	3	4	5	6

discrete random variables: probability mass function

A random variable is discrete if its range is a countable (finite or infinite) set.

If X is a discrete random variable, the function given by $f(x) = P(X = x)$ for each x within the range of X is called the probability distribution of X, also called probability mass function (pmf)

A function can serve as the probability distribution of a discrete random variable X if and only if its values, $f(x)$, satisfy the conditions:

- $f(x) \geq 0$ for each value within its domain
- $\sum_x f(x) = 1$ where the sum if over all the x values within its domain.

discrete random variables: probability mass function

example (cont'd...)

Toss a coin 3 times: the sample space is $\Omega : \{H,T\} \times \{H,T\} \times \{H,T\}$
Define the random variable: $X =$ the number of heads
What is the probability distribution of X?

Outcome (ω)	HHH	HTH THH HHT	HTT THT TTH	TTT
$X(\omega)$	3	2	1	0

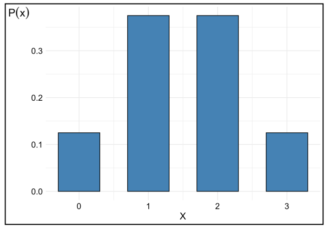
$$P(X=3) = \frac{1}{8} \qquad P(X=1) = \frac{3}{8}$$

$$P(X=2) = \frac{3}{8} \qquad P(X=0) = \frac{1}{8}$$

discrete random variables: probability mass function

example (cont'd...)

Toss a coin 3 times: the sample space is $\Omega : \{H,T\} \times \{H,T\} \times \{H,T\}$
Define the random variable: $X =$ the number of heads
What is the probability distribution of X?



expected value

The expected value, is the (probability) weighted average of the possible outcomes

$$E(X) = \sum_x x \cdot P(X = x)$$

the center of gravity of the PMF

The expected value rule:

Let X be a random variable with PMF $f(x)$ and let $g(X)$ be a function of X. Then,

$$E[g(X)] = \sum_x g(x) \cdot f(x)$$

variance

The variance is given by

$$V(X) = E[(X - E(X))^2 = \sum_x (x - E(X))^2 \cdot P(X = x) = E[X^2] - E[X]^2$$

the standard deviation $\sqrt{V(X)}$ is usually easier to interpret

The variance is always nonnegative

We can find $V(X)$ by calculating the mean of $Z = (X - E[X])^2$ via the expected value rule

When computing the variance often we use a different (equivalent) form of the variance equation:

$$V(X) = E[X^2] - E[X]^2$$

exercise 2

Prove this

expected value and variance

exercise 3

Toss a coin 3 times. Define the random variable: $X =$ the number of heads

What is the expected value and variance of X?

X	$f(x) = P(X = x)$
0	1/8
1	3/8
2	3/8
3	1/8

cumulative distribution function

If X is a discrete random variable, the function given by

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

for $-\infty < x < \infty$ is the cumulative distribution of X

The values $F(x)$ of the cumulative distribution of a discrete random variable X satisfies the conditions:

- $f(-\infty) = 0$ and $f(\infty) = 1$
- If $a < b$, then $F(a) \leq F(b)$ for any real numbers a and b

cumulative distribution function

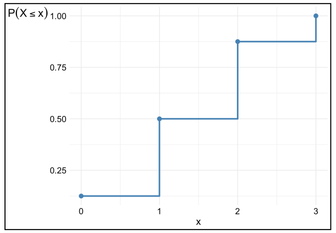
example (cont'd...)

Toss a coin 3 times: the sample space is Ω : {H,T} × {H,T} × {H,T}

Define the random variable: $X =$ the number of heads

What is the cdf of X?

x	$f(x) = P(X = x)$	$F(x) = P(X \leq x)$
0	1/8	1/8
1	3/8	4/8
2	3/8	7/8
3	1/8	8/8



summary: pmf and cdf

PMF of a discrete random variable X

gives the probability that X takes on a specific value

$$p(x_i) = P(X = x_i) = P(\{s \in S \mid X(s) = x_i\}$$

$$\sum_{x_i} p(x_i) = p(x_1) + p(x_2) + \cdots = 1$$

$$p(x_i) \geq 0 \quad \forall x_i$$

$$P(X \in A) = \sum_{x_i \in A} p(x_i)$$

CDF of a discrete random variable X

gives the probability X takes on a value that is less than or equal to a specific value

$$F(x) = P(X \leq x), \quad \text{for any } x \in \mathbb{R}$$

$$F(x) = P(X \leq x) = P(X \in A) = \sum_{x_i \leq x} p(x_i)$$

$$p(x_i) \geq 0 \quad \forall x_i$$

$$P(X \in A) = \sum_{x_i \in A} p(x_i)$$

joint, marginal and conditional distributions

Contingency table based on relative frequencies

example

Suppose we randomly select a family from a large population. Let:

X = number of boys in the family

Y = number of girls in the family

where X + Y ≤ 4.

joint, marginal and conditional distributions

Contingency table based on relative frequencies

example cont'd

Joint and marginal probability distributions

	Number of girls (Y)					
Number of boys (X)	**0**	**1**	**2**	**3**	**4**	$P(X = x) = \sum_y P(x, y)$
0	0.38	0.16	0.04	0.01	0.01	0.60
1	0.17	0.08	0.02	–	–	0.27
2	0.05	0.02	0.01	–	–	0.08
3	0.02	0.01	–	–	–	0.03
4	0.02	–	–	–	–	0.02
$P(Y = y) = \sum_x P(x, y)$	0.64	0.27	0.07	0.01	0.01	1.00

joint, marginal and conditional distributions

Contingency table based on relative frequencies

example

Conditional distributions.

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

Number of boys (X)	Number of girls (Y) 0	1	2	3	4	P(X = x)
0	0.59	0.59	0.57	1.00	1.00	0.60
1	0.27	0.30	0.29	–	–	0.27
2	0.08	0.07	0.14	–	–	0.08
3	0.03	0.04	–	–	–	0.03
4	0.02	–	–	–	–	0.02
	1.00	1.00	1.00	1.00	1.00	1.00

joint, marginal and conditional distributions

Contingency table based on relative frequencies

example

Conditional distributions.

are X and Y independent?

Number of boys (X)	Number of girls (Y) 0	1	2	3	4	P(X = x)
0	0.59	0.59	0.57	1.00	1.00	0.60
1	0.27	0.30	0.29	–	–	0.27
2	0.08	0.07	0.14	–	–	0.08
3	0.03	0.04	–	–	–	0.03
4	0.02	–	–	–	–	0.02
	1.00	1.00	1.00	1.00	1.00	1.00

vampirical vs. empirical

Odds Of Having Boy Or Girl Aren't Exactly 50-50, Study Reveals

The study found that some families are more likely to have children of the same sex than would be expected by chance.

Edited by: Srishti Singh Sisodia | Feature | Jul 21, 2025 17:53 pm IST

Read Time: 3 mins

Share

Representative image.

https://www.science.org/doi/full/10.1126/sciadv.adu7402

Bernoulli random variable

$$P(X = x \mid p) = \begin{cases} p, & \text{if } x = 1 \\ 1 - p, & \text{if } x = 0 \end{cases}$$

- A random variable for modeling binary events
- Two possible outcomes:
  - Success: value 1
  - Failure: value 0
- Single parameter p, probability of a success
- multiple Bernoulli r.v. can be combined to model more complex random variables
- Shorthand notation: $X \sim \text{Bern}(p)$
- $E(X) = p, \; V(X) = p(1 - p)$

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

geometric random variable

- A r.v. modeling the number of (identical) Bernoulli trials needed to obtain the first success
- Infinite outcomes $\{1,2,3\ldots,\infty\}$
- Single parameter p, probability of a success for each trial
- Shorthand notation: $X \sim \text{Geo}(p)$
- $E(X) = \frac{1}{p},\ V(X) = \frac{1-p}{(p^2)}$

$$P(X = x \,|\, p) = p(1-p)^{x-1}$$

X	$P(X=x)$
1	p
2	$p(1-p)$
3	$p(1-p)^2$
4	$p(1-p)^3$
$\vdots$	$\vdots$
∞	$p(1-p)^{\infty} \approx 0$

exercise 4

What is the probability of flipping a coin more than 4 times before getting a heads?

exercise 5

What is the expected number of rolls it will take to get a 7 when rolling two dice?

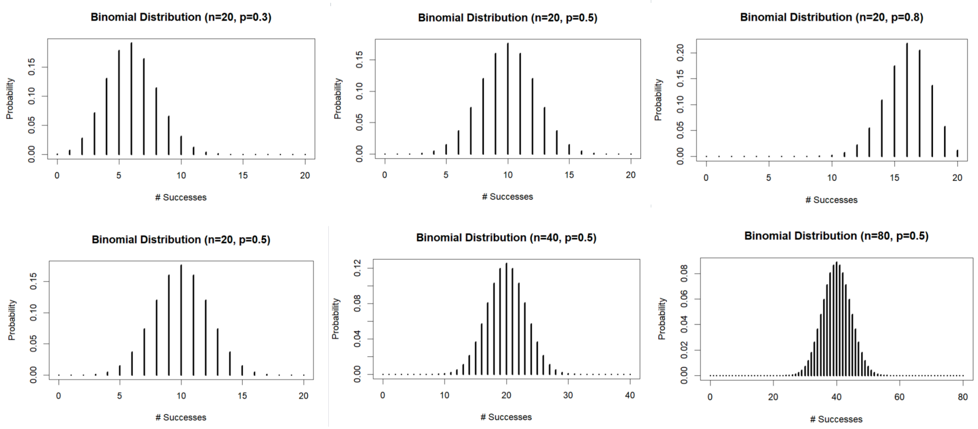
binomial random variable

- A r.v. modeling the number of successes in a fixed number of independent Bernoulli trials.
- Discrete outcomes $\{0,1,2,3\ldots,n\}$
- Two parameter
  - p : probability of a success for each trial
  - n : number of trials
- Shorthand notation: $X \sim \text{Binom}(n,p)$
- $E(X) = np,\ V(X) = np(1-p)$

$$P(X = x \,|\, n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

X	$P(X=x)$
0	$\binom{n}{0} p^0 (1-p)^n$
1	$\binom{n}{1} p^1 (1-p)^{n-1}$
2	$\binom{n}{2} p^2 (1-p)^{n-2}$
$\vdots$	$\vdots$
n-1	$\binom{n}{(n-1)} p^{n-1} (1-p)^1$
n	$\binom{n}{n} p^n (1-p)^0$

binomial random variable: parameters shaping pmf



binomial random variable

example (cont'd...)

Toss a coin 3 times: the sample space is Ω : {H,T} $\times$ {H,T} $\times$ {H,T}

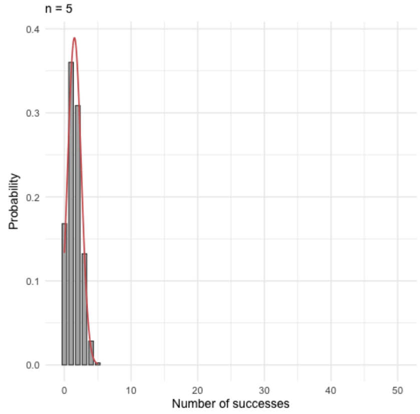
Define the random variable: $X =$ the number of heads

What is the probability distribution of X?

$$X \sim Bin(n = 3, p = 0.5)$$

$$\implies P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} 0.5^x (0.5)^{3-x}$$

linking binomial and normal



multinomial random variable

- The generalization of the binomial distribution with more than two possible outcomes
- n independent trials
- Each trial results in one of k outcomes that are mutually exclusive
- For any trial, the probabilities of the k outcomes $p_1, \ldots, p_k$ are mutually exclusive and collectively exhaustive
- Shorthand notation: $(X_1, X_2, \ldots, X_k) \sim$ Multinomial$(n, p_1, p_2, \ldots, p_k)$, where $\sum_{i=1}^{k} p_i = 1$
- $E(X_i) = np_i$, $V(X_i) = np_i(1 - p_i)$

$$P\big((X_1 = x_1) \cap \cdots \cap (X_k = x_k)\big) = \begin{cases} \dfrac{n!}{x_1! \cdots x_k!} \prod_{i=1}^{k} p_i^{x_i}, & \text{when } \sum_{i=1}^{k} x_i = n, \\ 0, & \text{otherwise}. \end{cases}$$

Poisson random variable

- A r.v. that expresses the probability of how many times an event occurs in a fixed period of time if these events
  - occur with known average rate of λ
  - and independently of each other
- Discrete outcomes $\{0,1,2,3\ldots\}$
- Shorthand notation: $X \sim$ Poisson(λ)
- $E(X) = V(X) = \lambda$

$$P(X = x \mid \lambda) = e^{-\lambda}\frac{\lambda^x}{x!}$$

X	$P(X = x)$
0	$e^{-\lambda}$
1	$e^{-\lambda}\lambda$
2	$e^{-\lambda}\frac{\lambda^2}{2}$
$\vdots$	$\vdots$

- If the data shows overdispersion (variance > mean) or underdispersion (variance < mean), other models like the Negative Binomial

linking binomial and Poisson

What if we take the limit of $P(X = x)$ as n approaches infinity?

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \longrightarrow \quad \lim_{n\to\infty} P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} = \frac{\lambda^x}{x!}e^{-\lambda}.$$

Let's replace p with $\frac{\lambda}{n}$ and $q = 1 - p$ with $1 - \frac{\lambda}{n}$ so that $P(X = x) = \binom{n}{x}\left(\frac{\lambda}{n}\right)^x\left(1 - \frac{\lambda}{n}\right)^{n-x}$.

Write out the binomial coefficient and pmf becomes:

$$P(X = x) = \frac{n(n-1)(n-2)\cdots(n-x+1)}{x!} \cdot \frac{\lambda^x}{n^x}\left(1 - \frac{\lambda}{n}\right)^{n-x}$$

→ There are exactly x factors in the first numerator!

linking binomial and Poisson

Use the power of multiplication and swap denominators between the first and second fraction:

$$P(X=x)=\left[\frac{n}{n}\cdot\frac{n-1}{n}\cdots\frac{n-x+1}{n}\right]\cdot\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^{n-x}.$$

We split up the last factor using the rules of exponents: $\left(1-\frac{\lambda}{n}\right)^{n-x}=\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-x}$, which gives

$$P(X=x)=\left[\frac{n}{n}\cdot\frac{n-1}{n}\cdots\frac{n-x+1}{n}\right]\cdot\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-x}.$$

Now take the limit as $n\to\infty$

Each factor $\frac{n-k}{n}\to 1$ for fixed k, $\quad\left(1-\frac{\lambda}{n}\right)^{n}\to e^{-\lambda}$, $\quad\left(1-\frac{\lambda}{n}\right)^{-x}\to 1$ for fixed x.

linking binomial and Poisson

So the whole expression tends to

$$P(X=x)\to\frac{\lambda^x}{x!}e^{-\lambda}\text{ which is the Poisson pmf.}$$

We showed that as $n\to\infty$ and $p=\frac{\lambda}{n}$, the Binomial pmf converges to the Poisson pmf.

negative binomial random variable

- A generalization of the geometric distribution Pascal(1,p)=Geometric(p)
  - How many trials do we need to run until we observe the r^{th} success?
  - How many failures happen before the r^{th} success occurs?
- $X=$ number of trials needed to get r successes with possible $x=r,r+1,r+2,\ldots$
- Two parameter
  - the number of successes we are waiting for
  - the probability that a single experiment gives a "success"
- Shorthand notation: $X\sim\text{NegBin}(r,p)$
- $E(X)=\frac{r}{p}$, $V(X)=\frac{r(1-p)}{p^2}$

negative binomial random variable

PMF (success-based form):

$$P(X=x\mid r,p)=\binom{x-1}{r-1}p^r(1-p)^{x-r}$$

- The last trial (the x-th) must be a success, completing the r-th success.
- The first $x-1$ trials must contain:
  - $r-1$ earlier successes
  - $x-r$ failures
- These can be arranged in $\binom{x-1}{r-1}$ different ways.
- Thus the probability is: (number of valid sequences) $\times\, p^r(1-p)^{x-r}$.

negative binomial random variable

PMF (failure-based form):

$$P(K = k \mid r, p) = \binom{k + r - 1}{k} p^r (1 - p)^k.$$

where:

- $k = x - r$ is the number of failures,
- the last trial again must be a success,
- the first $k + r - 1$ trials contain k failures and $r - 1$ successes.

negative binomial random variable

exercise 6

On a (American) roulette wheel, there are 38 spaces: 18 black, 18 red, and 2 green. You've been at the casino for a while now and decide to leave after you have won 3 bets on red. What is the probability that you leave the casino after placing exactly 5 bets on red?

so much more...