Tutorial – Mathematics for Social Scientists

Winter semester 2024/25

Probability | Discrete distributions

To do

- Weekly recap
- Real world applications
- Hands on practice
- Questions

Chapter 10 | Discrete distributions

Contingency tables and marginal distributions

from margo, marginis – ``Rand“ or border/edge

Where are the marginal distributions located?

- What is the probability of being East-German?

$$P(East - German) = \frac{65}{131}$$

$$\approx 0.496 \rightarrow 49.6\%$$

	Barbie	**Oppenheimer**	**Total**
East-German	45	20	65
West-German	17	49	66
Total	62	69	131

Contingency tables and joint distributions

Where are the joint distributions located?

- what is the probability of watching Barbie and being East-German?

$$P(Barbie \cap EG) = \frac{45}{131}$$

$$\approx 0.344 \rightarrow 34.4\%$$

	Barbie	**Oppenheimer**	**Total**
East-German	45	20	65
West-German	17	49	66
Total	62	69	131

Contingency tables and conditional distributions

Where are the conditional distributions located?

- what is the probability of watching Barbie and being East-German, **given** that the person is East-German?

$$P(Barbie \cap EG|EG) = \frac{\frac{45}{131}}{\frac{65}{131}} = \frac{45}{65}$$

$$\approx 0.692 \rightarrow \ 69.2\%$$

	Barbie	**Oppenheimer**	**Total**
East-German	45	20	65
West-German	17	49	66
Total	62	69	131

Hands on – Contingency tables

$P(WG \cap Oppie \mid WG) = \frac{\frac{49}{131}}{\frac{66}{131}} \qquad \frac{49}{\cancel{131}} \cdot \frac{\cancel{131}}{66} = \frac{49}{66}$ //

Task:

1) Find $P(WG)$ $= \frac{66}{131}$
2) Find $P(Oppenheimer)$ $= \frac{69}{131}$
3) Find the probability of being West-German and watching Oppenheimer

   $P(WG \cap Oppie) = \frac{49}{131}$
4) Find the probability of being West-German and watching Oppenheimer, given one is West-German?

   $\frac{P(A \cap B)}{P(B)}$
5) Are the events 'Oppenheimer' and 'West-German' independent? $P(A) = P(A \mid B)$

   $P(Oppie) = P(Oppie \cap WG \mid WG)$

   $\frac{66}{131} \neq \frac{49}{66}$
6) Are the events 'Oppenheimer' and 'West-German' disjoint? → mutually exclusive? $\boxed{!}$

   $P(Oppie \cap WG) = 0$?

   $= \frac{49}{131} \neq 0$

	Barbie	Oppenheimer	Total
East-German	45	20	65
West-German	17	49	66
Total	62	69	131

Hands on –contingency tables

Solution:

1) $P(WG) = \frac{66}{131} \approx 0.50$

2) $P(Oppenheimer) = \frac{69}{131} \approx 0.53$

3) $P(WG \cap Oppie) = \frac{49}{131} \approx 0.37$

4) $\frac{P(WG \cap Oppie)}{P(WG)} = \frac{\frac{49}{131}}{\frac{69}{131}} \approx 0.71$

5) $\frac{P(Oppie \cap WG)}{P(WG)} \approx 0.71 \neq P(Oppie) \approx 0.526$
→ *NOT independent*

6) $P(Oppie \cap WG) \approx 0.37 \neq 0$
→ NOT disjoint

	Barbie	**Oppenheimer**	**Total**
East-German	45	20	65
West-German	17	49	66
Total	62	69	131

Useful definitions

Random variable

- takes on different values based on outcome of **random event**
- describes **probability distribution** of possible outcomes

Random event (reminder)

- an uncertain outcome that cannot be predicted with certainty
- follows **probability distributions** and often described using likelihood

Probability distribution

- mathematical function providing **probabilities** of different **outcomes**

Parameter

- describes features/characteristics of a population/**probability distribution**
- defines shape, centre and spread of the distribution

Data generating process – DGP

The **DGP** includes all structures and dynamics of relationships & interactions that shape (**social**) phenomena

- social organization
- cultural norms and values
- power dynamics
- social institutions

→ the 'way' our data goes until we get to see an outcome, which is in turn our data ☺

stochastic component

$$Y \sim f(y|\theta)$$

with

$$\theta = g(X, \beta)$$

deterministic component

Discrete random variables

Discrete random variables…

- take on a finite or countably infinite number of values

→‘countably infinite’: values can be expressed as one-to-one mapping with positive integers

Bernoulli RV is ‘simplest’ RV

- two possible outcomes with $\mathrm{p}_1 = P(X = 1) = \frac{1}{2}$ and $\mathrm{p}_2 = P(X = 0) = \frac{1}{2}$
- individual probabilities sum up to 1

Discrete random variables – PMF & CDF

PMF of a discrete random variable X

→ yields the probability that X takes on a specific value

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

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

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

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

CDF of a discrete random variable X

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

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

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

Hands on - PMF

Task:

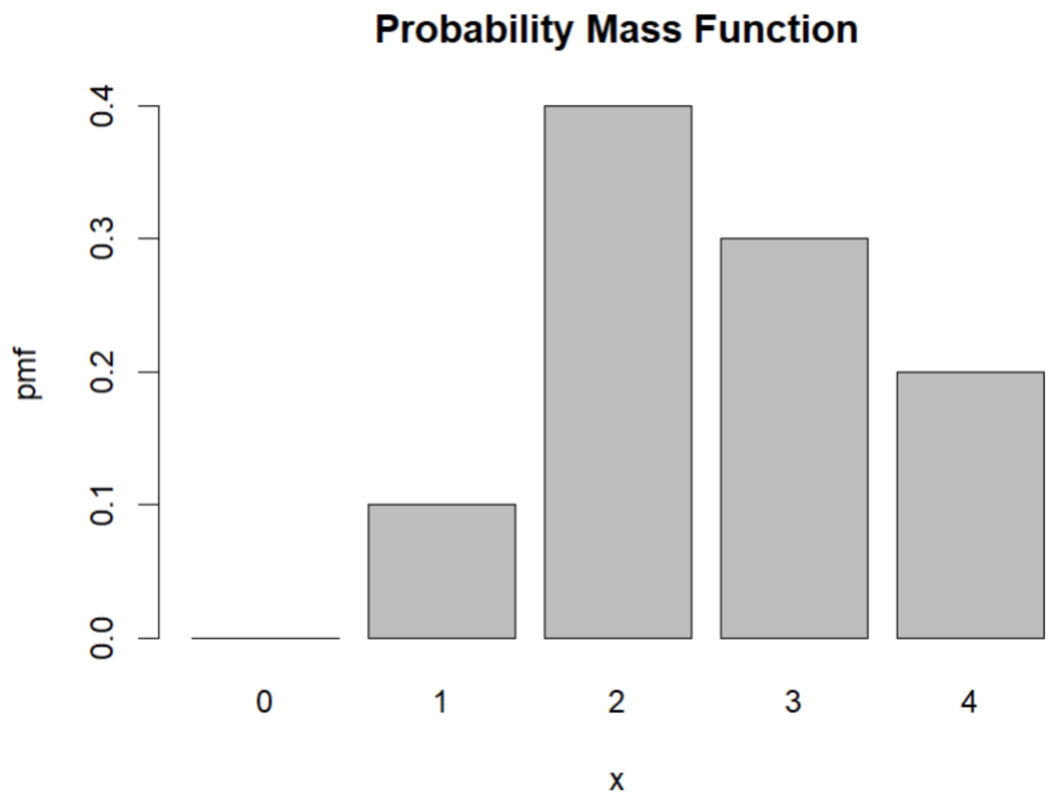
1. Sketch the PMF
2. Find $P(X = C)$

X	A	B	C	D
$P(X = x_i)$	0.1	0.4	0.3	0.2

Hands on - PMF

Solution:

1. Sketch the PMF
2. $P(X = C) = 0.3$



Expectations of RVs and variance

The expectation of a RV

- average value, weighted by probability distribution

$$E_X[X] = \sum_i x_i(P(X = x_i))$$

Example

- A die is rolled 20 times, X denotes the nr of ones that are rolled. What is E(X)?

$$E(X) = np = 20 \cdot \frac{1}{6} = 3.\overline{33}$$

- Suppose that 54% of Konstanz Uni's students are female. If we select 10 students at random, what is the expected nr of female students in this sample?

$$E(X) = np = 10 \cdot 0.54 = 5.4$$

Binomial distribution

- The number of **successes y** over **n** independent **Bernoulli trials** has a **Binomial distribution**
- **PMF**: $P(Y = y|n, p) = \binom{n}{y} p^y (1 - p)^{n-y}$

  - Bernoulli trials have two **outcomes**
    - 1 =: success
    - 0 =: failure
    - probability of success p remains constant over n trials

$$\Pr(Y = y|p) = \begin{cases} 1 - p \; for \; y = 0 \\ p \qquad for \; y = 1 \end{cases}$$

PMF: $P(Y = y|p) = p^y (1 - p)^{1-y}$

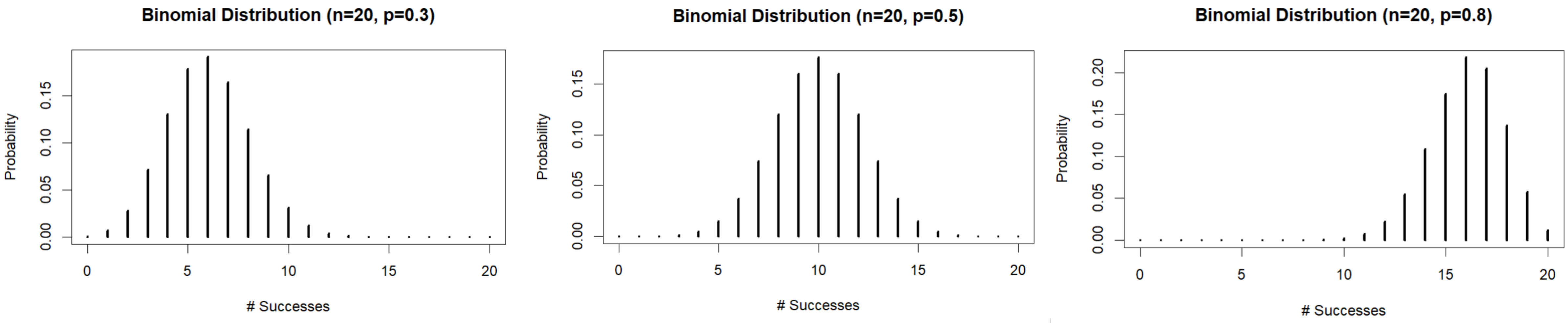
→ remember: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Moments around the mean

- The **binomial distribution** has two parameters
  - n nr of trials
  - p probability of success

- **Parameters** shape probability distributions
  - the centre of a binomial distribution shifts around the **mean**: $E(X) = np$
  - the spread of a binomial distribution is expressed via the **variance**: $Var(X) = np(1-p)$

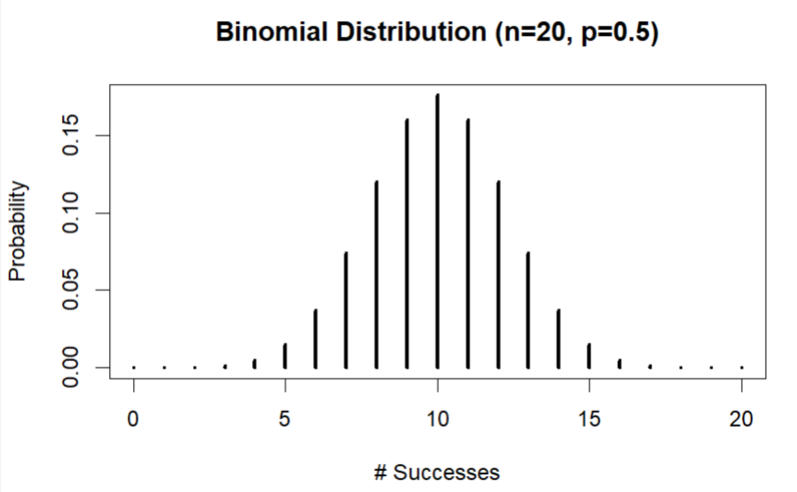
Moments around the mean

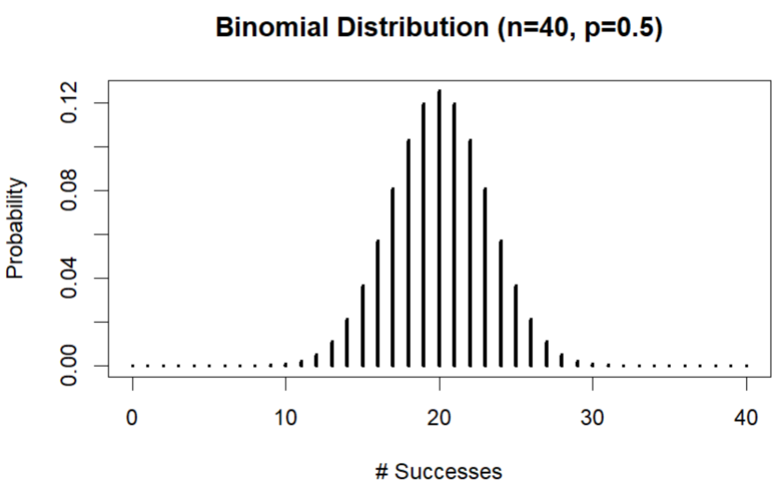


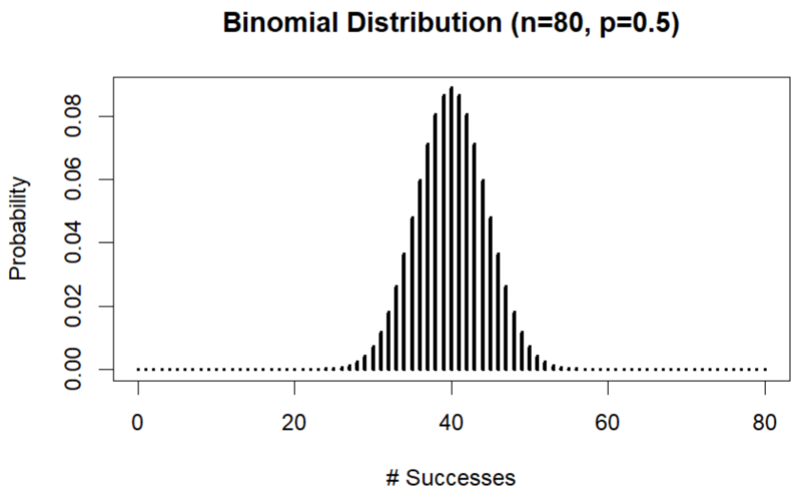
$\boldsymbol{E(X_1) = 20 \cdot 0.3 = 6}$ $\boldsymbol{E(X_2) = 20 \cdot 0.5 = 10}$ $\boldsymbol{E(X_3) = 20 \cdot 0.8 = 16}$

- notice how the spread of the distribution barely changes – but its centre shifts!
- why? n remains constant, but p increases!

Moments around the mean – variance



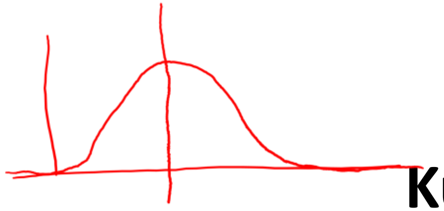




$\boldsymbol{Var(X_1) = 40 \cdot 0.5 \cdot 0.5 = 5}$ $\boldsymbol{Var(X_2) = 40 \cdot 0.5 \cdot 0.5 = 10}$ $\boldsymbol{Var(X_3) = 80 \cdot 0.5 \cdot 0.5 = 20}$

- notice the increase in n and how it manipulates the spread of the distributions while p remains constant!

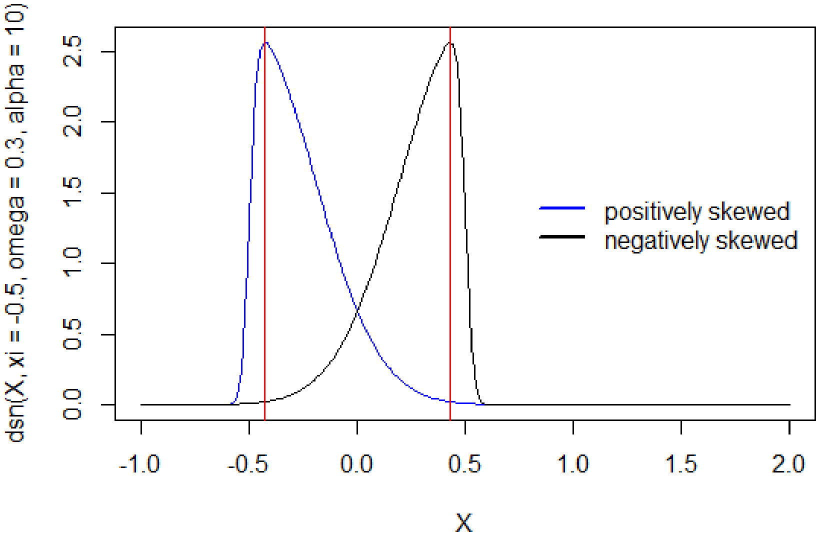
Skewness & Kurtosis



Skewness

- ‘lopsidedness’ of a distribution

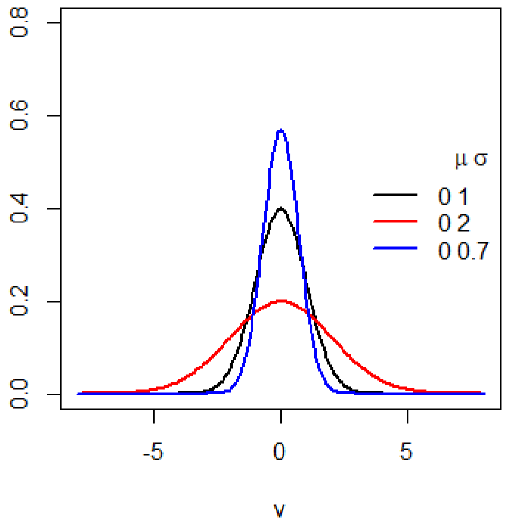
$$\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu^3}{\sigma^3}$$



Kurtosis

- peakedness/flatness of a distribution

$$\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mu^4}{\sigma^4}$$



Hands on – Binomial distribution

Task: Let y be the number of trials a five is rolled on a die after 3 rolls. What is the probability that a five is rolled twice?

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

Hints:

- find nr. of total trials $n = 3$
- find nr. of successes $y = 2$
- find probability of success $p = \frac{1}{6}$

$$P(Y=2 \mid n=3, p=\tfrac{1}{6}) = \binom{3}{2} \cdot \frac{1}{6}^2 \cdot \left(1-\frac{1}{6}\right)^{3-2}$$

total trials $n=3$

nr. successes $Y=2$

prob. of success $p=\frac{1}{6}$

$$= \binom{3}{2} \cdot \frac{1}{6}^2 \cdot \frac{5}{6}^1$$

$$= \frac{1\cdot 2\cdot 3}{1\cdot 2\cdot 1} \cdot \frac{1}{6}^2 \cdot \frac{5}{6}^1$$

$$\approx 0.0694$$

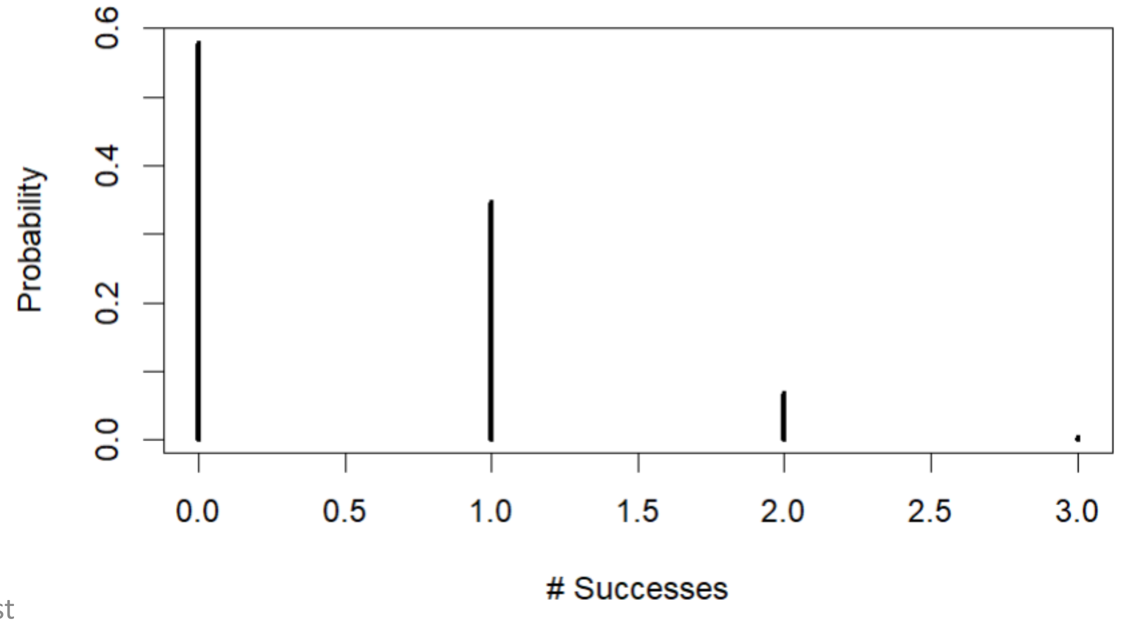
$$\approx 0.07$$

Hands on – Binomial distribution

Solution:

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

- $P\left(Y = 2\middle|3, \frac{1}{6}\right) = \binom{3}{2}\left(\frac{1}{6}\right)^2\left(1 - \frac{1}{6}\right)^{3-2}$
- $P\left(Y = 2\middle|3, \frac{1}{6}\right) = \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1$
- $P\left(Y = 2\middle|3, \frac{1}{6}\right) = 3 \cdot (0.02777778) \cdot (0.8333333)$
- $P\left(Y = 2\middle|3, \frac{1}{6}\right) = 0.06944445 \approx 0.07$



Multinomial distribution

- The generalization of the binomial distribution with more than two possible outcomes
- **PMF**: $P((Y_1 = y_1) \cap \cdots \cap (Y_k = y_k)) = \begin{cases} \frac{n!}{y_1! \ldots y_k!} \prod_{i=1}^{k} p_i^{y_i} & when \sum_{i=1}^{k} y_i = n\,, \\ 0 & otherwise. \end{cases}$
- **Properties:**
  - **n** independent **trials**
  - each trial results in one of **k outcomes** that are mutually exclusive
  - for any trial, the **probabilities** of k outcomes $\boldsymbol{p_1}, \ldots, \boldsymbol{p_k}$ are mutually exclusive and collectively exhaustive

Hands on – Multinomial distribution

20

Task: A bag is filled with 8 red balls, 3 yellow and 9 white ones. 6 balls are randomly selected with replacement. Find the probability that 2 are red, 1 is yellow and 3 are white!

$$P((Y_1 = y_1) \cap \cdots \cap (Y_k = y_k)) = \begin{cases} \frac{n!}{y_1! \ldots y_k!} \prod_{i=1}^{k} p_i^{y_i} & when \sum_{i=1}^{k} y_i = n, \\ 0 & otherwise. \end{cases}$$

Hints:

- find total nr. of trials $n = 6$
- find individual nr. of successes y_k
- find individual probabilies of success p_i

$Y = \{ Y_{red} = 2, Y_{yellow} = 1, Y_{white} = 3 \}$

$P = \{ P_{red} = \frac{8}{20}, P_{yellow} = \frac{3}{20}, P_{white} = \frac{9}{20} \}$

Hands on – Multinomial distribution

Solution:

$$P((Y_1 = y_1) \cap \cdots \cap (Y_k = y_k)) = \begin{cases} \frac{n!}{y_1! \ldots y_k!} \prod_{i=1}^{k} p_i^{y_i} & when \sum_{i=1}^{k} y_i = n, \\ 0 & otherwise. \end{cases}$$

Hints:

- find $n = 6$
- find $y = \{y_{red} = 2,\ y_{yellow} = 1,\ y_{white} = 3\}$
- find $p = \{p_{red} = \frac{8}{20}, p_{yellow} = \frac{3}{20}, p_{white} = \frac{9}{20}\}$

- $P\left((Y_{red} = 2) \cap \left(Y_{yellow} = 1\right) \cap (Y_{white} = 3)\right) = \frac{6!}{2!1!3!}\left(\frac{8}{20}\right)^2\left(\frac{3}{20}\right)^1\left(\frac{9}{20}\right)^3$
- $P\left((Y_{red} = 2) \cap \left(Y_{yellow} = 1\right) \cap (Y_{white} = 3)\right) = 0.13122 \approx 0.13$

Poisson distribution

- The **Poisson distribution** depicts the probability of a number of events occurring in a fixed interval of time

- **PMF**: $P(Y = y|\overset{\lambda}{\mu}) = \frac{\mu^y}{y!\, e^{\mu}} = e^{-\mu}\frac{\mu^y}{y!}$ → $e^{-\mu} = \frac{1}{e^{\mu}}$

- **Properties**:
  - Parameters: $\mu = E(X) = Var(X)$ ← μ is often expressed using λ
  - mean rate of occurrence is constant
  - events occur independently from each other

  → we have no n of total 'trials' anymore… how large is n?

Hands on – Poisson distribution

Task: Find $P(Y = 5)$ given a Poisson random variable where $\mu = 3$

$$P(Y = y|\mu) = \frac{\mu^y}{y!\ e^{\mu}}$$

Hands on – Poisson distribution

Task: Find $P(Y = 5)$ given a Poisson random variable where $\mu = 3$

$$P(Y = y|\mu) = \frac{\mu^y}{y!\ e^{\mu}}$$

Solution:

$$P(Y = 5|\mu = 3) = \frac{\mu^y}{y!\, e^{\mu}} = \frac{3^5}{5!e^3} \approx 0.1008 = 10.08$$

Binomial and Poisson distribution – a link!

- What if we take the limit of $P(Y)$ as n approaches infinity?

$$P(Y) = \binom{n}{y} p^y (1-p)^{n-y} \rightarrow \lim_{n\to\infty} P(Y=y) = \frac{e^{-\mu}\mu^y}{y!} = \frac{\mu^y}{y!\, e^{\mu}}$$

- let's replace p with $\frac{\mu}{n}$ and $q = 1 - p$ with $1 - \frac{\mu}{n}$

$$P(Y) = \binom{n}{y} \left(\frac{\mu}{n}\right)^y \left(1 - \frac{\mu}{n}\right)^{n-y}$$

- write out the binomial coefficient

$$P(Y) = \frac{n(n-1)(n-2)\ldots(n-y+1)}{y!} \cdot \frac{\mu^y}{n^y}\left(1 - \frac{\mu}{n}\right)^{n-y}$$

→ there are exactly y factors in the first numerator!

Binomial and Poisson distribution – a link!

→ use the power of multiplication and swap denominators between first and second fraction!

$$P(Y) = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \ldots \cdot \frac{n-y+1}{n} \cdot \frac{\mu^y}{y!}\left(1-\frac{\mu}{n}\right)^{n-y}$$

- we split up the last factor into two using the rules of exponents

$$P(Y) = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \ldots \cdot \frac{n-y+1}{n} \cdot \frac{\mu^y}{y!}\left(1-\frac{\mu}{n}\right)^{n}\left(1-\frac{\mu}{n}\right)^{-y}$$

→these factors approach one!

→this we recognize already as part of the Poisson distribution! $P(Y) = \frac{e^{-\mu}\mu^y}{y!}$

→this factor approaches $e^{-\lambda}$!

Binomial and Poisson distribution – a link!

We end up with the special case of the binomial distribution as n approaches infinity: the Poission distribution!

$$P(Y) = \binom{n}{y} p^y (1-p)^{n-y} \rightarrow \lim_{n \to \infty} P(Y = y) = \frac{e^{-\mu} \mu^y}{y!}$$

Negative binomial distribution

Describes the number of independent and identically distributed **Bernoulli trials** needed to achieve a fixed number of failures before a specified number of successes

Based on failures:

- PMF: $P(X = k|r, p) = \binom{k+r-1}{k} p^r (1-p)^k$

Based on successes:

PMF: $P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$

With:

- **n** independent **trials**
- **r successes** – experiment stops once yth success occurred
- **k failures** before rth success occurs
- constant **probability** of success **p** per trial

Negative binomial distribution

Based on successes:

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

- **Mean**:

$$E(X) = \frac{r}{p}$$

Based on failures:

$$P(X = k|r,p) = \binom{k+r-1}{k} p^r (1-p)^k$$

- **Mean**:

$$E(X) = \frac{r}{p} - r = \frac{r(1-p)}{p}$$

- **Variance for both**: $Var(X) = \frac{r(1-p)}{p^2}$

Note: often used to deal with overdispersed count data, where the variance is greater than that of a Poisson distribution!

Hands on - Negative Binomial distribution

Task: Given a bag with $\overbrace{8 \text{ red balls and } 12}^{20}$ white balls, what is the probability that a red ball is pulled for the 4th time on the 10th draw, assuming replacement?

$$P(X = k|r,p) = \binom{k+r-1}{k} p^r (1-p)^k$$

Hints:

- find total number of **cases n**
- find number of **failures k**
- find number of **successes r**
- find **probability** of **success p** – what does replacement mean for us?

nr of total trials $n = 10$

nr of failures $k = n - r = 10 - 4 = 6$

nr of successes $r = 4$

prob. of success $p = \frac{8}{20}$

$$P(X = 6 \mid r = 4, p = \frac{8}{20}) = \binom{6+4-1}{6} \cdot \left(\frac{8}{20}\right)^4 \cdot \left(1 - \frac{8}{20}\right)^6$$

$$= \binom{9}{6} \cdot \left(\frac{8}{20}\right)^4 \cdot \left(\frac{12}{20}\right)^6$$

$$= \frac{9!}{6!(9-6)!} \cdot 0.4^4 \cdot 0.6^6$$

$$= \frac{9!}{6!3!} \cdot 0.4^4 \cdot 0.6^6$$

$$\approx 0.1003 \approx 10.03$$

Hands on - Negative Binomial distribution

Solution:

- number of **total cases** $\boldsymbol{n} = 10$
- number of **successes r** $= 4$
- number of **failures** $\boldsymbol{k} = 6 = 10 \ - 4$
- **probability** of **success p =** $\frac{8}{20} = 0.4$

$$P(X = k|r, p) = \binom{k+r-1}{k} p^r (1-p)^k$$

$$P(X = 6|r = 4, p = 0.4) = \binom{6+4-1}{6} 0.4^4 (1-0.4)^6$$

$$P(X = 6|r = 4, p = 0.4) = \binom{9}{6} 0.4^4 (0.6)^6$$

$$= \frac{9!}{6! \cdot 3!} \cdot 0.0256 \cdot 0.046656$$

$$\approx 0.1003$$

$$\approx 10.03$$

with: $\binom{n}{k} = \frac{n!}{k!(n-k)!} \rightarrow \binom{9}{6} = \frac{9!}{6!(9-6)!}$

→ replacement: probability stays constant!

Time for your questions

- Any questions during the week?
  - joerdis.strack@uni-konstanz.de