# Product of two normal random variables

## Checking the distribution functions

### Generating two normal random variables

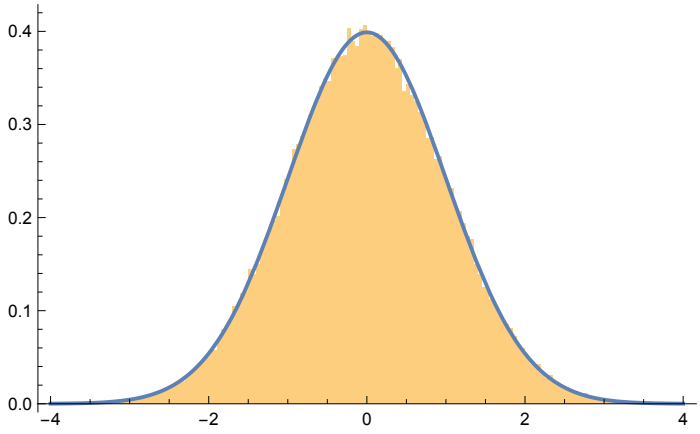
In[1]:=
```
n0 = 100 000; (* numnber of samples *)
```

In[2]:=
```
data1 = RandomVariate[NormalDistribution[0, 1], n0];
```

In[3]:=
```
data2 = RandomVariate[NormalDistribution[0, 1], n0];
```

In[4]:=
```
Show[Histogram[data1, {-4, 4, 0.05}, "ProbabilityDensity"],
 Plot[PDF[NormalDistribution[0, 1], x],
  {x, -4, 4}, PlotStyle → Thick, PlotRange → All]]
```

Out[4]=

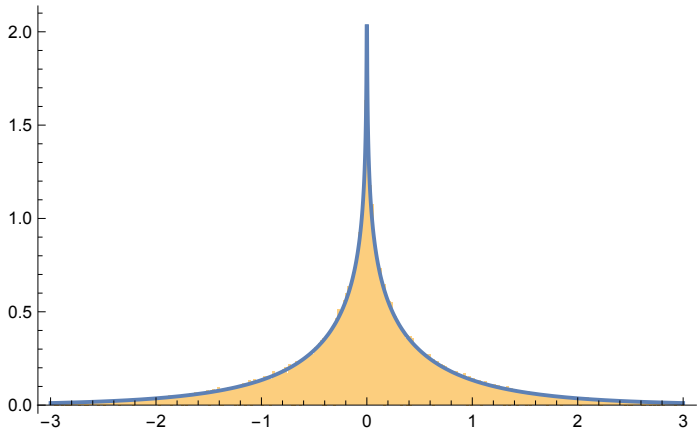



### Generating the product of two normal random variables

In[5]:=
```
data3 = data1 * data2;
```

In[6]:=
```
Show[Histogram[data3, {-3, 3, 0.02}, "ProbabilityDensity"],
 Plot[BesselK[0, Abs[x]] / π, {x, -3, 3}, PlotStyle → Thick, PlotRange → All]]
```

Out[6]=



# P-values for normal and the product of two normals

## Numerical evaluations

In[7]:=
```
precision0 = 500; (* precision for computing p-values *)
```

In[8]:=
```
SetPrecision0[x_] := SetPrecision[x, precision0]
```

In[9]:=
```
precision1 = 20; (* precision for outputs *)
```

In[10]:=
```
N1[x_] := N[x, precision1]
```

## Tail probability of normal random variable

In[11]:=
```
tailnormal[x_] = Integrate[PDF[NormalDistribution[0, 1], t], {t, Abs[x], ∞}]
```

Out[11]=

$$\frac{1}{2}\,\text{Erfc}\left[\frac{\text{Abs}[x]}{\sqrt{2}}\right]$$

In[12]:=
```
tailnormal[SetPrecision0[10.0]]
```

Out[12]=

$$7.619853024160526065973343251599308363504033277956960578035355462896615622059\backslash$$
$$6481703341513851828046716081630382279898298971981706143381826035228935122022\backslash$$
$$4590607303114569218382246575889987628945983087201467435763059338028661837282\backslash$$
$$3290499778661334681686088451113874500815807794766957498793093705124482221553\backslash$$
$$8148064382161426086451954772967628375408702034819605320092626303479247411111\backslash$$
$$7309810626001221936551334469548912147659706213685301067966246882993494705325\backslash$$
$$8629736292723371230 \times 10^{-24}$$

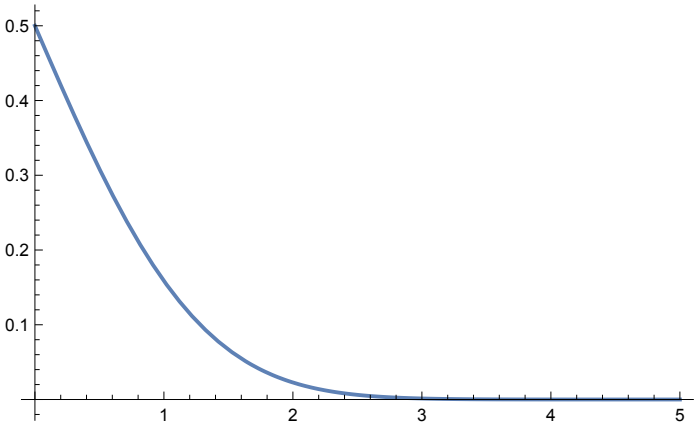
In[13]:=
```
N1[tailnormal[SetPrecision0[10.0]]]
```

Out[13]=

$$7.6198530241605260660 \times 10^{-24}$$

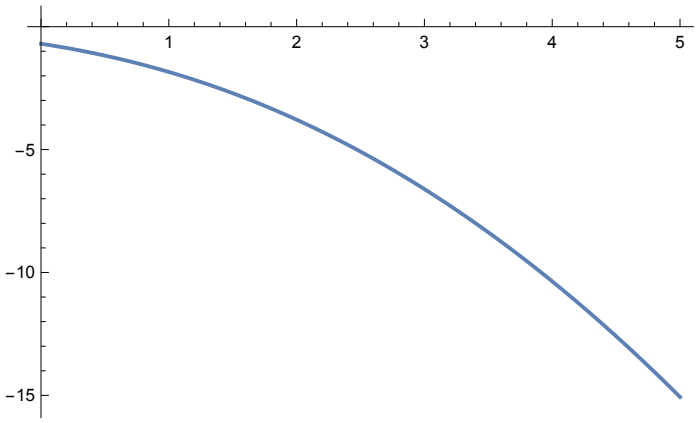
In[14]:=
```
Plot[tailnormal[x], {x, 0, 5}]
```

Out[14]=



In[15]:=
```
Plot[Log[tailnormal[x]], {x, 0, 5}]
```

Out[15]=




## Tail probability of the product of two normal random variables

In[16]:=
```
tailbessel[x_] = 1/2 - Integrate[BesselK[0, t], {t, 0, Abs[x]}] / π
```

Out[16]=

$$\frac{1}{2} - \frac{1}{2}\,\text{Abs}[x]\,(\text{BesselK}[0, \text{Abs}[x]]\,\text{StruveL}[-1, \text{Abs}[x]] + \text{BesselK}[1, \text{Abs}[x]]\,\text{StruveL}[0, \text{Abs}[x]])$$

In[17]:=
```
tailbessel[SetPrecision0[100.0]]
```

Out[17]=

$$1.4749658472608068797590286143824440538875872559798033825359379836525036304443403296523461815455830674816321133347874152820995489362592481282019196096478780991550833725044840010082534000722307531671945015571090165435848464940461485225183891027914962403574899998696283909739584454482173055202715755168226333670442653373747246793253401446572064409392670807673278647254849803556667833771973525217762643039078578308433862250458611542641651149100574656174377060 \times 10^{-45}$$

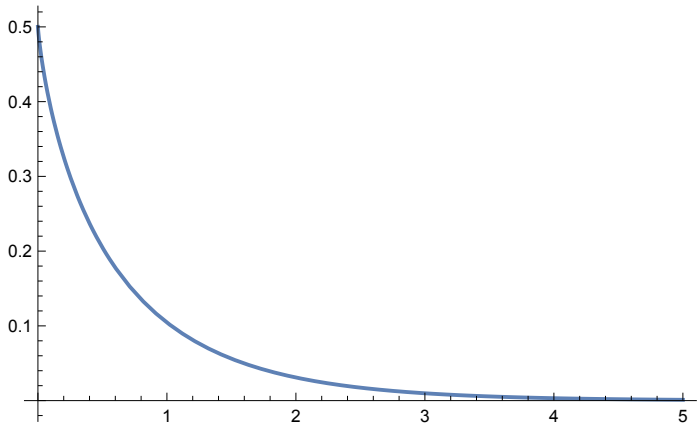
In[18]:=
```
N1[tailbessel[SetPrecision0[100.0]]]
```

Out[18]=

$$1.4749658472608068798 \times 10^{-45}$$

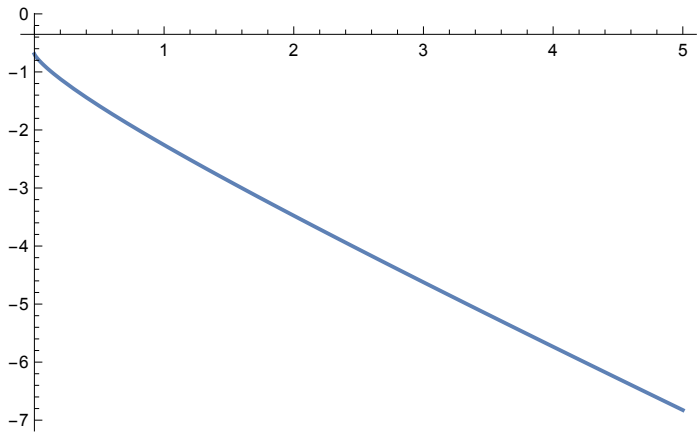
In[19]:=
```
Plot[tailbessel[x], {x, 0, 5}, PlotRange → All]
```

Out[19]=




In[20]:=
```
Plot[Log[tailbessel[x]], {x, 0, 5}, PlotRange → All]
```

Out[20]=




## P-values for embeddings

one-sided p-value for the elements of embeddings with dimensions d

In[21]:=
```
PvalueNormal[x_, d_] := tailnormal[x Sqrt[d]]
```

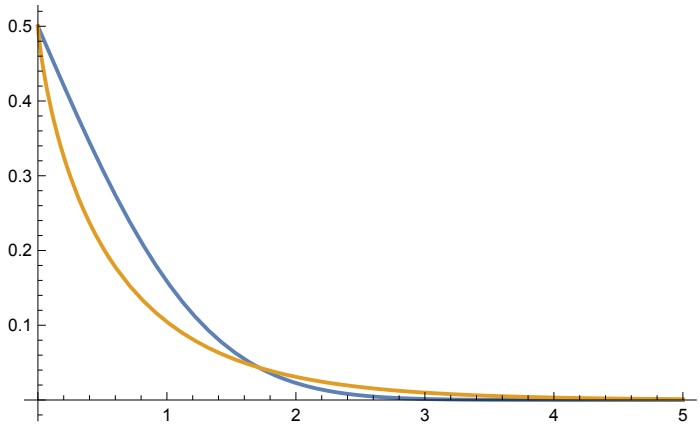
one-sided p-value for the products of two elements of embeddings with dimensions d

In[22]:=
```
PvalueProduct[x_, d_] := tailbessel[x d]
```

## Compare

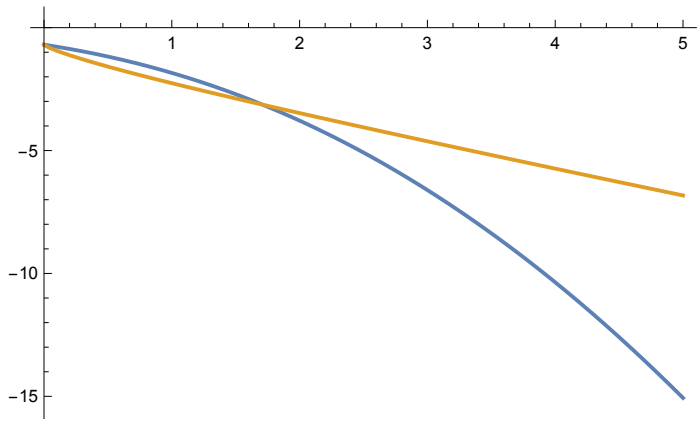
In[23]:=
```
Plot[{tailnormal[x], tailbessel[x]}, {x, 0, 5}, PlotRange → All]
```

Out[23]=



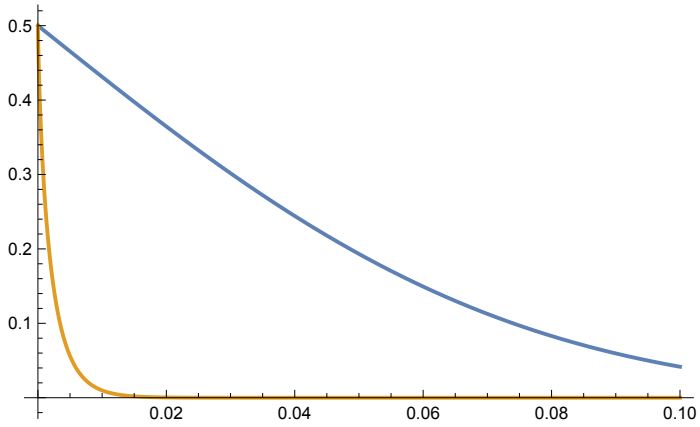

In[24]:=
```
Plot[{Log[tailnormal[x]], Log[tailbessel[x]]}, {x, 0, 5}, PlotRange → All]
```

Out[24]=



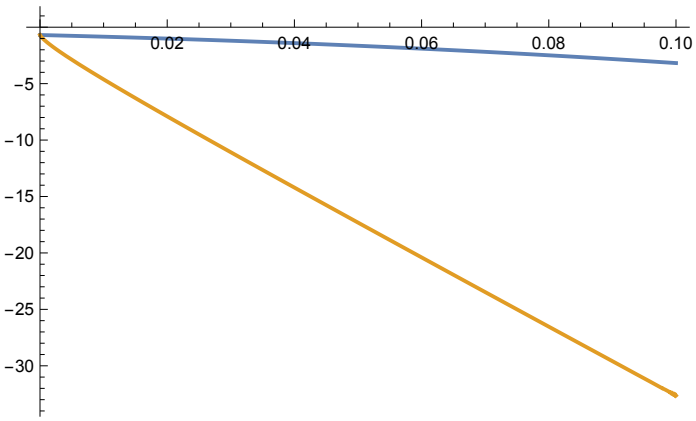

In[25]:=
```
Plot[{PvalueNormal[x, 300], PvalueProduct[x, 300]}, {x, 0, 0.1}, PlotRange → All]
```

Out[25]=



In[26]:=
```
Plot[{Log[PvalueNormal[x, 300]], Log[PvalueProduct[x, 300]]},
 {x, 0, 0.1}, PlotRange → All]
```

Out[26]=




# GloVe ICA

## setting

In[27]:=
```
filename = "data/glove_ica"; (* base name for input csv *)
```

In[28]:=
```
pvnum = 10; (* number of p-values shown in this notebook *)
```

## loading embeddings

In[29]:=
```
embeddings = Import[FileNameJoin[{NotebookDirectory[], filename <> ".csv"}]];
```

In[30]:=
```
Dimensions[embeddings]
```

Out[30]=
```
{2, 301}
```

In[31]:=
```
d0 = Dimensions[embeddings][[2]] - 1 (* dimensions *)
```

Out[31]=
```
300
```

In[32]:=
```
embeddings[[1 ;; 2, 1]]
```

Out[32]=
```
{ultraviolet, light}
```

In[33]:=
```
embedding1 = Table[embeddings[[1]][[i + 1]], {i, 1, d0}];
```

In[34]:=
```
embedding2 = Table[embeddings[[2]][[i + 1]], {i, 1, d0}];
```

In[35]:=
```
product = embedding1 embedding2;
```

## largest elements

In[36]:=
```
Reverse[Sort[Abs[embedding1]]][[1 ;; pvnum]]
```

Out[36]=
```
{0.534778, 0.26525, 0.219918, 0.142215,
 0.140345, 0.135273, 0.123291, 0.118807, 0.112031, 0.10384}
```

In[37]:=
```
Reverse[Sort[Abs[embedding2]]][[1 ;; pvnum]]
```

Out[37]=
```
{0.553637, 0.3788, 0.23018, 0.114424, 0.113606,
 0.110458, 0.108315, 0.105231, 0.103104, 0.102278}
```

In[38]:=
```
Reverse[Sort[Abs[product]]][[1 ;; pvnum]]
```

Out[38]=
```
{0.296073, 0.030351, 0.0224457, 0.0140396, 0.00698963,
 0.00558748, 0.0053346, 0.00516984, 0.00512847, 0.00509834}
```

## computing p-values

In[39]:=
```
pvalue1 = Table[PvalueNormal[SetPrecision0[embedding1[[i]], d0], {i, 1, d0}];
```

In[40]:=
```
pvalue2 = Table[PvalueNormal[SetPrecision0[embedding2[[i]], d0], {i, 1, d0}];
```

In[41]:=
```
pvalue3 = Table[PvalueProduct[SetPrecision0[product[[i]], d0], {i, 1, d0}];
```

## truncate digits

In[42]:=
```
pvalue1 = N1[pvalue1];
```

In[43]:=
```
pvalue2 = N1[pvalue2];
```

In[44]:=
```
pvalue3 = N1[pvalue3];
```

## smallest p - values

In[45]:=
```
Sort[pvalue1][[1 ;; pvnum]]
```

Out[45]=
```
{9.9727102109255499839 × 10^-21, 2.1713253147956397236 × 10^-6,
 0.000069741544648986716679, 0.006884602839660907199 1,
 0.007531645242357326808 4, 0.009564591112190209271 6, 0.016361578793168615862,
 0.019804368158821912148, 0.026163923335786963767, 0.036043532415242638207}
```

In[46]:=
```
Sort[pvalue2][[1 ;; pvnum]]
```

Out[46]=
```
{4.4352624838040532268 × 10^-22, 2.6721118617412933276 × 10^-11,
 0.000033480324678891397434, 0.023746355809461928765,
 0.024550463178974927537, 0.027862357739569101300, 0.030323093636972657604,
 0.034177578312293119273, 0.037064199102258047676, 0.038238320398020305428}
```

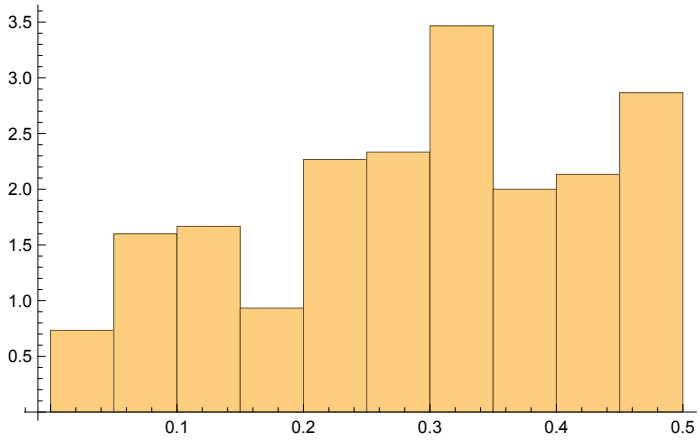
In[47]:=
```
Sort[pvalue3][[1 ;; pvnum]]
```

Out[47]=
$$\{1.1188250686729021116 \times 10^{-40},\ 0.000013818835165210664030,$$
$$0.00016900066437121847150,\ 0.002559616891209596188 8,$$
$$0.027602292082712734552,\ 0.045359478539450394649,\ 0.049683428069175252811,$$
$$0.052734854137595203625,\ 0.053532084761371115616,\ 0.054120792586203003736\}$$

## histogram of p - values

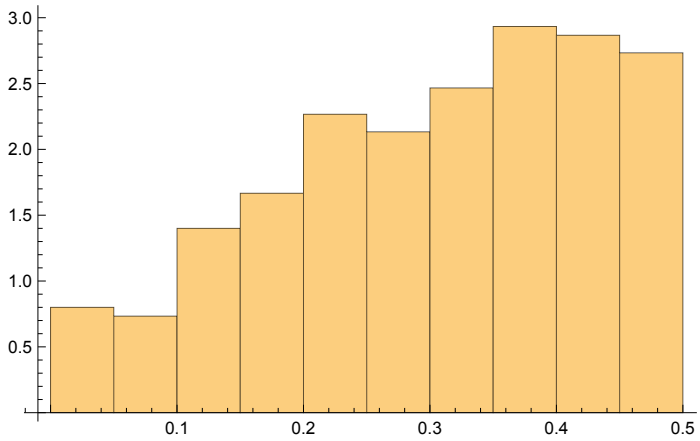
In[48]:=
```
Show[Histogram[pvalue1, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[48]=



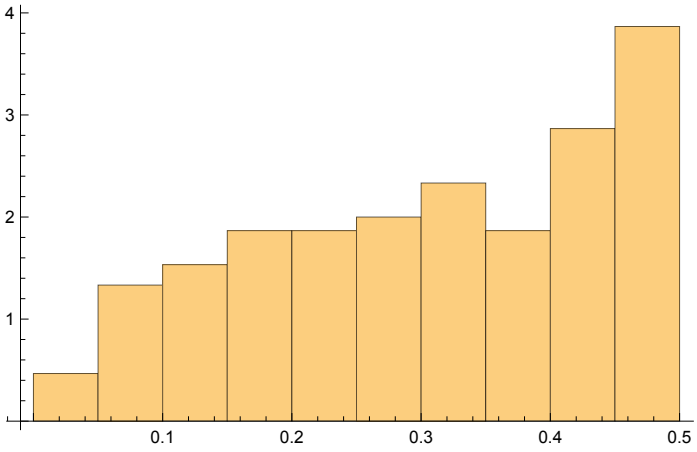

In[49]:=
```
Show[Histogram[pvalue2, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[49]=



In[50]:=
```
Show[Histogram[pvalue3, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[50]=




### output p - values

In[51]:=
```
pvalues = {pvalue1, pvalue2, pvalue3};
```

In[52]:=
```
Export[FileNameJoin[{NotebookDirectory[], filename <> "-pvalue.csv"}],
  pvalues, "CSV"];
```

---

# GloVe PCA

### setting

In[53]:=
```
filename = "data/glove_pca"; (* base name for input csv *)
```

In[54]:=
```
pvnum = 10; (* number of p-values shown in this notebook *)
```

### loading embeddings

In[55]:=
```
embeddings = Import[FileNameJoin[{NotebookDirectory[], filename <> ".csv"}]];
```

In[56]:=
```
Dimensions[embeddings]
```

Out[56]=
```
{2, 301}
```

In[57]:=
```
d0 = Dimensions[embeddings][[2]] - 1 (* dimensions *)
```

Out[57]=
```
300
```

In[58]:=
```
embeddings[[1 ;; 2, 1]]
```

Out[58]=
```
{ultraviolet, light}
```

In[59]:=
```
embedding1 = Table[embeddings[[1]][[i + 1]], {i, 1, d0}];
```

In[60]:=
```
embedding2 = Table[embeddings[[2]][[i + 1]], {i, 1, d0}];
```

In[61]:=
```
product = embedding1 embedding2;
```

## largest elements

In[62]:=
```
Reverse[Sort[Abs[embedding1]]][[1 ;; pvnum]]
```

Out[62]=
```
{0.202355, 0.197082, 0.174948, 0.172482,
 0.157429, 0.152461, 0.150836, 0.150187, 0.145999, 0.137225}
```

In[63]:=
```
Reverse[Sort[Abs[embedding2]]][[1 ;; pvnum]]
```

Out[63]=
```
{0.247417, 0.215653, 0.176051, 0.175901,
 0.174256, 0.158521, 0.145853, 0.139102, 0.132596, 0.132584}
```

In[64]:=
```
Reverse[Sort[Abs[product]]][[1 ;; pvnum]]
```

Out[64]=
```
{0.0238298, 0.0216126, 0.0199125, 0.0193315, 0.0170272,
 0.016883, 0.0165086, 0.0164595, 0.0160765, 0.0149764}
```

## computing p-values

In[65]:=
```
pvalue1 = Table[PvalueNormal[SetPrecision0[embedding1[[i]], d0], {i, 1, d0}];
```

In[66]:=
```
pvalue2 = Table[PvalueNormal[SetPrecision0[embedding2[[i]], d0], {i, 1, d0}];
```

In[67]:=
```
pvalue3 = Table[PvalueProduct[SetPrecision0[product[[i]], d0], {i, 1, d0}];
```

## truncate digits

In[68]:=
```
pvalue1 = N1[pvalue1];
```

In[69]:=
```
pvalue2 = N1[pvalue2];
```

In[70]:=
```
pvalue3 = N1[pvalue3];
```

## smallest p - values

In[71]:=
```
Sort[pvalue1][[1 ;; pvnum]]
```

Out[71]=
```
{0.00022839928633589803818, 0.00032060091970176385135,
 0.00122202982981003653 10, 0.001406472233835289 4975,
 0.003198069931795996 6585, 0.004136785064782957 8809, 0.004493378977993170 5359,
 0.004643273815113837 9601, 0.005723131158785605 4694, 0.008731490431435110 2950}
```

In[72]:=
```
Sort[pvalue2][[1 ;; pvnum]]
```

Out[72]=
```
{9.1211573884210215832 × 10^-6, 0.00009377599542006898 2650,
 0.001146916503340714 8393, 0.001156842784369252 6073,
 0.001271406481799953 5993, 0.003019360154834635 9915, 0.005764536259826029 8737,
 0.007991241858854561 0080, 0.010820121694564395 228, 0.010825748986848787 497}
```

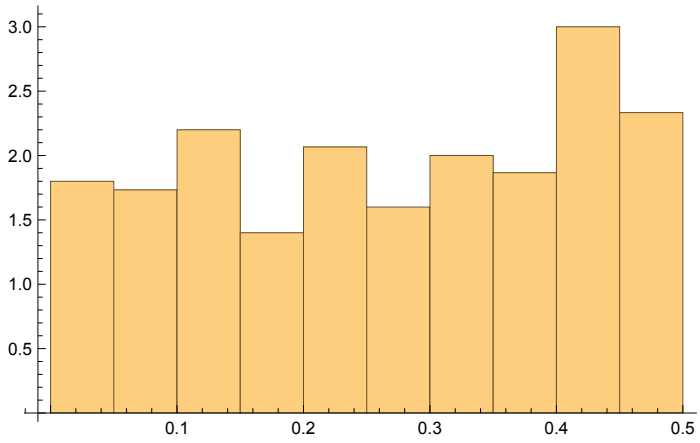
In[73]:=
```
Sort[pvalue3][[1 ;; pvnum]]
```

Out[73]=
```
{0.00010872410728436849796, 0.00022054231403184558623,
 0.00038034992569077376276, 0.00045848895053047518300,
 0.00096525251477809166805, 0.00101148223846126793l4, 0.00114228490408233331301,
 0.00116067221522736585830, 0.00131469809902760057610, 0.00188259815297226256250}
```

## histogram of p - values

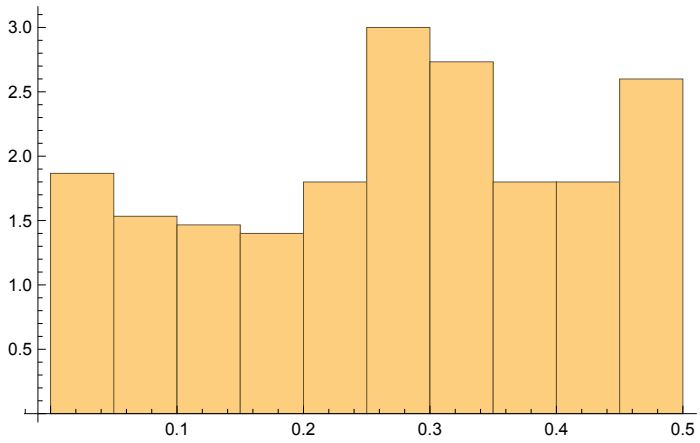
In[74]:=
```
Show[Histogram[pvalue1, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[74]=



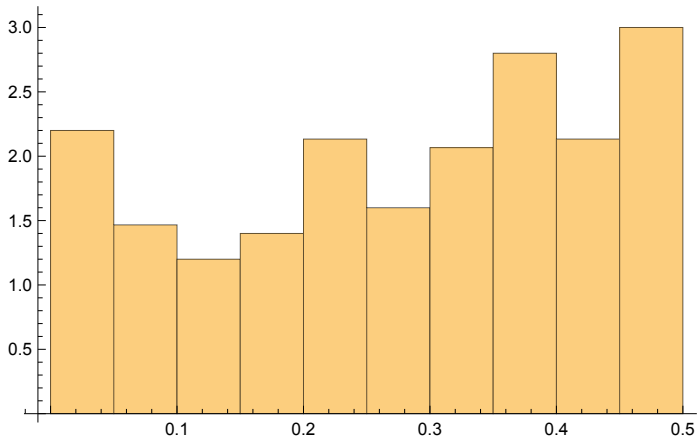

In[75]:=
```
Show[Histogram[pvalue2, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[75]=



In[76]:=
```
Show[Histogram[pvalue3, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[76]=




### output p - values

In[77]:=
```
pvalues = {pvalue1, pvalue2, pvalue3};
```

In[78]:=
```
Export[FileNameJoin[{NotebookDirectory[], filename <> "-pvalue.csv"}],
  pvalues, "CSV"];
```

---

# Pythia ICA

### setting

In[79]:=
```
filename = "data/EleutherAI-pythia-160m_ica"; (* base name for input csv *)
```

In[80]:=
```
pvnum = 10; (* number of p-values shown in this notebook *)
```

### loading embeddings

In[81]:=
```
embeddings = Import[FileNameJoin[{NotebookDirectory[], filename <> ".csv"}]];
```

In[82]:=
```
Dimensions[embeddings]
```

Out[82]=
```
{2, 769}
```

In[83]:=
```
d0 = Dimensions[embeddings][[2]] - 1 (* dimensions *)
```

Out[83]=
```
768
```

In[84]:=
```
embeddings[[1 ;; 2, 1]]
```

Out[84]=
```
{ultraviolet, light}
```

In[85]:=
```
embedding1 = Table[embeddings[[1]][[i + 1]], {i, 1, d0}];
```

In[86]:=
```
embedding2 = Table[embeddings[[2]][[i + 1]], {i, 1, d0}];
```

In[87]:=
```
product = embedding1 embedding2;
```

## largest elements

In[88]:=
```
Reverse[Sort[Abs[embedding1]]][[1 ;; pvnum]]
```

Out[88]=
```
{0.601093, 0.149198, 0.115575, 0.107548, 0.0933143,
 0.0887963, 0.0874292, 0.0872673, 0.0863517, 0.0861818}
```

In[89]:=
```
Reverse[Sort[Abs[embedding2]]][[1 ;; pvnum]]
```

Out[89]=
```
{0.645316, 0.150569, 0.122209, 0.116203, 0.11412,
 0.0897541, 0.087483, 0.0742343, 0.0738345, 0.0729211}
```

In[90]:=
```
Reverse[Sort[Abs[product]]][[1 ;; pvnum]]
```

Out[90]=
```
{0.387895, 0.0224646, 0.0101595, 0.00820432, 0.00636319,
 0.00564241, 0.00505327, 0.00499439, 0.00494187, 0.00470642}
```

## computing p-values

In[91]:=
```
pvalue1 = Table[PvalueNormal[SetPrecision0[embedding1[[i]], d0], {i, 1, d0}];
```

In[92]:=
```
pvalue2 = Table[PvalueNormal[SetPrecision0[embedding2[[i]], d0], {i, 1, d0}];
```

In[93]:=
```
pvalue3 = Table[PvalueProduct[SetPrecision0[product[[i]], d0], {i, 1, d0}];
```

## truncate digits

In[94]:=
```
pvalue1 = N1[pvalue1];
```

In[95]:=
```
pvalue2 = N1[pvalue2];
```

In[96]:=
```
pvalue3 = N1[pvalue3];
```

## smallest p - values

In[97]:=
```
Sort[pvalue1][[1 ;; pvnum]]
```

Out[97]=
```
{1.3239142906559974835 × 10^-62, 0.000017771101896645251444,
 0.0006802489713048409541 6, 0.001439052971490637221 9,
 0.004854802818808639343 9, 0.006931495958530492361 0, 0.007698407020437201907 4,
 0.007794014930490877234 4, 0.008354497380855494412 6, 0.008462345341666391663 6}
```

In[98]:=
```
Sort[pvalue2][[1 ;; pvnum]]
```

Out[98]=
```
{7.9232621737520507911 × 10^-72, 0.000015051031511772283297,
 0.000353605864957957903 10, 0.000640274475048589931 64,
 0.000781852085951022641 53, 0.006435120949924865604 6, 0.007666863610204012603 2,
 0.019831526361069609108, 0.020370185096259333318, 0.021647709013704370874}
```

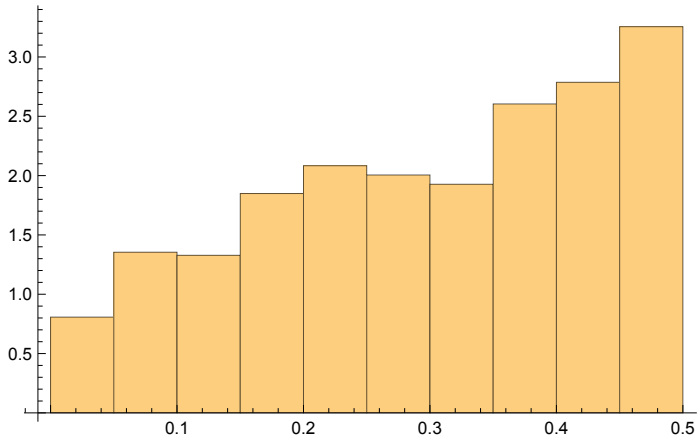
In[99]:=
```
Sort[pvalue3][[1 ;; pvnum]]
```

Out[99]=
```
{9.6611899412734605431 × 10^-132, 2.9852622069772292607 × 10^-9,
 0.00005443876618829030529 4, 0.00026806005323228060740,
 0.00122652985138823333 76, 0.00224089094649027463 86, 0.00368235774616938335 87,
 0.00387063623059305534 59, 0.00404687686157582054 33, 0.00494313176241929372 95}
```

## histogram of p - values

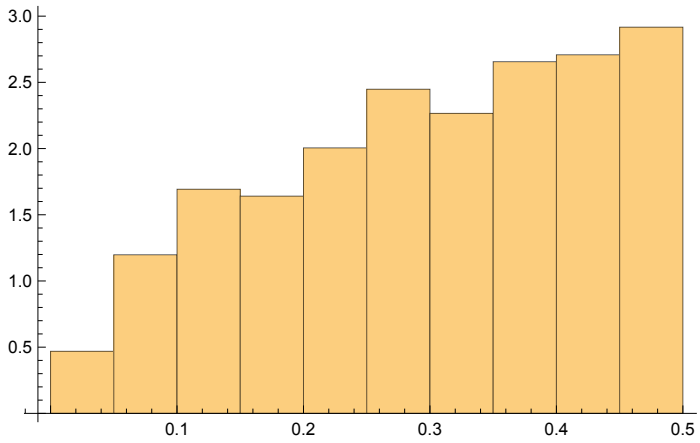
In[100]:=
```
Show[Histogram[pvalue1, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[100]=



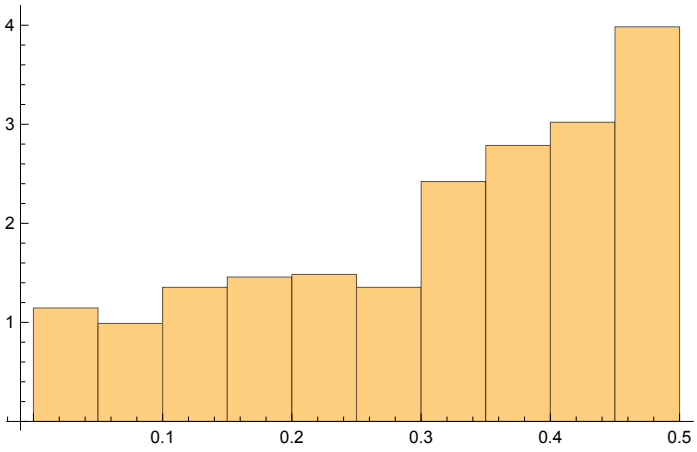

In[101]:=
```
Show[Histogram[pvalue2, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[101]=



In[102]:=
```
Show[Histogram[pvalue3, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[102]=




## output p - values

In[103]:=
```
pvalues = {pvalue1, pvalue2, pvalue3};
```

In[104]:=
```
Export[FileNameJoin[{NotebookDirectory[], filename <> "-pvalue.csv"}],
  pvalues, "CSV"];
```

# Pythia PCA

## setting

In[105]:=
```
filename = "data/EleutherAI-pythia-160m_pca"; (* base name for input csv *)
```

In[106]:=
```
pvnum = 10; (* number of p-values shown in this notebook *)
```

## loading embeddings

In[107]:=
```
embeddings = Import[FileNameJoin[{NotebookDirectory[], filename <> ".csv"}]];
```

In[108]:=
```
Dimensions[embeddings]
```

Out[108]=
```
{2, 769}
```

In[109]:=
```
d0 = Dimensions[embeddings][[2]] - 1 (* dimensions *)
```

Out[109]=
```
768
```

In[110]:=
```
embeddings[[1 ;; 2, 1]]
```
Out[110]=
```
{ultraviolet, light}
```

In[111]:=
```
embedding1 = Table[embeddings[[1]][[i + 1]], {i, 1, d0}];
```

In[112]:=
```
embedding2 = Table[embeddings[[2]][[i + 1]], {i, 1, d0}];
```

In[113]:=
```
product = embedding1 embedding2;
```

## largest elements

In[114]:=
```
Reverse[Sort[Abs[embedding1]]][[1 ;; pvnum]]
```
Out[114]=
```
{0.125502, 0.121365, 0.0973242, 0.0957124, 0.0929038,
 0.0920286, 0.0914271, 0.0909184, 0.0898081, 0.0876936}
```

In[115]:=
```
Reverse[Sort[Abs[embedding2]]][[1 ;; pvnum]]
```
Out[115]=
```
{0.128636, 0.113207, 0.10435, 0.103102, 0.101709,
 0.101052, 0.0995926, 0.0989615, 0.0986276, 0.0978361}
```

In[116]:=
```
Reverse[Sort[Abs[product]]][[1 ;; pvnum]]
```
Out[116]=
```
{0.0156119, 0.0129396, 0.00768797, 0.00764188, 0.00667392,
 0.00666742, 0.00645442, 0.006219, 0.0061793, 0.00616578}
```

## computing p-values

In[117]:=
```
pvalue1 = Table[PvalueNormal[SetPrecision0[embedding1[[i]], d0], {i, 1, d0}];
```

In[118]:=
```
pvalue2 = Table[PvalueNormal[SetPrecision0[embedding2[[i]], d0], {i, 1, d0}];
```

In[119]:=
```
pvalue3 = Table[PvalueProduct[SetPrecision0[product[[i]], d0], {i, 1, d0}];
```

## truncate digits

In[120]:=
```
pvalue1 = N1[pvalue1];
```

In[121]:=
```
pvalue2 = N1[pvalue2];
```

In[122]:=
```
pvalue3 = N1[pvalue3];
```

## smallest p - values

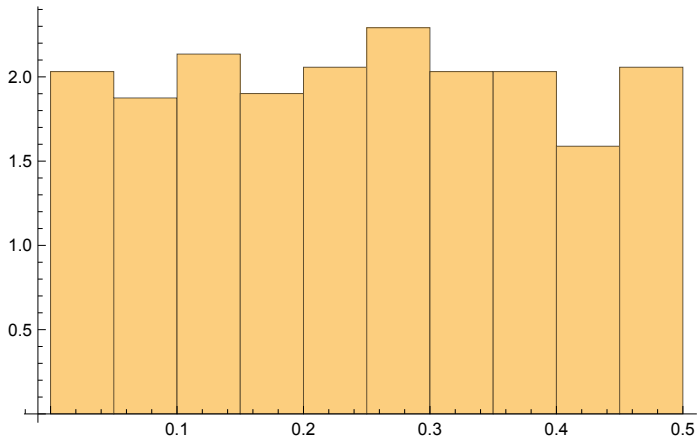
In[123]:=
```
Sort[pvalue1][[1 ;; pvnum]]
```
Out[123]=
```
{0.0002525603882206207 3907, 0.0003849738178977 5003943,
 0.0034970315579463093691, 0.0039953940157082905983,
 0.0050174489478795269111, 0.0053803973158651800971, 0.0056432276757008489033,
 0.0058743490944086947131, 0.0064081171035014159542, 0.0075445040780099606809}
```

In[124]:=
```
Sort[pvalue2][[1 ;; pvnum]]
```
Out[124]=
```
{0.0001820278133446619 1715, 0.0008525683421259011 4958,
 0.0019149984644905975316, 0.0021366729418871678656,
 0.0024113984950979447977, 0.0025516287425164403500, 0.0028901409804209738951,
 0.0030486483008225889579, 0.0031356652741908659164, 0.0033508684914258616032}
```

In[125]:=
```
Sort[pvalue3][[1 ;; pvnum]]
```
Out[125]=
$$\{6.8193423690938106586 \times 10^{-7},\ 5.7812168629015303849 \times 10^{-6},$$
```
 0.0004096999241353486 2805, 0.0004255456667072833 24162,
 0.0009472558036194627 4880, 0.0009523852759709575 0162,
 0.0011368312340156486596, 0.0013831515735792064068,
 0.0014297337542016547266, 0.0014459480832412269150}
```

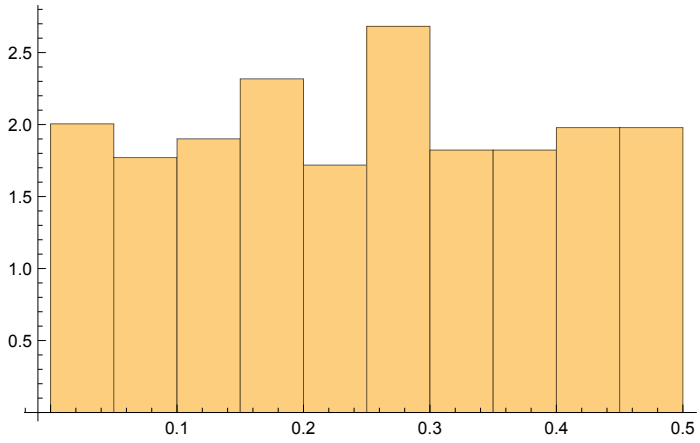
## histogram of p - values

In[126]:=
```
Show[Histogram[pvalue1, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```
Out[126]=



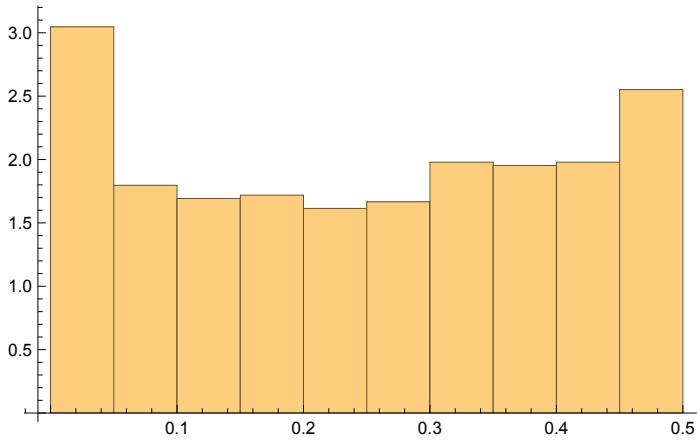
In[127]:=
```
Show[Histogram[pvalue2, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[127]=




In[128]:=
```
Show[Histogram[pvalue3, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[128]=




### output p - values

In[129]:=
```
pvalues = {pvalue1, pvalue2, pvalue3};
```

In[130]:=
```
Export[FileNameJoin[{NotebookDirectory[], filename <> "-pvalue.csv"}],
  pvalues, "CSV"];
```

# GPT2 ICA

### setting

In[131]:=
```
filename = "data/gpt2_ica"; (* base name for input csv *)
```

In[132]:=
```
pvnum = 10; (* number of p-values shown in this notebook *)
```

## loading embeddings

In[133]:=
```
embeddings = Import[FileNameJoin[{NotebookDirectory[], filename <> ".csv"}]];
```

In[134]:=
```
Dimensions[embeddings]
```

Out[134]=
```
{2, 769}
```

In[135]:=
```
d0 = Dimensions[embeddings][[2]] - 1 (* dimensions *)
```

Out[135]=
```
768
```

In[136]:=
```
embeddings[[1 ;; 2, 1]]
```

Out[136]=
```
{ultraviolet, light}
```

In[137]:=
```
embedding1 = Table[embeddings[[1]][[i + 1]], {i, 1, d0}];
```

In[138]:=
```
embedding2 = Table[embeddings[[2]][[i + 1]], {i, 1, d0}];
```

In[139]:=
```
product = embedding1 embedding2;
```

## largest elements

In[140]:=
```
Reverse[Sort[Abs[embedding1]]][[1 ;; pvnum]]
```

Out[140]=
```
{0.412892, 0.274985, 0.156069, 0.123778, 0.117139,
 0.11502, 0.102941, 0.101257, 0.0992244, 0.0991805}
```

In[141]:=
```
Reverse[Sort[Abs[embedding2]]][[1 ;; pvnum]]
```

Out[141]=
```
{0.449752, 0.321907, 0.144877, 0.128259, 0.122076,
 0.121777, 0.110015, 0.103546, 0.100394, 0.0971766}
```

In[142]:=
```
Reverse[Sort[Abs[product]]][[1 ;; pvnum]]
```

Out[142]=
```
{0.185699, 0.0885194, 0.0190056, 0.0179326, 0.0127208,
 0.0102743, 0.0099823, 0.00776384, 0.00774396, 0.00518704}
```

## computing p-values

In[143]:=
```
pvalue1 = Table[PvalueNormal[SetPrecision0[embedding1[[i]], d0], {i, 1, d0}];
```

In[144]:=
```
pvalue2 = Table[PvalueNormal[SetPrecision0[embedding2[[i]], d0], d0], {i, 1, d0}];
```

In[145]:=
```
pvalue3 = Table[PvalueProduct[SetPrecision0[product[[i]], d0], d0], {i, 1, d0}];
```

## truncate digits

In[146]:=
```
pvalue1 = N1[pvalue1];
```

In[147]:=
```
pvalue2 = N1[pvalue2];
```

In[148]:=
```
pvalue3 = N1[pvalue3];
```

## smallest p - values

In[149]:=
```
Sort[pvalue1][[1 ;; pvnum]]
```

Out[149]=
$$\{1.2834480284074826699\times10^{-30},\ 1.2625316958177614424\times10^{-14},\ 7.6225439022690687637\times10^{-6},\ 0.00030151394852150871834,\ 0.00058468433670341731773,\ 0.00071744930048294055338,\ 0.002166961257037646119 1,\ 0.00250704471636819840 33,\ 0.00298170289896250855 84,\ 0.00299277596192871115 53\}$$

In[150]:=
```
Sort[pvalue2][[1 ;; pvnum]]
```

Out[150]=
$$\{5.8756053956656604701\times10^{-36},\ 2.3117635174655319148\times10^{-19},\ 0.00002973061748348791912 2,\ 0.00018941500365791267579,\ 0.00035839769444402885142,\ 0.00036937558852008958569,\ 0.00114867394577673376 09,\ 0.00205522201528577738 40,\ 0.00269960423332171148 31,\ 0.00354022888112170618 08\}$$

In[151]:=
```
Sort[pvalue3][[1 ;; pvnum]]
```

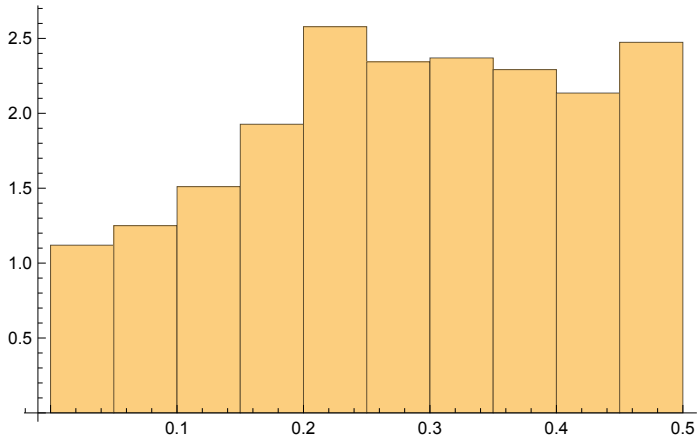
Out[151]=
$$\{3.8397613293868146788\times10^{-64},\ 1.4328633660188202127\times10^{-31},\ 4.5974783925943615532\times10^{-8},\ 1.0766629659811503079\times10^{-7},\ 6.8915619827463771374\times10^{-6},\ 0.00004959903817369698 1040,\ 0.00006285774740485998 0855,\ 0.00038490385259725407147,\ 0.00039125151305702764793,\ 0.00328843948192158799 27\}$$

## histogram of p - values

In[152]:=

```
Show[Histogram[pvalue1, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

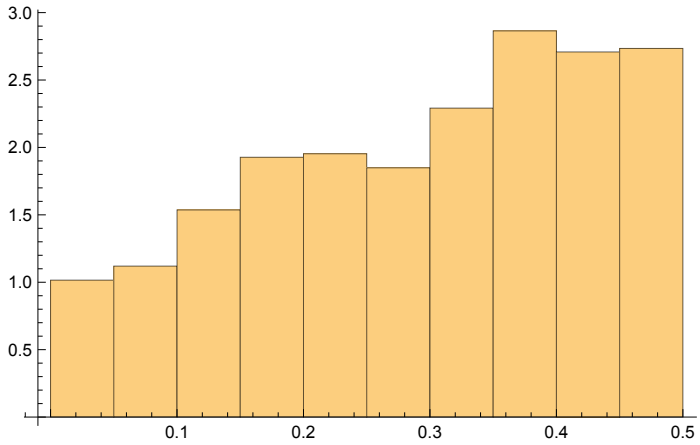
Out[152]=




In[153]:=

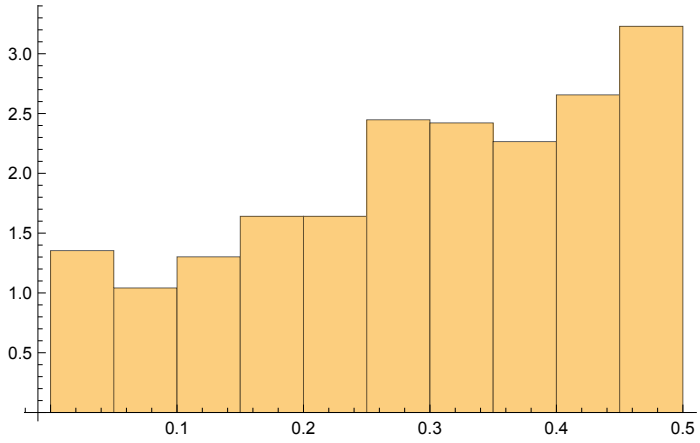
```
Show[Histogram[pvalue2, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[153]=



In[154]:=
```
Show[Histogram[pvalue3, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[154]=




### output p - values

In[155]:=
```
pvalues = {pvalue1, pvalue2, pvalue3};
```

In[156]:=
```
Export[FileNameJoin[{NotebookDirectory[], filename <> "-pvalue.csv"}],
  pvalues, "CSV"];
```

# GPT2 PCA

### setting

In[157]:=
```
filename = "data/gpt2_pca"; (* base name for input csv *)
```

In[158]:=
```
pvnum = 10; (* number of p-values shown in this notebook *)
```

### loading embeddings

In[159]:=
```
embeddings = Import[FileNameJoin[{NotebookDirectory[], filename <> ".csv"}]];
```

In[160]:=
```
Dimensions[embeddings]
```

Out[160]=
```
{2, 769}
```

In[161]:=
```
d0 = Dimensions[embeddings][[2]] - 1 (* dimensions *)
```

Out[161]=
```
768
```

In[162]:=
```
embeddings[[1 ;; 2, 1]]
```
Out[162]=
```
{ultraviolet, light}
```

In[163]:=
```
embedding1 = Table[embeddings[[1]][[i + 1]], {i, 1, d0}];
```

In[164]:=
```
embedding2 = Table[embeddings[[2]][[i + 1]], {i, 1, d0}];
```

In[165]:=
```
product = embedding1 embedding2;
```

## largest elements

In[166]:=
```
Reverse[Sort[Abs[embedding1]]][[1 ;; pvnum]]
```
Out[166]=
```
{0.108743, 0.101386, 0.0995225, 0.095572, 0.0955223,
 0.0949627, 0.0922943, 0.0883422, 0.086621, 0.083538}
```

In[167]:=
```
Reverse[Sort[Abs[embedding2]]][[1 ;; pvnum]]
```
Out[167]=
```
{0.127528, 0.114267, 0.101612, 0.101204, 0.100394,
 0.0979855, 0.0977297, 0.0936943, 0.090686, 0.0897086}
```

In[168]:=
```
Reverse[Sort[Abs[product]]][[1 ;; pvnum]]
```
Out[168]=
```
{0.0095899, 0.00946041, 0.00935614, 0.00921621, 0.00715787,
 0.00666351, 0.00657745, 0.00601056, 0.00589385, 0.00566825}
```

## computing p-values

In[169]:=
```
pvalue1 = Table[PvalueNormal[SetPrecision0[embedding1[[i]], d0], {i, 1, d0}];
```

In[170]:=
```
pvalue2 = Table[PvalueNormal[SetPrecision0[embedding2[[i]], d0], {i, 1, d0}];
```

In[171]:=
```
pvalue3 = Table[PvalueProduct[SetPrecision0[product[[i]], d0], {i, 1, d0}];
```

## truncate digits

In[172]:=
```
pvalue1 = N1[pvalue1];
```

In[173]:=
```
pvalue2 = N1[pvalue2];
```

In[174]:=
```
pvalue3 = N1[pvalue3];
```

## smallest p - values

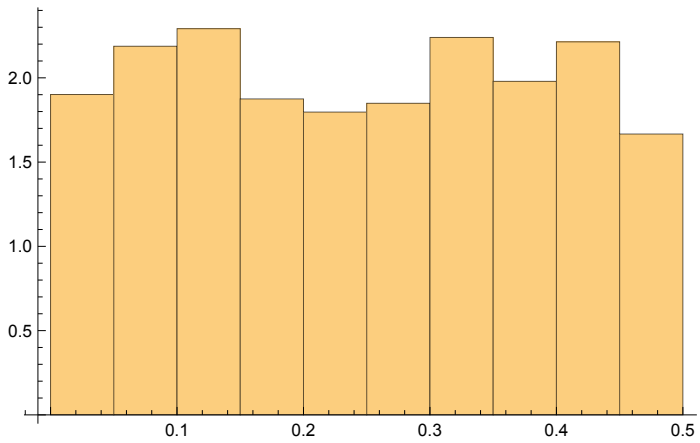
In[175]:=
```
Sort[pvalue1][[1 ;; pvnum]]
```
Out[175]=
```
{0.0012909648877135171132, 0.0024794282531022847861,
 0.0029073704813826130227, 0.0040416670610177599312,
 0.0040581838063236048641, 0.0042481446866775299171, 0.0052678388480166798726,
 0.0071783988190064237491, 0.0081860786565549845988, 0.010304490946062993381}
```

In[176]:=
```
Sort[pvalue2][[1 ;; pvnum]]
```
Out[176]=
```
{0.00020453589951324273447, 0.00077103772939661159025,
 0.0024315389463037215391, 0.0025185733333578072587,
 0.0026994821066789520628, 0.0033092381255084655907, 0.0033807768748035388519,
 0.0047085044653103563300, 0.0059826855260951623546, 0.006457984941733437250 0}
```

In[177]:=
```
Sort[pvalue3][[1 ;; pvnum]]
```
Out[177]=
```
{0.00008646531688710925 6062, 0.00009607353557411403 0709,
 0.00010458580245612786567, 0.00011721540928606733259,
 0.00063439377173455608527, 0.00095548025250574422175,
 0.00102628590850489665 71, 0.00164611831966127576 07,
 0.00181494793715835779 33, 0.00219279564900813599 83}
```

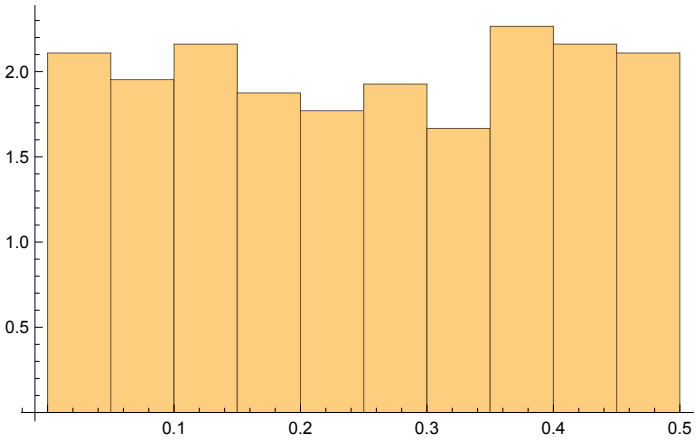
## histogram of p - values

In[178]:=
```
Show[Histogram[pvalue1, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```
Out[178]=



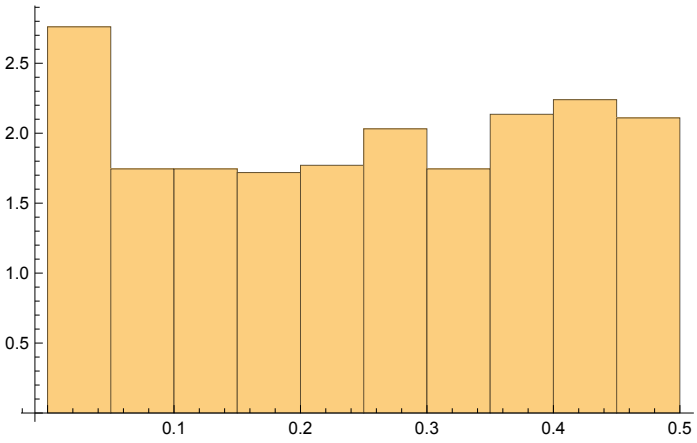
In[179]:=
```
Show[Histogram[pvalue2, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[179]=




In[180]:=
```
Show[Histogram[pvalue3, 10, "ProbabilityDensity"],
 PlotStyle → Thick, PlotRange → All]
```

Out[180]=




## output p - values

In[181]:=
```
pvalues = {pvalue1, pvalue2, pvalue3};
```

In[182]:=
```
Export[FileNameJoin[{NotebookDirectory[], filename <> "-pvalue.csv"}],
  pvalues, "CSV"];
```