Tutorial – Mathematics for Social Scientists

Winter semester 2024/25

Continuous distributions

To do

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

- Upcoming Deadline: 14.01.2024 – Assignment 3 | Probability

Chapter 11 | Continuous distributions

Continuous distributions – PDF and CDF

PDF

- ‘function describing the smooth curve that connects the various probabilities of a specific values of a sample’ ← Moore and Siegel, 2013
- density of the probability within range instead of the ‘mass’ of probability at a particular value

CDF

- the CDF of a real-valued RV X accumulates the probabilities of X being smaller than x when evaluated at x

Continuous distributions PDF and CDF

PDF

- consider an integral – continuous analogue to ‘sums‘
- no area in a line – so no probability assigned to RV taking on a specific value

$f(x) \geq 0, for\ all\ x \in \mathbb{R}$

f is piecewise continuous

$\int_{-\infty}^{\infty} f(x)dx = 1$

$P(a \ \leq X \leq b) = \int_{a}^{b} f(x)dx$

CDF

- CDF is found by integrating the PDF
- PDF is found by differentiating the CDF
- the CDF is always non-decreasing

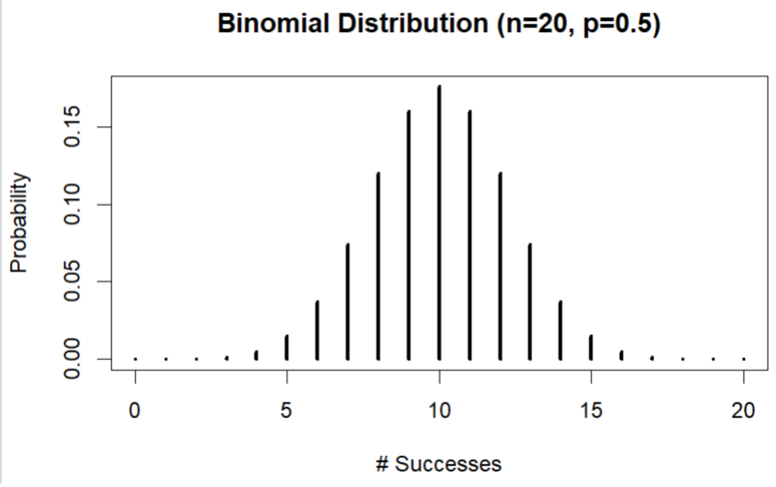
$F(x) = \int_{-\infty}^{x} f(y)dy$ for $-\infty < x < \infty$

- Why do we use y instead of x?

PDF and PMF – what‘s the difference?

PMF – Probability Mass Function

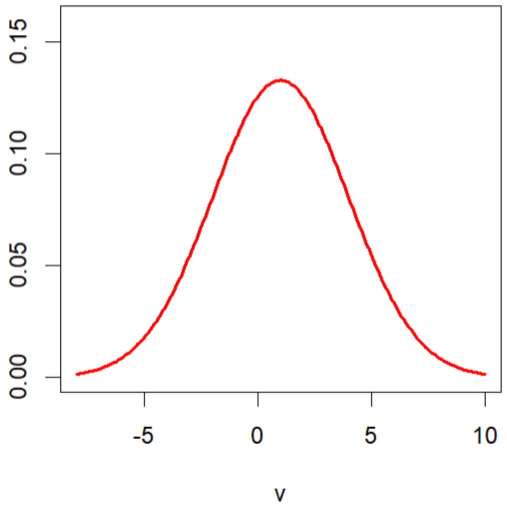
- individual spikes/peaks for countable # of events



PDF – Probability Density Function

- one smooth function for pot. ∞ # of events

Standard Normal Distribution μ = 1 & sd = 3



Some continuous distributions

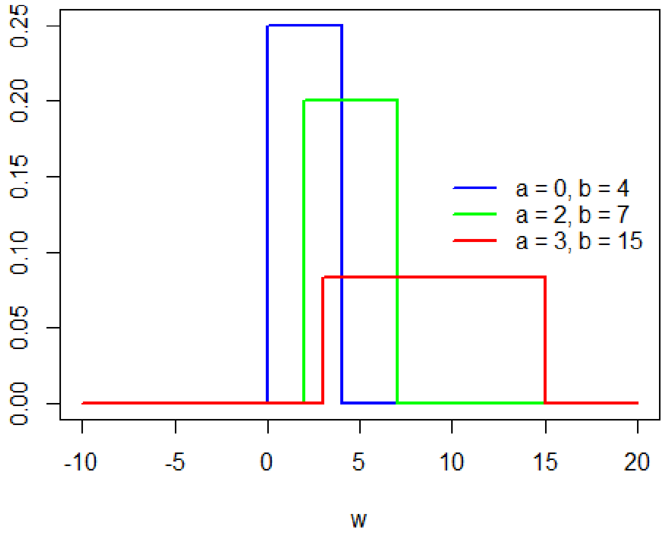
We will not cover every distribution listed in the book in the tutorial.

Therefore, you are required to read up on the leftovers in the book yourself.

Uniform distribution

The uniform distribution is a special case of the beta distribution, where $\alpha = 1$ and $\beta = 1$

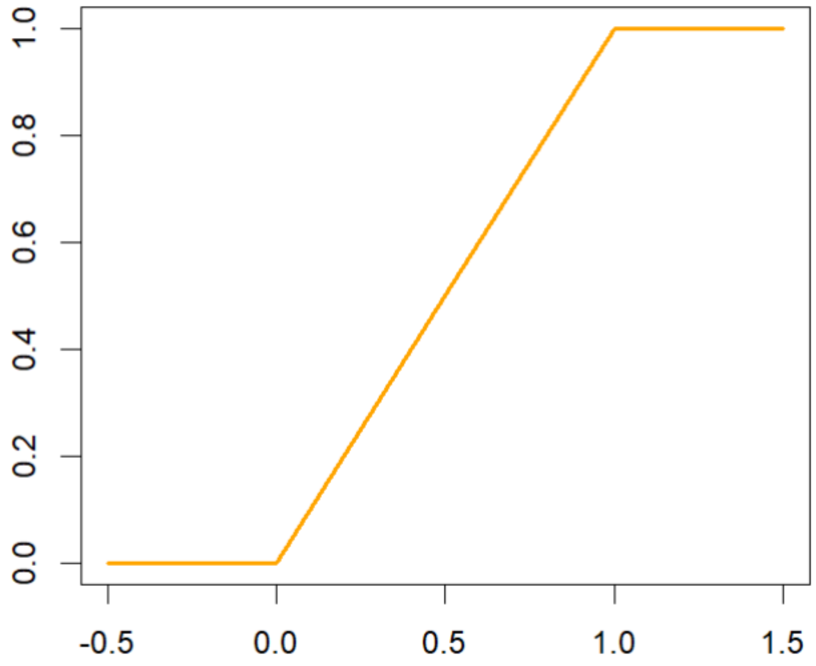
- suited for scenarios where there‘s equal probability for all events in sample space
- **PDF**: $f(x|\alpha,\beta) = \begin{cases} \frac{1}{\beta-\alpha} & if \quad x \in [\alpha,\beta] \\ 0 & otherwise \end{cases}$
- **mean** $= \frac{a+b}{2}$
- **variance** $= \frac{(b-a)^2}{12}$



Uniform distribution

The uniform distribution is a special case of the beta distribution, where $\alpha = 1 \;\; and \;\; \beta = 1$

- **CDF**: $$F(x) = \begin{cases} 0 & \text{if} \quad x < \alpha, \\ \frac{x-\alpha}{\beta-\alpha} & \text{if} \quad x \in [\alpha, \beta], \\ 1 & \text{if} \quad x > \beta. \end{cases}$$



Hands on – Uniform distribution

Task: Let $X = [0, 60]$ be a random variable denoting minutes per hour. It is polling day and a person is travelling by bus to the polling station. Busses leave once every hour at the same time. Make a sketch!

1) find the probability that the person will have to wait fewer than 24 minutes!
2) find the probability that the person will have to wait between 20 and 40 minutes!
3) how long does the person have to wait on average?

Hints:

mean $= \frac{a+b}{2}$ and $$f(x|\alpha,\beta) = \begin{cases} \frac{1}{\beta-\alpha} & if \quad x \in [\alpha,\beta] \\ 0 & otherwise \end{cases}$$

Hands on – Uniform distribution

Solution: $X = [0, 60]$

1) find the probability that the person will have to wait fewer than 24 minutes!

$$P(X \leq x) = \frac{1}{\beta - \alpha} \cdot x \rightarrow P(X \leq 24) = \frac{1}{60-0} \cdot 24 = \frac{24}{60} \rightarrow \frac{24}{60} \cdot 100\% = 40\%$$

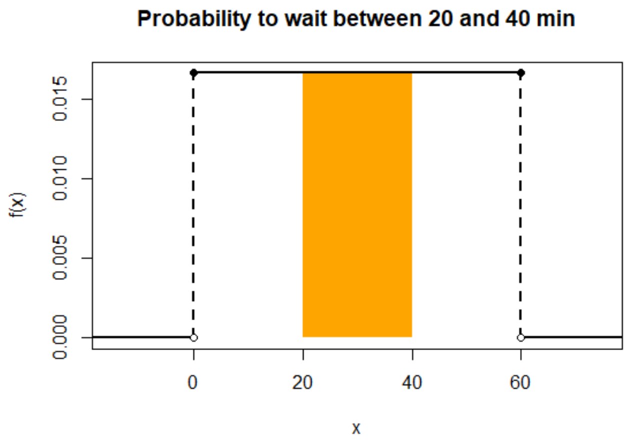
2) find the probability that the person will have to wait between 20 and 40 minutes!

$$P(20 \leq x \leq 40) = \frac{1}{60-0} \cdot 40 - \frac{1}{60-0} \cdot 20 = \frac{20}{60} \rightarrow \frac{20}{60} \cdot 100\% = 33,\overline{33}\%$$

3) how long does the person have to wait on average?

$$E[X] = \frac{\alpha + \beta}{2} \rightarrow E[X] = \frac{0+60}{2} = 30$$

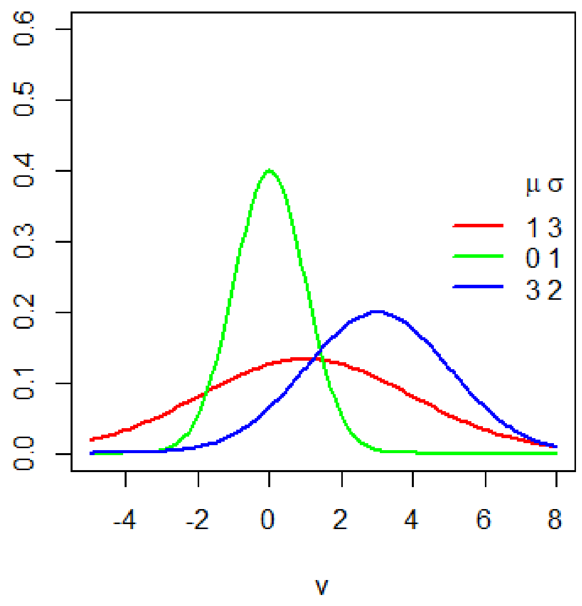
→ the person will have to wait 30 minutes on average



The Normal distribution

One of the most used distributions among the Gaussian distribution family

- normality is a crucial assumption for most of our hypotheses tests!
- **PDF**: $f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- **CDF**: … cannot be compared sensibly – we must standardize it!



Normal transformations

Standard normal distribution

- **mean** = 0
- **variance** = 1
- **PDF**: $f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}$
- **z-scores** $= \frac{x-\mu}{\sigma}$
  - positive → $x > \mu$
  - negative → $x < \mu$
  - $z = \mu$ → $\mathrm{x} = \mu$

log-normal distribution

- **mean** = $= \mathrm{e}^{(\mu+\frac{1}{2}\sigma^2)}$
- **variance** = $(e^{\sigma^2} - 1)\, e^{2\mu+\sigma^2}$
- **PDF**: $f(x \mid \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \mathrm{e}^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$

- NOTE: $0 \leq x \leq \infty$

→useful for right-skewed and asymmetric data, e.g., income

Standard normal distribution – Z-tables

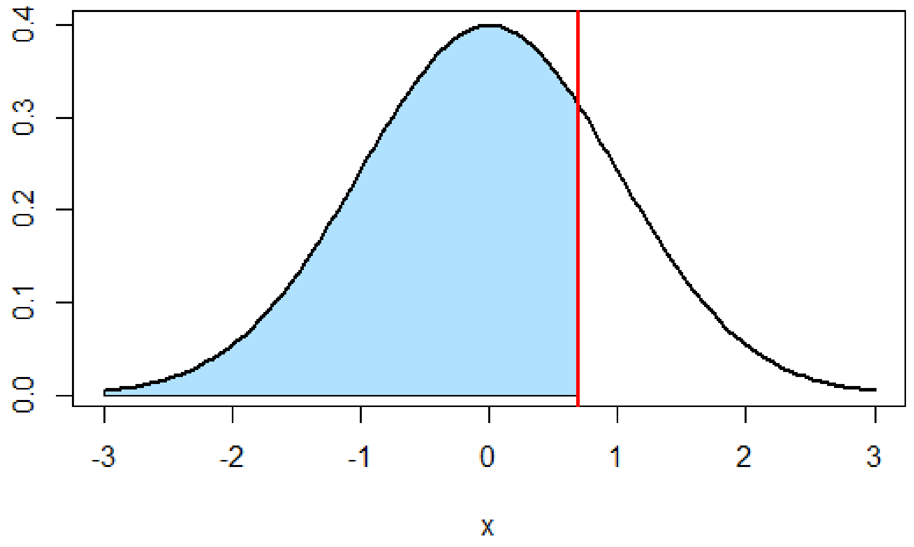
Example: $\frac{5.5 - 2}{5} = 0.7$

- note: z is positive → $x > \mu$
- we check the z-score for z =0.7 and 0 df
- 75.8% of the distribution lie below X = 5.5

→ the blue-shaded area

TABLE G.1 (Continued)

z	0	1	2
−0.4	0.3446	0.3409	0.3372
−0.3	0.3821	0.3783	0.3745
−0.2	0.4207	0.4168	0.4129
−0.1	0.4602	0.4562	0.4522
−0.0	0.5000	0.4960	0.4920
0.0	0.5000	0.5040	0.5080
0.1	0.5398	0.5438	0.5478
0.2	0.5793	0.5832	0.5871
0.3	0.6179	0.6217	0.6255
0.4	0.6554	0.6591	0.6628
0.5	0.6915	0.6950	0.6985
0.6	0.7257	0.7291	0.7324
0.7	0.7580	0.7611	0.7642
0.8	0.7881	0.7910	0.7939
0.9	0.8159	0.8186	0.8212
1.0	0.8413	0.8438	0.8461
1.1	0.8643	0.8665	0.8686



Wooldridge, 2019 p. 832

Hands on – (Standard) normal distribution

Task: **z-scores** $= \frac{x-\mu}{\sigma}$

- positive → $x > \mu$
- negative → $x < \mu$
- $z = \mu$ → $\mathrm{x} = \mu$

1) Suppose, $\mu = 20\ and\ \sigma = 5$, what is the percentage of the distribution that is above X = 30?
2) Find $P(Z \leq 1.27), P(0.82 \ \leq X \leq 1.27), P(Z \geq 1.27)$

TABLE G.1 (Continued)

z	0	1	2	3	4	5	6	7	8	9
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Wooldridge, 2019 p. 832

Hands on – (Standard) normal distribution

Solution: **z-scores** $= \frac{x-\mu}{\sigma}$

1) Suppose, $\mu = 20 \; and \; \sigma = 5$, what is the percentage of the distribution that is above X = 30?

$z = \frac{30-20}{5} = 2.00$ → refer to Z-table to find $P(X \leq 30) = 0.9772$

$P(X \geq 30) = 1 - P(X \leq 30) = 1 - 0.9772 = 0.0228 \rightarrow 0.0228 \cdot 100\%$

$= 2.28\%$

→ 2.28% of the distribution are above X = 30

Hands on – (Standard) normal distribution

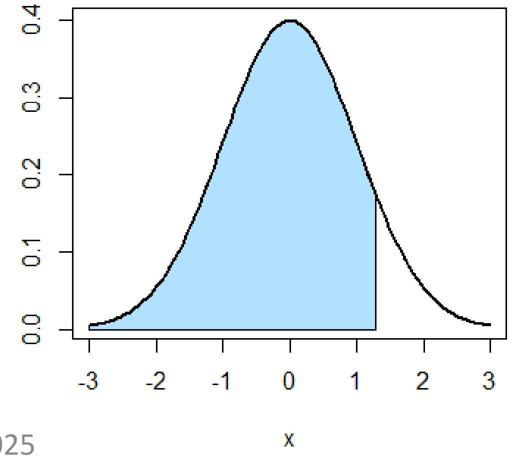
Solution: **z-scores** $= \frac{x-\mu}{\sigma}$

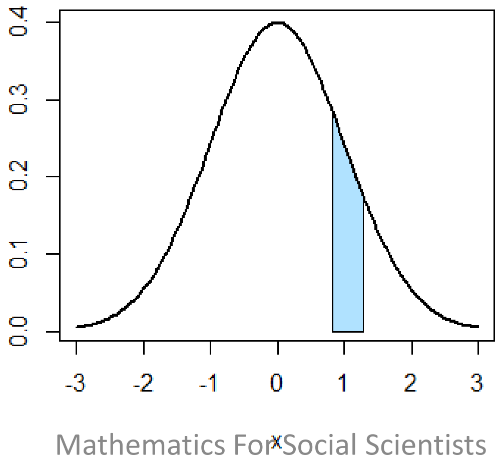
$P(Z \leq 1.27) = 0.8980$

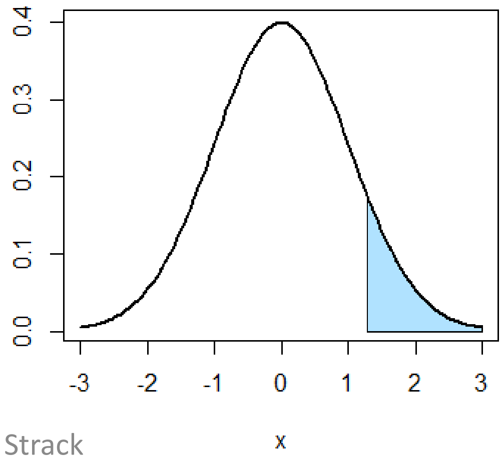
$P(0.82 \leq X \leq 1.27) = 0.8980 - 0.7939 = 0.1041$

$P(Z \leq 0.82) = 0.7939$

$P(Z \geq 1.27) = 1 - 0.8980 = 0.1020$







Central limit theorem – CLT

Regardless of the shape of a population‘s distribution, if the sample size n is large enough ($n \geq 30$) and there is finite variance, then…

- the distribution of the sample means will be approx. normal

→shape of distribution of X becomes more bell-shaped and symmetric

- centre of the distribution of $\bar{X}$ remains μ
- the spread of the distribution increases and it becomes more ‘peaked’

Law of large numbers – LLN

The law of large numbers states that for an increasing number of trials, the sample average should approach the population average

- guarantee for long-term and stable results of random events
- necessary for statistical modelling → remember asymptotic normality, efficiency and consistency from Statistics / RD II!

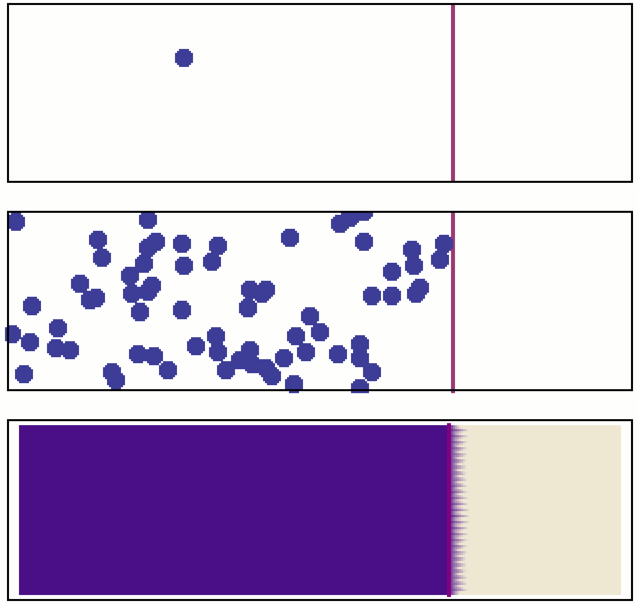
Exceptions:

- samples from Cauchy and some Pareto distributions ($\alpha<1$) may not converge as n becomes increases!

→ often due to is heavy tails [Skewness… ☺]

Real world applications – LLN & Diffusion

- diffusion is a process, in which different solutes get ‘mixed’
- the solutes consist of singular molecules, which move through a container
  - at first the two solutes will not appear very mixed
  - → the molecules move through the container in random
  - but with more time, the solutes will become visibly more difficult to distinguish from each other fluctuations
  - →movement becomes more uniformly until the solutes have ‘mixed’
- in reality, chemists can actually make predictions about diffusion behaviour



https://en.wikipedia.org/wiki/Law_of_large_numbers, 04.10.2023

χ^2 distribution

The χ^2 distribution is another special case of the gamma distribution, where $\alpha = \frac{n}{2}$ and $\beta = 2$. It is used to test if categorical data are independent/significantly different from expectation

- **PDF**: $f(x|n) = \begin{cases} \dfrac{x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{\pi}{2}} \Gamma\left(\frac{n}{2}\right)} & for\ x \geq 0, \\ 0 & for\ x < 0 \end{cases}$

- test statistic $\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$
- mean of individual cells: $E_i = \frac{row\ r\ total \cdot column\ c\ total}{N}$

TABLE G.4 Critical Values of the Chi-Square Distribution

		Significance Level		
		.10	.05	.01
Degrees of Freedom	1	2.71	3.84	6.63
	2	4.61	5.99	9.21
	3	6.25	7.81	11.34
	4	7.78	9.49	13.28
	5	9.24	11.07	15.09
	6	10.64	12.59	16.81
	7	12.02	14.07	18.48
	8	13.36	15.51	20.09
	9	14.68	16.92	21.67
	10	15.99	18.31	23.21
	11	17.28	19.68	24.72
	12	18.55	21.03	26.22
	13	19.81	22.36	27.69
	14	21.06	23.68	29.14
	15	22.31	25.00	30.58
	16	23.54	26.30	32.00
	17	24.77	27.59	33.41
	18	25.99	28.87	34.81
	19	27.20	30.14	36.19
	20	28.41	31.41	37.57
	21	29.62	32.67	38.93
	22	30.81	33.92	40.29
	23	32.01	35.17	41.64
	24	33.20	36.42	42.98
	25	34.38	37.65	44.31
	26	35.56	38.89	45.64
	27	36.74	40.11	46.96
	28	37.92	41.34	48.28
	29	39.09	42.56	49.59
	30	40.26	43.77	50.89

Wooldridge, 2019, p.837

Hands on – χ^2 distribution

Task: You hypothesize that participation in reafforestation projects depends on the Land (Bundesland) one lives in. Carry out a χ^2-test for independence at the 95% significance level to test, if participation is independent from residence!

Land	**Bavaria**	**Lower Saxony**	**Bremen**	**Ba-Wü**	**Hessen**	**Total**
Participates	30	55	22	42	37	186
Does NOT participate	40	48	26	30	38	182
Total	70	103	48	72	75	368

Hands on – χ^2 distribution

Hints:

Land	**Bavaria**	**Lower Saxony**	**Bremen**	**Ba-Wü**	**Hessen**	**Total**
Participates	30	55	22	42	37	186
Does NOT participate	40	48	26	30	38	182
Total	70	103	48	72	75	368

- find k and degrees of freedom $= (r-1)(c-1) = df$

→find critical value $\chi^2_{0.05}(df)$ for 95% significance level

- compute test-statistic TS χ^2 by computing the cell's means E_i

$$E_i = \frac{row\ r\ total \cdot column\ c\ total}{N}$$

- compare TS with $\chi^2_{0.05}(df)$

→ if TS $> \chi^2_{0.05}\ (df)$ we can reject the H_0 and assume significant changes!

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Hands on – χ^2distribution

Solution:

- $(5-1)(2-1) = 4 \cdot 1 = 4\, df$
- critical value: $\chi^2_{0.05}(4) = 9.49$
- compute E_i using $E_i = \frac{row\ r\ total \cdot column\ c\ total}{N}$

Land	**Bavaria**	**Lower Saxony**	**Bremen**	**Ba-Wü**	**Hessen**	**Total**
Participates	$\frac{70 \cdot 186}{368}$	$\frac{103 \cdot 186}{368}$	$\frac{48 \cdot 186}{368}$	$\frac{72 \cdot 186}{368}$	$\frac{75 \cdot 186}{368}$	186
Does NOT participate	$\frac{70 \cdot 182}{368}$	$\frac{103 \cdot 182}{368}$	$\frac{48 \cdot 182}{368}$	$\frac{72 \cdot 182}{368}$	$\frac{75 \cdot 182}{368}$	182
Total	70	103	48	72	75	368

Hands on – χ^2 distribution

Solution:

- TS $\chi^2 = \frac{(30-35.38)^2}{35.38} + \frac{(55-52.06)^2}{52.06} + \frac{(22-24.26)^2}{24.26} + \frac{(42-36.39)^2}{36.39} + \frac{(37-37.91)^2}{37{,}91} + \frac{(40-34{,}62)^2}{34{,}62} + \frac{(48-50.94)^2}{50.94} + \frac{(26-23.74)^2}{23.74} + \frac{(30-35.61)^2}{35.61} + \frac{(38-37.09)^2}{37.09} = 4.208383 \approx 4.21$

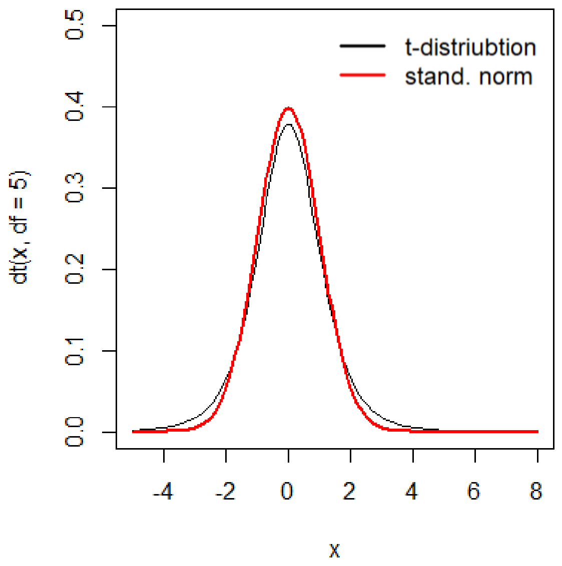
- $\chi^2_{0.05}(4) = 9.49 > 4.21 = TS\ \chi^2$

→ since our TS < 9.49 we fail to reject the null hypothesis – participation in reafforestation projects appears to be independent from one's Bundesland of residence.

t-Distribution

The t-distribution looks similar to the normal distribution but has thicker tails! This is useful for hypothesis testing under small sample sizes

- **PDF**: $f(x|n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$



- **test statistic**:
  - one-sample t-test
    $t = \frac{X - \Delta}{SE(\widehat{\beta})}$ with $df = n - k - 1$ with k # of variables
  - where Δ is the mean-difference postulated by Null-Hypothesis

Hands on - t-distribution

Task: Perform a two-tailed t-test to check, if the elasticity of nox is different from -1 at 95% significance level (Wooldridge, 2019, pp. 133)

Model:

$$\widehat{\log(price)} = \underset{(0.32)}{11.08} - \underset{(.117)}{.954}\log(nox) - \underset{(.043)}{.134}\log(dist) + \underset{(.019)}{.255}\,rooms - \underset{(.006)}{.052}\,stratio$$

$$n = 506,\ R^2 = .581.$$

Hints:

- find df
- find t_{df}
- compute t and compare to t_{df}
- $t = \dfrac{X - \Delta}{SE(\widehat{\beta})}$

TABLE G.2 Critical Values of the *t* Distribution

	Significance Level				
1-Tailed:	.10	.05	.025	.01	.005
2-Tailed:	.20	.10	.05	.02	.01
Degrees of Freedom 1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
90	1.291	1.662	1.987	2.368	2.632
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

Wooldridge, 2019, p.833

Hands on - t-distribution

Solution: (Wooldridge, 2019, pp. 133)

$$\widehat{\log(price)} = 11.08 - .954 \log(nox) - .134 \log(dist) + .255\, rooms - .052\, stratio$$
$$(0.32)\ (.117) \qquad (.043) \qquad (.019) \qquad (.006)$$
$$n = 506,\ R^2 = .581.$$

1) find $df = 506 - 4 - 1 = 501$
2) find $t_{df} = 1.96$
3) compute and compare t with t_{df}

- $t = \frac{X - \Delta}{\sqrt{\sigma_x}} = \frac{-0.954-(-1)}{0.117} = 0.393$
- $t_{df} = 1.96 > t = 0.393$

→ we fail to reject the Null-Hypothesis, the elasticity of nox is unlikely different from -1

TABLE G.2 Critical Values of the t Distribution

		Significance Level		
1-Tailed:		.10	.05	.025
2-Tailed:		.20	.10	.05
	1	3.078	6.314	12.706
	2	1.886	2.920	4.303
	3	1.638	2.353	3.182
	4	1.533	2.132	2.776
	5	1.476	2.015	2.571
	6	1.440	1.943	2.447
	7	1.415	1.895	2.365
	8	1.397	1.860	2.306
	9	1.383	1.833	2.262
Degrees of Freedom	10	1.372	1.812	2.228
	11	1.363	1.796	2.201
	12	1.356	1.782	2.179
	13	1.350	1.771	2.160
	14	1.345	1.761	2.145
	15	1.341	1.753	2.131
	16	1.337	1.746	2.120
	17	1.333	1.740	2.110
	18	1.330	1.734	2.101
	19	1.328	1.729	2.093
	20	1.325	1.725	2.086
	21	1.323	1.721	2.080
	22	1.321	1.717	2.074
	23	1.319	1.714	2.069
	24	1.318	1.711	2.064
	25	1.316	1.708	2.060
	26	1.315	1.706	2.056
	27	1.314	1.703	2.052
	28	1.313	1.701	2.048
	29	1.311	1.699	2.045
	30	1.310	1.697	2.042
	40	1.303	1.684	2.021
	60	1.296	1.671	2.000
	90	1.291	1.662	1.987
	120	1.289	1.658	1.980
	∞	1.282	1.645	$\boxed{1.960}$

Wooldridge, 2019, p.833

F-Distribution

The ratio of two RV, each distributed according to a chi-squared distribution and scaled to respective degrees of freedom

→ used for hypothesis testing to test, if several variables are not jointly zero/insignificant

- **PDF**: $f(x|n_1, n_2) = \frac{\sqrt{\frac{(n_1 X)^{n_1} n_2^{n_2}}{(n_1 x + n_2)^{n_1+n_2}}}}{xB\left(\frac{n_1}{2}, \frac{n_2}{2}\right)}$ where $B(x|y) = \frac{\Gamma(x)\Gamma(\mathrm{y})}{\Gamma(x+y)}$ is the beta function
- **test statistic**: $F = \frac{\frac{SSR_r - SSR_{ur}}{q}}{\frac{SSR_{ur}}{(n-k-1)}}$ where
  - $q =$ numerator df $= df_r - df_{ur}$ and
  - $n - k - 1 =$ denominator df $= df_{ur}$

Hands on – F distribution

Task: Perform an F-test using these two models (Wooldridge, 2019, pp. 144)

unrestricted:
$$\widehat{\log(salary)} = \underset{(0.29)}{11.19} + \underset{(.0121)}{.0689}\, years + \underset{(.0026)}{.0126}\, gamesyr + \underset{(.00110)}{.00098}\, bavg + \underset{(.0161)}{.0144}\, hrunsyr + \underset{(.0072)}{.0108}\, rbisyr$$
$$n = 353,\ \mathrm{SSR} = 183.186,\ R^2 = .6278,$$

restricted:
$$\widehat{\log(salary)} = \underset{(.11)}{11.22} + \underset{(.0125)}{.0713}\, years + \underset{(.0013)}{.0202}\, gamesyr$$
$$n = 353,\ \mathrm{SSR} = 198.311,\ R^2 = .5971.$$

- **test statistic**: $F = \dfrac{\frac{SSR_r - SSR_{ur}}{q}}{\frac{SSR_{ur}}{(n-k-1)}}$ where
  - $q =$ numerator df $= df_r - df_{ur}$ and
  - $n - k - 1 =$ denominator df $= df_{ur}$

TABLE G.3b 5% Critical Values of the *F* Distribution

Denominator Degrees of Freedom	Numerator Degrees of Freedom: 1	2	3	4	5	6	7	8	9	10
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

Wooldridge, 2019, p.835

Hands on – F distribution

Solution: (Wooldridge, 2019, pp. 144)

1) compute numerator/ denominator df!

$$n - k - 1 = 353 - 5 - 1 = 347$$

$$q = \#\, variables\ excluded = 3$$

2) compute F statistic

$$F = \frac{\frac{198.311 - 183.186}{3}}{\frac{183.186}{347}} = \frac{198.311 - 183.186}{183.186} \cdot \frac{347}{3} \approx 9.55$$

3) find/compare to critical value

$$F_{0.05} = 2.60 < F = 9.55$$

→ we can reject the Null-Hypothesis, the set of variables appears to be jointly significant

TABLE G.3b 5% Critical Values of the F Distr

Denominator Degrees of Freedom	Num 1	2	3	4
10	4.96	4.10	3.71	3.48
11	4.84	3.98	3.59	3.36
12	4.75	3.89	3.49	3.26
13	4.67	3.81	3.41	3.18
14	4.60	3.74	3.34	3.11
15	4.54	3.68	3.29	3.06
16	4.49	3.63	3.24	3.01
17	4.45	3.59	3.20	2.96
18	4.41	3.55	3.16	2.93
19	4.38	3.52	3.13	2.90
20	4.35	3.49	3.10	2.87
21	4.32	3.47	3.07	2.84
22	4.30	3.44	3.05	2.82
23	4.28	3.42	3.03	2.80
24	4.26	3.40	3.01	2.78
25	4.24	3.39	2.99	2.76
26	4.23	3.37	2.98	2.74
27	4.21	3.35	2.96	2.73
28	4.20	3.34	2.95	2.71
29	4.18	3.33	2.93	2.70
30	4.17	3.32	2.92	2.69
40	4.08	3.23	2.84	2.61
60	4.00	3.15	2.76	2.53
90	3.95	3.10	2.71	2.47
120	3.92	3.07	2.68	2.45
∞	3.84	3.00	2.60	2.37

Wooldridge, 2019, p.835

Exponential distribution

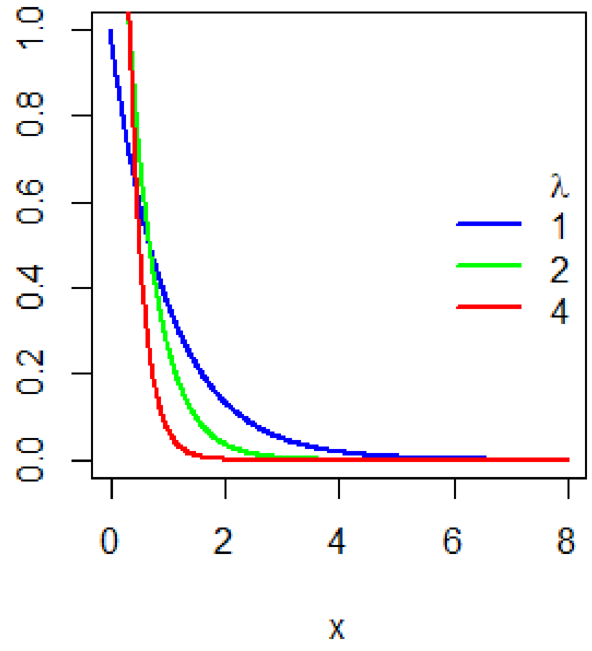
The **exponential distribution** is a special case of the gamma distribution, where $\alpha = 1 \; and \; \beta = \frac{1}{\lambda}$

- **PDF**: $f(x|\lambda) = \lambda e^{-\lambda x}$
- **CDF**: $F(x|\lambda) = 1 - e^{-\lambda x}$
- **Moore & Siegel 2013**: $f(x, \mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$

→appropriate to model the time for a single outcome to occur if events occur independently and at a constant rate

  - e.g., storms, catastrophes, accidents… particle movement

- mean = $\frac{1}{\lambda}$
- $sd = \frac{1}{\lambda}$



Hands on – exponential distribution

Task: If particles arrive independently at a counter at a rate of 1 per second, find the probability that a particle will arrive in 2 seconds?

$$f(x) = \lambda e^{-\lambda x} and\ F(X) = 1 - e^{-\lambda x}$$

Hints:

- find λ or μ
- find x!
- will you need to use the PDF or CDF?

Hands on – exponential distribution

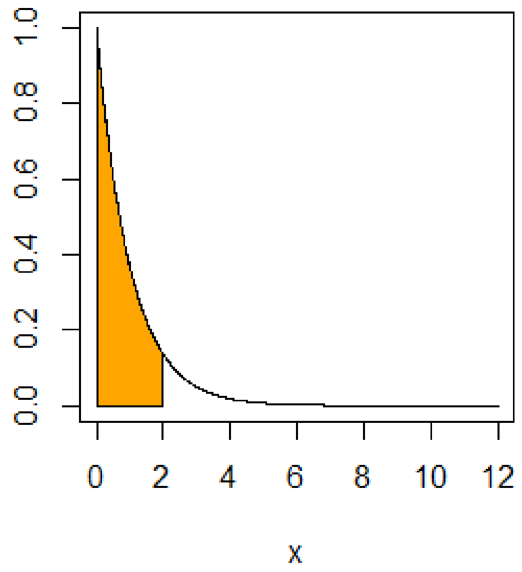
Task: If particles arrive independently at a counter at a rate of 1 per second, find the probability that a particle will arrive in 2 seconds?

$$f(x) = \lambda e^{-\lambda x} and\ F(X) = 1 - e^{-\lambda x}$$

Hints:

- λ or μ= 1
- $x = 2$
- we are interested in the CDF!

- $F(2) = 1 - e^{(-1 \cdot 2)} = 0.8646647 \approx 0.8647 \rightarrow 86.47\%$



Continuous and discrete distributions – A computational comparison

	Discrete distribution	**Continuous distribution**
Expected Value/ mean $\boldsymbol{\mu}$ $\boldsymbol{E(X)}$	$E[X] = \sum_i x_i \Pr(X = x_i)$	$E[X] = \int_{-\infty}^{\infty} x\, f(x) dx$
Variance $\boldsymbol{\sigma^2}$ $\boldsymbol{Var(X)}$	$E[(X - \mu)^2] = \sum_i (x_i - \mu)^2 \Pr(X = x_i)$	$E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$
Example	A die is rolled twice …	X has the following exponential distribution: $f(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0, & x < 0 \end{cases}$ Another example: $f(x) = \begin{cases} 0.5x & 0 \leq x \leq 2 \\ 0 & elsewhere \end{cases}$

Hands on – ‘Continuous’ variance

Task: Let X have the PDF $f(x) = \begin{cases} 0.5x & 0 \leq x \leq 2 \\ 0 & elsewhere \end{cases}$

1) What is E(X)?
2) What is $\mathrm{Var}[X]$?

Hints:

- $E[X] = \int_{-\infty}^{\infty} x\, f(x) dx$
- $E[(X - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$

Hands on – ‘Continuous’ variance

Solution: What is $E[X]$?

$$E[X] = \int_{-\infty}^{\infty} x\, f(x) dx$$

$$E[X] = \int_0^2 xf(x)\, dx$$ ← start by plugging in the bounds

$$E[X] = \int_0^2 x0.5x\, dx$$ ← plug in $f(x)$

$$E[X] = \int_0^2 0.5x^2\, dx$$ ← simplify and integrate

$$E[X] = \frac{x^3}{6}\Big|_0^2 = \frac{2^3}{6} - \frac{0^3}{6} = \frac{8}{6} = \frac{4}{3}$$ ← evaluate antiderivative at bounds

Hands on – 'Continuous' variance

Solution: What is $\mathrm{Var}[X]$?

$$E[(X - \mu)^2] = \int_{-\infty}^{\infty}(x-\mu)^2 f(x)dx$$

$$\mathrm{Var}[X] = \int_0^2 (x-\mu)^2 f(x)\, dx$$ ← plug in bounds

$$\mathrm{Var}[X] = \int_0^2 (x^2 - 2x\mu + \mu^2) 0.5x\, dx$$ ← plug in $f(x)$

$$\mathrm{Var}[X] = \int_0^2 0.5x^3\, dx - \int_0^2 x^2\mu\, dx + \int_0^2 0.5x\mu^2\, dx$$ ← FToC, simplify and solve

$$\mathrm{Var}[X] = \left[\frac{x^4}{8} - \frac{x^3}{3}\mu + \frac{x^2}{4}\mu^2\right]_0^2 = [0.125x^4 - 0.\overline{33}x^3\mu + 0.25x^2\mu^2]_0^2$$

$$\mathrm{Var}[X] = \frac{16}{8} - \frac{8}{3}\mu + \frac{4}{4}\mu^2 = \frac{16}{8} - \frac{8}{3}\cdot\frac{4}{3} + 1\cdot\left(\frac{4}{3}\right)^2 = \frac{16}{8} - \frac{32}{9} + \frac{16}{9}$$

$$\mathrm{Var}[X] = \frac{18}{9} - \frac{32}{9} + \frac{16}{9} = \frac{2}{9}$$

Hands on – 'Continuous' variance

Alternative solution: What is $\mathrm{Var}[X]$? → $E[X^2] - E[X]^2$

$$E[X^2] \quad = \int_0^2 x^2\, f(x)dx$$
$$E[X^2] \quad = \int_0^2 x^2\, 0.5x\, dx$$
$$E[X^2] \quad = \int_0^2 0.5x^3\, dx$$
$$E[X^2] \quad = \frac{x^4}{8}$$
$$E[X^2] \quad = \frac{2^4}{8} - \frac{0^4}{8} = \frac{16}{8} = 2$$

$$E[X]^2 = \left(\frac{4}{3}\right)^2$$
$$E[X^2] - E[X]^2 = \frac{18}{9} - \frac{16}{9} = \frac{2}{9}$$

Time for your questions

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