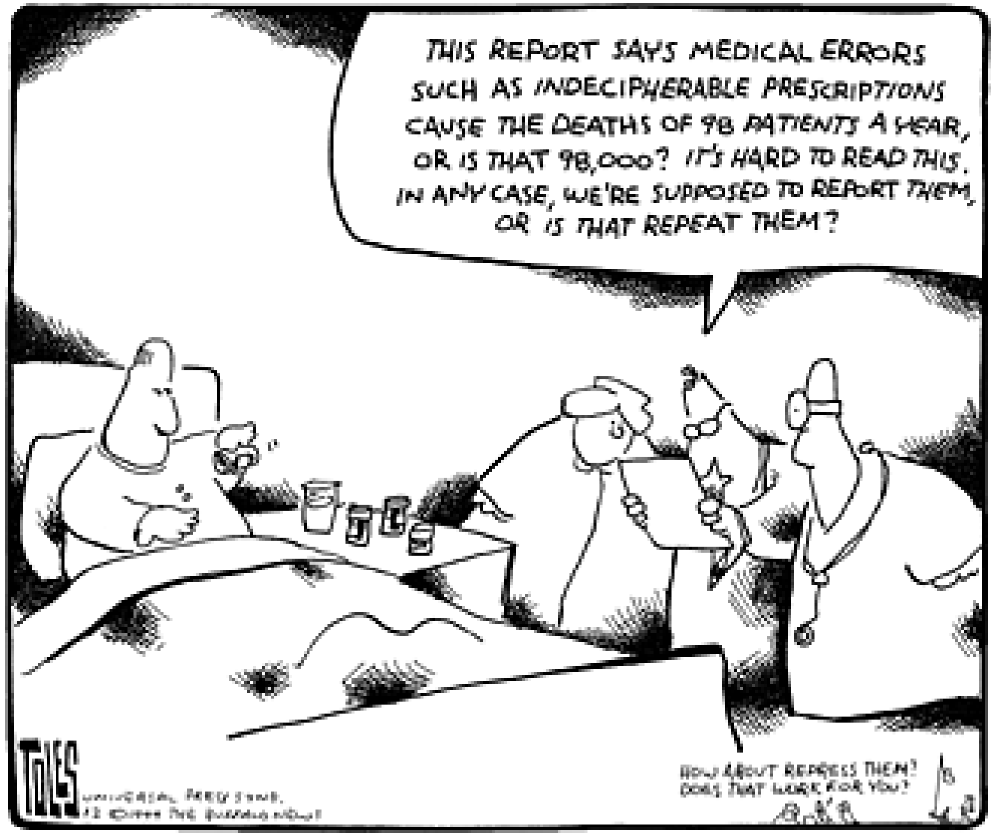
Berkeley
UNIVERSITY OF CALIFORNIA

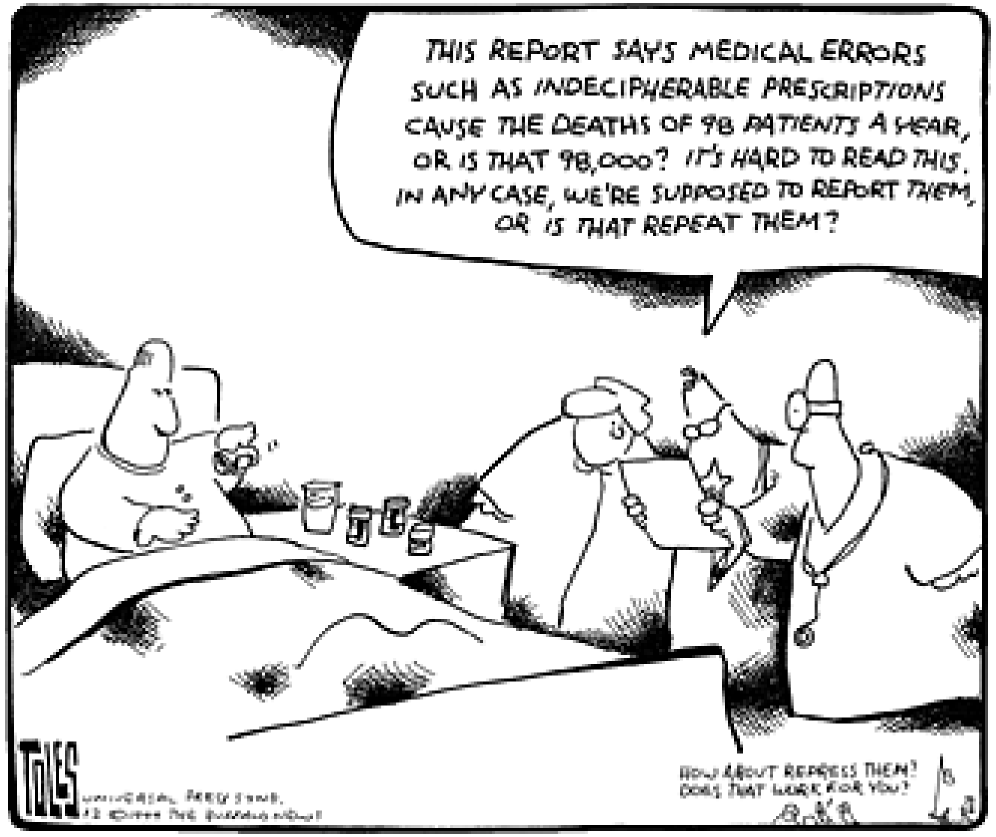
M. Hohle:

Physics 77: Introduction to Computational Techniques in Physics



Introduction to Computational Techniques in Physics: syllabus

Week	**Date**	**Topic**
1	June 12th	Programming Environment & UIs for Python, Programming Fundamentals
2	*June 19th*	Basic Types in Python
3	June 26th	Parsing, Data Processing and File I/O, Visualization
4	July 3rd	Functions, Map & Lambda
5	**July 10th**	**Random Numbers & Probability Distributions, Interpreting Measurements**
6	July 17th	Numerical Integration and Differentiation
7	July 24th	Root finding, Interpolation
8	July 31st	Systems of Linear Equations, Ordinary Differential Equations (ODEs)
9	Aug 7th	Stability of ODEs, Examples
10	Aug 14th	Final Project Presentations



Outline:

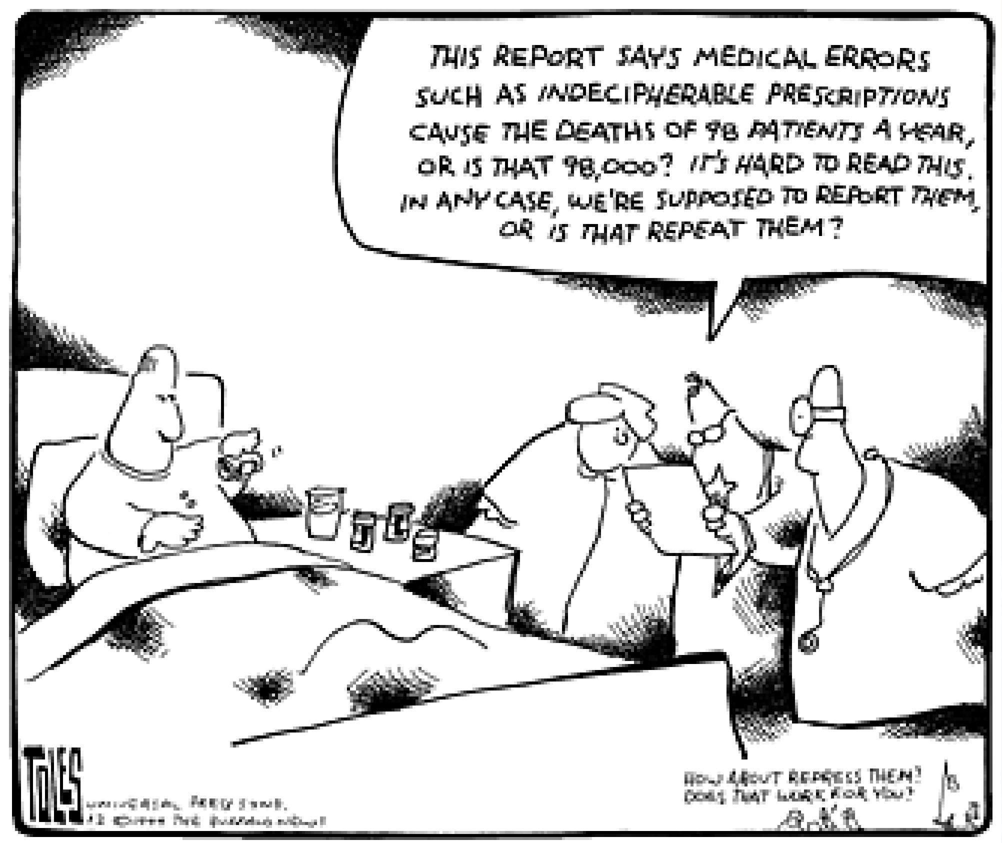
Basics

Most Common PDFs

- uniform
- binomial
- Poissonian
- Normal/Gaussian

Error Estimation

Bayesian Statistics



Outline:

Basics

Most Common PDFs

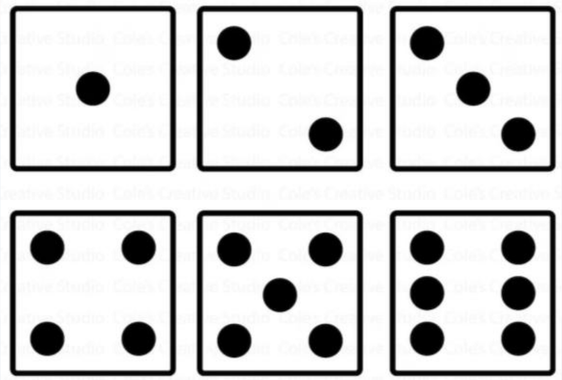
- uniform
- binomial
- Poissonian
- Normal/Gaussian

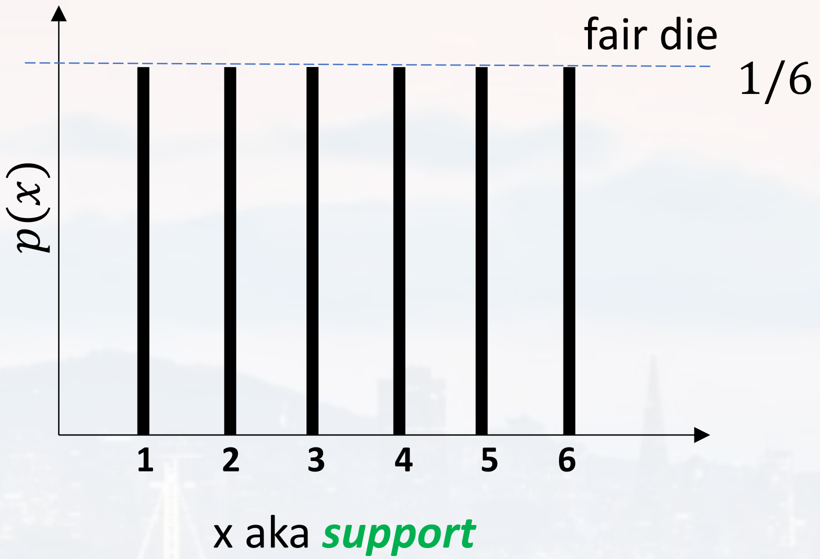
Error Estimation

Bayesian Statistics

distributions

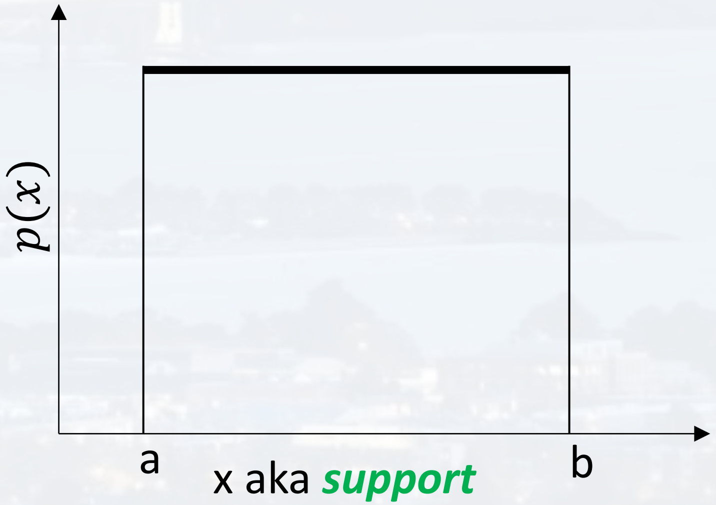
discrete (= countable)





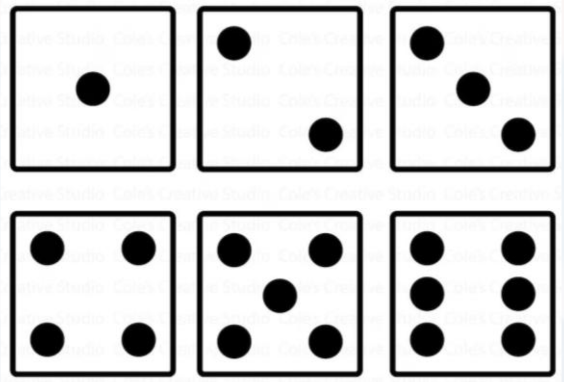
continuous

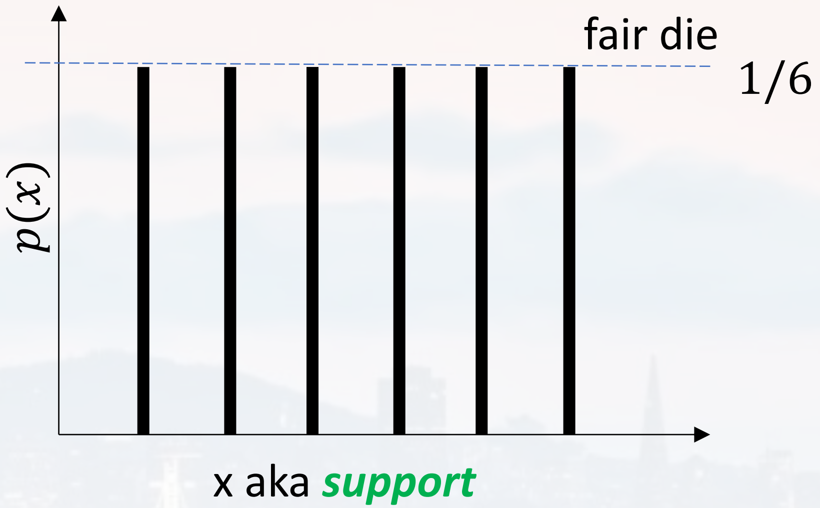
[a ≤ x ≤ b]



distributions

discrete (= countable)





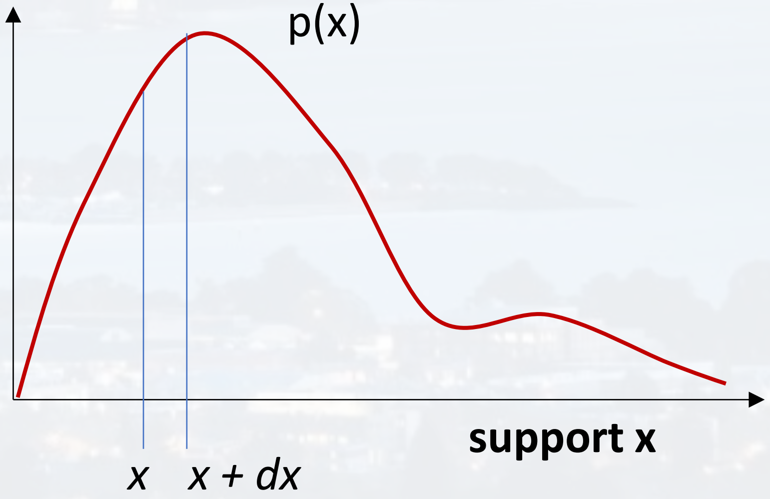
continuous

[a ≤ x ≤ b]

p(x) doesn't make sense

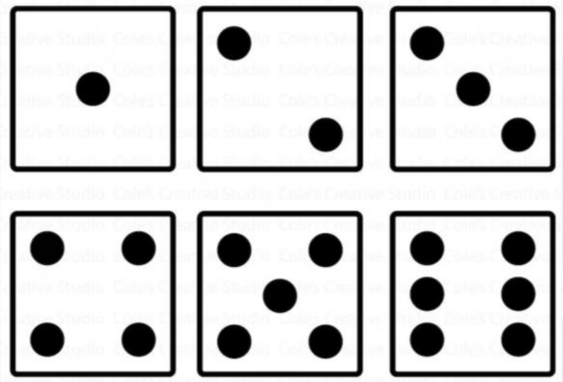
→ *p(x) dx*

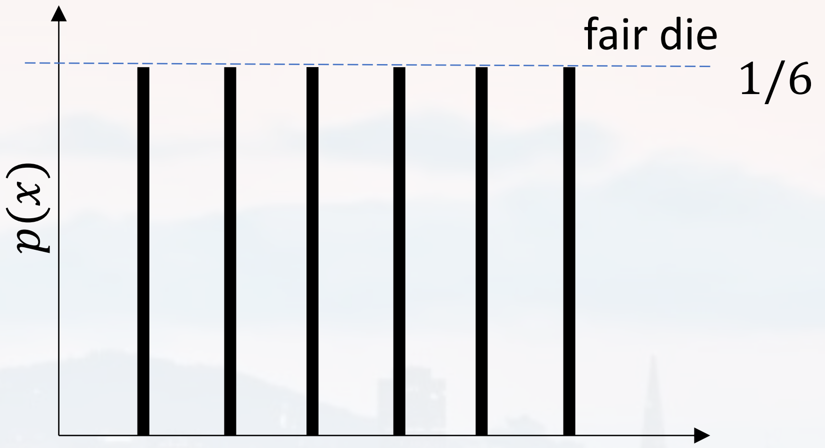
probability **d**ensity **f**unction



distributions

discrete (= countable)



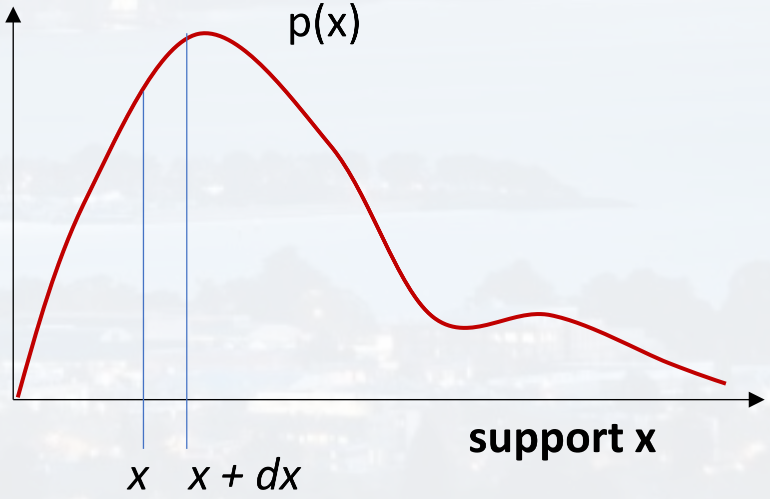


continuous

p(x) doesn't make sense
→ *p(x) dx*

probability **d**ensity **f**unction

dx defines the probability!



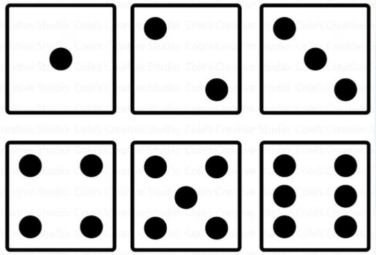
*What is the probability to find a person who is **EXACTLY** 6ft tall?*
*Depends on how **accurate** (dx) you measure!*

	the mean μ	**the variance σ^2**
	(barycenter)	(natural scatter)
discrete (= countable)	$\mu = E(x) = \sum_i x_i \, p(x_i)$	$\sigma^2 = var(x) = \sum_i (x_i - \mu)^2 \, p(x_i)$
continuous	$\mu = E(x) = \int x \, p(x) \, dx$	$\sigma^2 = var(x) = \int (x - \mu)^2 \, p(x) \, dx$

uniform distribution: discrete vs continuous

the mean μ

discrete (= countable)



$$\mu = \sum_i x_i \, p(x_i) = \sum_{i=1}^{6} i \, \frac{1}{6} = \mathbf{3.5}$$

continuous [a ≤ x ≤ b]

$$\mu = \int x \, p(x) \, dx$$

$p(x) = const$ (uniform)

$$= const \int_a^b x \, dx = const \, \frac{1}{2} (b^2 - a^2)$$

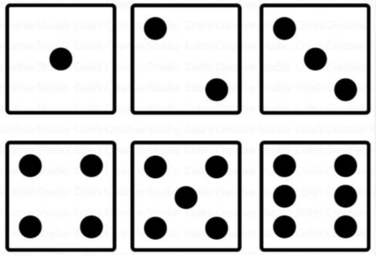
2nd axiom $\int_a^b p(x) \, dx = 1$

$$const \int_a^b dx = 1 \quad \rightarrow const = \frac{1}{b-a}$$

uniform distribution: discrete vs continuous

the mean μ

discrete (= countable)



$$\mu = \sum_i x_i \, p(x_i) = \sum_{i=1}^{6} i \, \frac{1}{6} = \mathbf{3.5}$$

continuous [a ≤ x ≤ b]

$$\mu = \int x \, p(x) \, dx = const \, \frac{1}{2}(b^2 - a^2)$$

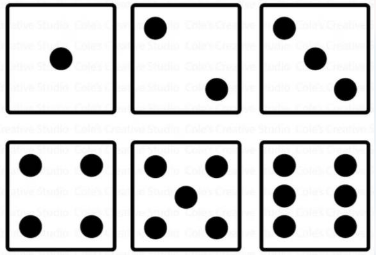
$$const = \frac{1}{b-a}$$

$$\boxed{\mu = \frac{1}{2}\frac{b^2 - a^2}{b-a}}$$

= **3.5** for a = 1 and b = 6

the variance σ^2

discrete (= countable)



$$\sigma^2 = \sum_i (x_i - \mu)^2 \, p(x_i) = \frac{1}{6}\sum_{i=1}^{6}(i - 3.5)^2 \approx \mathbf{2.9}$$

continuous

[a ≤ x ≤ b]

$$\sigma^2 = \int (x - \mu)^2 \, p(x) \, dx = \frac{1}{b-a}\int_a^b (x - 3.5)^2 dx$$

$$\boxed{\sigma^2 = \frac{1}{12}\,(b-a)^2}$$

= 25/12 ≈ **2.1** for a = 1 and b = 6

$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

$$var(x) = \int (x-\mu)^2\, p(x)\, dx = E([x-\mu]^2)$$

variance can be interpreted as **mean of $[x-\mu]^2$**

$$= E(x^2 - 2x\mu + \mu^2)$$

$$= \int [x^2 - 2x\mu + \mu^2]\, p(x)\, dx$$

$$= \int x^2\, p(x)\, dx - 2\mu \int x\, p(x)\, dx + \mu^2 \int p(x)\, dx$$

$$= E(x^2) - 2\mu\, E(x) + \mu^2\, E(1)$$

2nd axiom $\int p(x)\, dx = 1$

$$= E(x^2) - 2\mu\, E(x) + \mu^2$$

$\mu = E(x)$

$$\boxed{\sigma^2 = E(x^2) - E(x)^2}$$

$$\boldsymbol{\sigma^2 = E(x^2) - E(x)^2}$$

$$\mu = E(x) = \int x\, p(x)\, dx$$

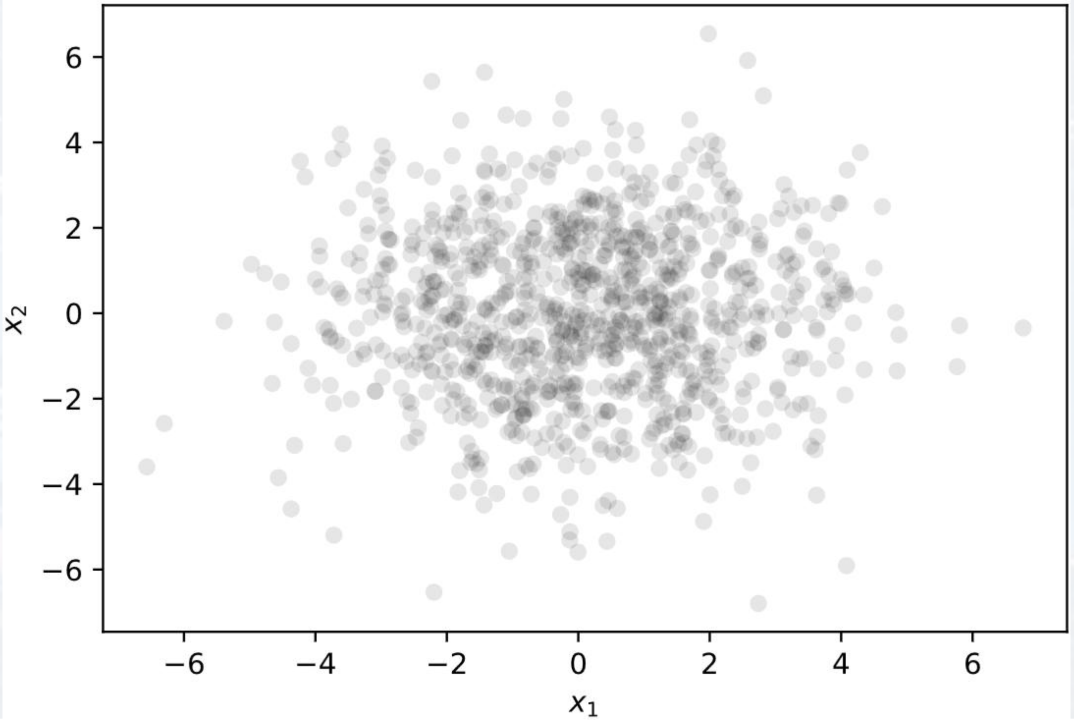
$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

plotting two sets of random number: x_1 and x_2

```
x1 = np.random.normal(0,2,(1000,))
x2 = np.random.normal(0,2,(1000,))

plt.scatter(x1, x2, color = 'k', alpha = 0.1, edgecolor = 'none')
plt.xlabel('$x_1$')
plt.ylabel('$x_2$')
```

x_1 and x_2 are unrelated and mutually **independent**

→ featureless data cloud



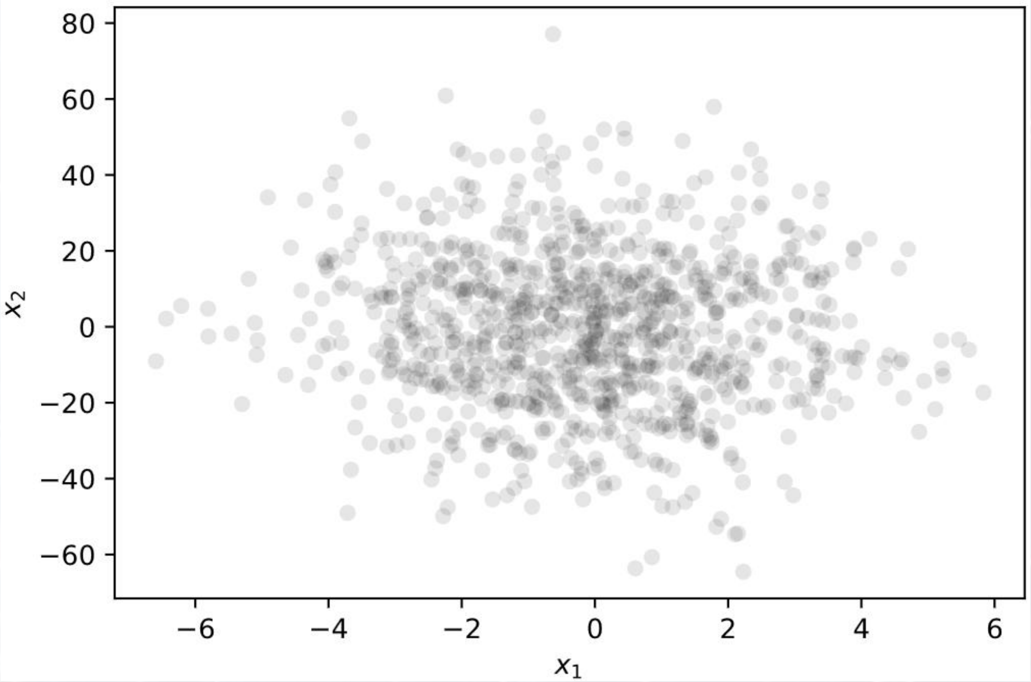
$$\boxed{\sigma^2 = E(x^2) - E(x)^2}$$

$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

plotting two sets of random number: x_1 and x_2

```
x1 = np.random.normal(0,2,(1000,))
x2 = np.random.normal(0,20,(1000,))

plt.scatter(x1, x2, color = 'k', alpha = 0.1, edgecolor = 'none')
plt.xlabel('$x_1$')
plt.ylabel('$x_2$')
```



x_1 and x_2 are unrelated and mutually **independent**
→ featureless data cloud

$$\boxed{\sigma^2 = E(x^2) - E(x)^2}$$

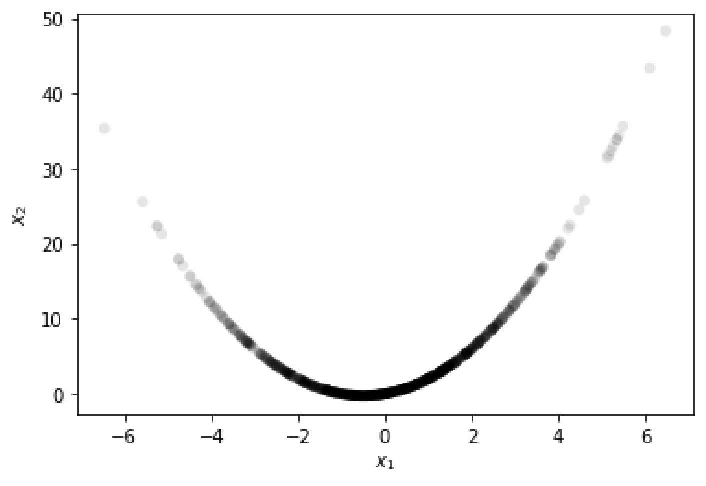
$$\mu = E(x) = \int x\, p(x)\, dx$$

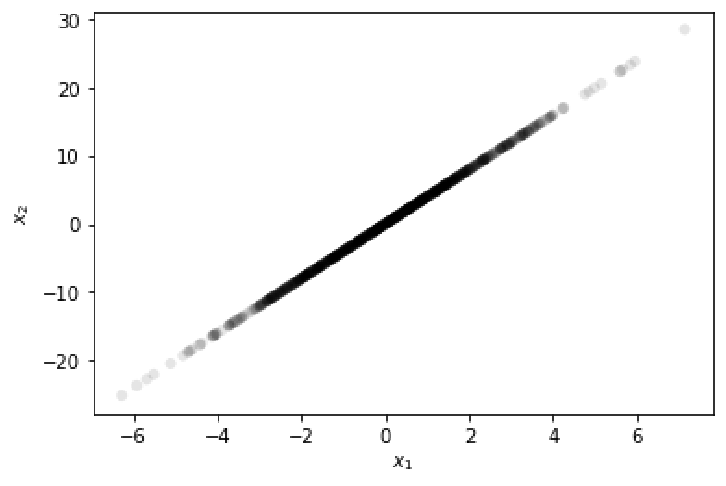
$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

plotting two sets of random number: x_1 and x_2

```
x1 = np.random.normal(0,2,(1000,))
x2 = x1**2 + x1
#x2 = 4*x1

plt.scatter(x1, x2, color = 'k', alpha = 0.1, edgecolor = 'none')
plt.xlabel('$x_1$')
plt.ylabel('$x_2$')
```

based on the shape of the data cloud
→ prediction how x_1 and x_2 are related, i. e. how they **correlate**



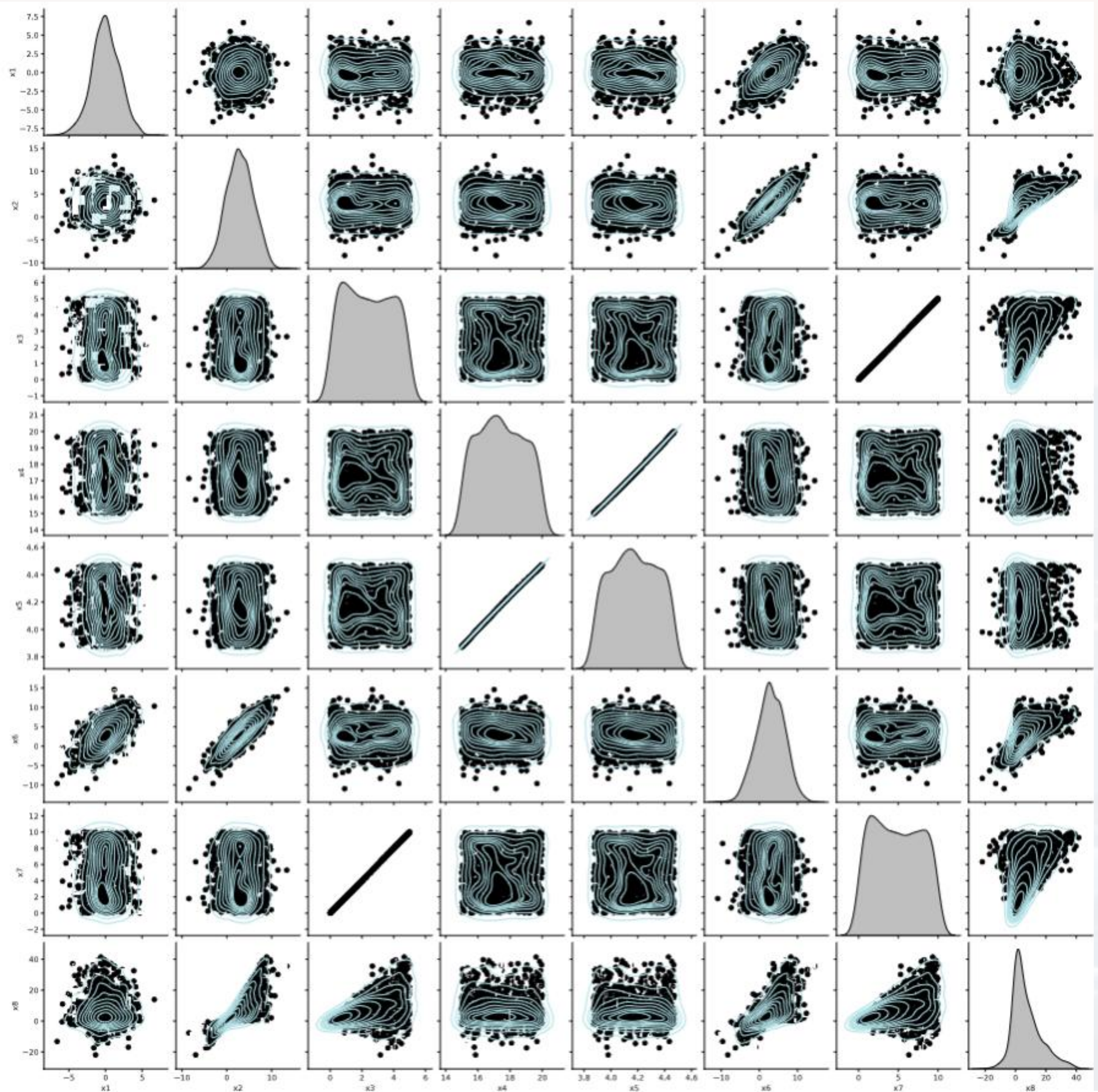


$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

$$\boldsymbol{\sigma^2 = E(x^2) - E(x)^2}$$

```
x1 = np.random.normal(0,2,(1000,))
x2 = np.random.normal(3,3,(1000,))

x3 = np.random.uniform(0,5,(1000,))
x4 = 5*np.random.uniform(3,4,(1000,))

x5 = np.sqrt(x4)
x6 = x1 + x2
x7 = 2*x3
x8 = x3*x2

All  = np.vstack((x1, x2, x3, x4, x5, x6, x7, x8))
data = pd.DataFrame(All.transpose(),
                    columns = ['x1', 'x2', 'x3', 'x4', 'x5', 'x6', 'x7', 'x8'])

out = sns.pairplot(data, kind = "kde", \
                   plot_kws = {'color':[176/255, 224/255, 230/255]}, \
                   diag_kws = {'color':'black'})
out.map_offdiag(plt.scatter, color = 'black')
```



$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

based on the shape of the data cloud

→ prediction how x_1 and x_2 are related, i. e. how they **correlate**

→ how to quantify?

a, b = const

$$\boxed{\sigma^2 = E(x^2) - E(x)^2}$$

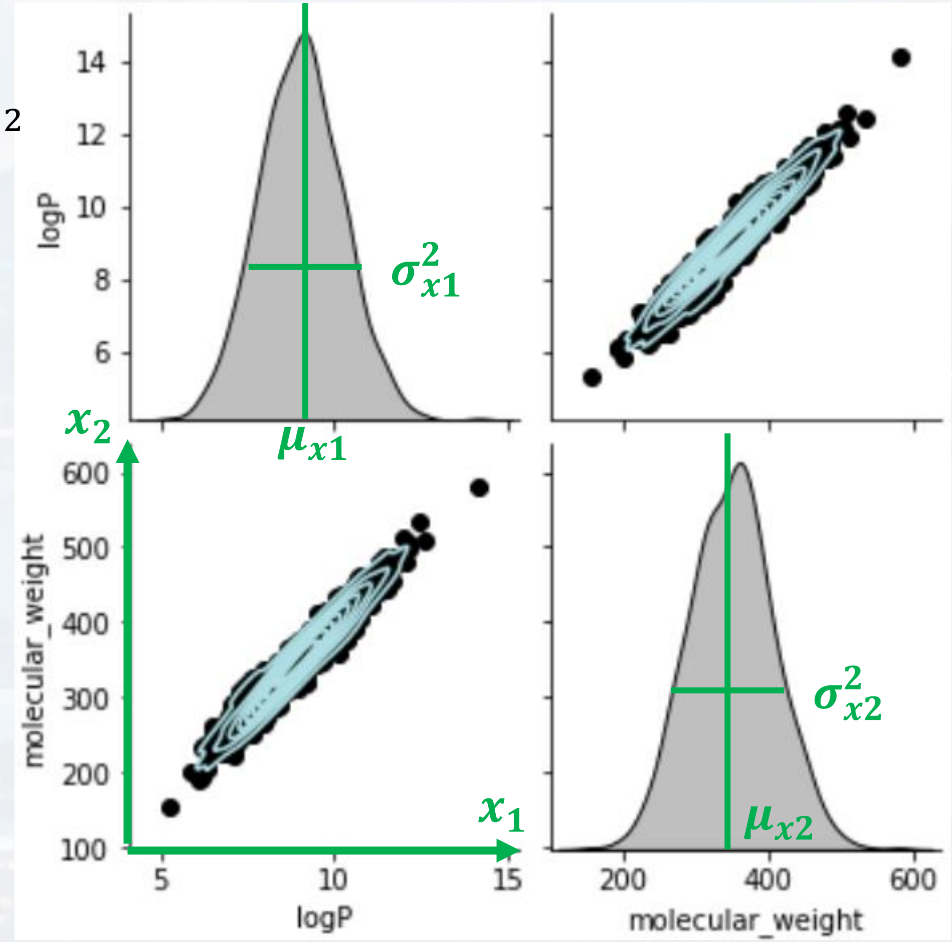
$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

$$var([a\,x_1 + b\,x_2]) = E([a\,x_1 + b\,x_2]^2) - E(a\,x_1 + b\,x_2)^2$$

$$= E(a^2\,x_1^2 + 2ab\,x_1 x_2 + b^2\,x_2^2) - E(a\,x_1 + b\,x_2)^2$$

$$= a^2 E(x_1^2) + 2ab\,E(x_1 x_2) + b^2\,E(x_2^2) - E(a\,x_1 + b\,x_2)^2$$



$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

a, b = const

$$\boxed{\boldsymbol{\sigma^2 = E(x^2) \; - \; E(x)^2}}$$

$$var([a\,x_1 + b\,x_2]) = E([a\,x_1 + b\,x_2]^2) - \; E(a\,x_1 + b\,x_2)^2$$

$$= E(a^2\,x_1^2 + 2ab\,x_1x_2 + b^2\,x_2^2) \; - \; E(a\,x_1 + b\,x_2)^2$$

$$= a^2E(x_1^2) + 2ab\,E(x_1x_2) + b^2\,E(x_2^2) \; - \; E(a\,x_1 + b\,x_2)^2$$

$$= a^2E(x_1^2) + 2ab\,E(x_1x_2) + b^2\,E(x_2^2) \; - \; [aE(x_1) + b\,E(x_2)]^2$$

$$= \underbrace{a^2E(x_1^2) - a^2E(x_1)^2}_{\boldsymbol{a^2\,var(x_1)}} + \underbrace{b^2\,E(x_2^2) \; - b^2E(x_2)^2}_{\boldsymbol{b^2\,var(x_2)}} + \underbrace{2ab\,E(x_1x_2) - 2abE(x_1)E(x_2)}_{\boldsymbol{2ab\,cov(x_1,x_2)}}$$

$$= \boldsymbol{a^2\,var(x_1)} + \boldsymbol{b^2\,var(x_2)} + \boldsymbol{2ab\,cov(x_1,x_2)}$$

$$\boxed{\boldsymbol{cov(x_1,x_2)} = \boldsymbol{cov(x_2,x_1)} = E(x_1x_2) - E(x_1)E(x_2)}$$

covariance

$$\boldsymbol{cov(x_1, x_2) = cov(x_2, x_1)} = E(x_1 x_2) - E(x_1)E(x_2)$$

$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

What does the covariance tell us?

1) geometrical interpretation → next lectures
2) arithmetical interpretation

$$\boldsymbol{cov(x_1, x_2) = cov(x_2, x_1)} = E(x_1 x_2) - E(x_1)E(x_2)$$

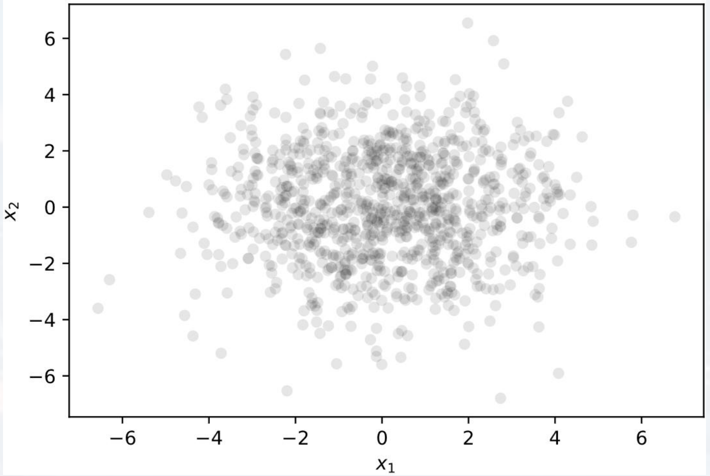
$$\mu = E(x) = \int x\, p(x)\, dx$$

$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

What does the covariance tell us?

1) geometrical interpretation → next lectures

2) arithmetical interpretation



a) x_1 and x_2 are independent

$$E(x_1 x_2) - E(x_1)E(x_2)$$

$$= \iint x_1 x_2\, p(x_1)\, p(x_2)\, dx_1 dx_2 - \int x_1\, p(x_1)\, dx_1 \int x_2\, p(x_2)\, dx_2$$

x_1 and x_2 are independent: x_1 is not a function of x_2 and vise verse

x_1 cannot be predicted by x_2 and vise verse

$$= \int x_1\, p(x_1)\, dx_1 \int x_2\, p(x_2)\, dx_2 - \int x_1\, p(x_1)\, dx_1 \int x_2\, p(x_2)\, dx_2 = 0$$

Covariance equals **zero** if samples are **independent**!

$$\boldsymbol{cov(x_1, x_2) = cov(x_2, x_1)} = E(x_1 x_2) - E(x_1)E(x_2)$$

$$\mu = E(x) = \int x\, p(x)\, dx$$

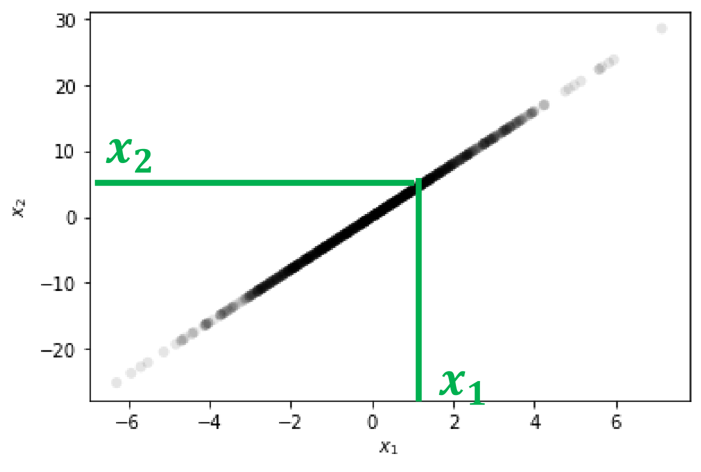
$$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

What does the covariance tell us?

1) geometrical interpretation → next lectures

2) arithmetical interpretation

b) x_1 and x_2 are **not** independent



$$E(x_1 x_2) - E(x_1)E(x_2)$$

$$= \iint x_1 x_2\, p(x_1)\, p(x_2)\, dx_1 dx_2 - \int x_1\, p(x_1)\, dx_1 \int x_2\, p(x_2)\, dx_2$$

x_1 and x_2 are **not** independent:

x_1 ***is*** a function of x_2 and vise verse

x_1 ***can*** be predicted by x_2 to certain degree and vise verse

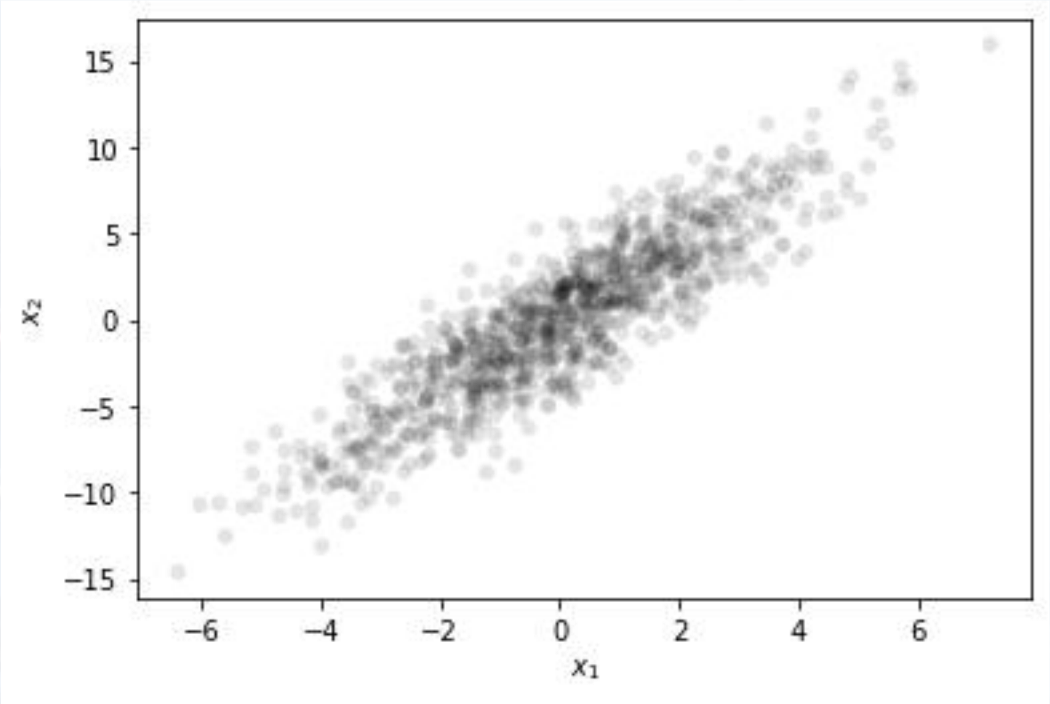
$$= \iint x_1\, p(x_1)\ \ x_2(x_1)\, p(x_2(x_1))\, dx_1 dx_2(x_1) - \int x_1\, p(x_1)\, dx_1 \int x_2\, p(x_2)\, dx_2$$

Covariance **does not** equal **zero**!

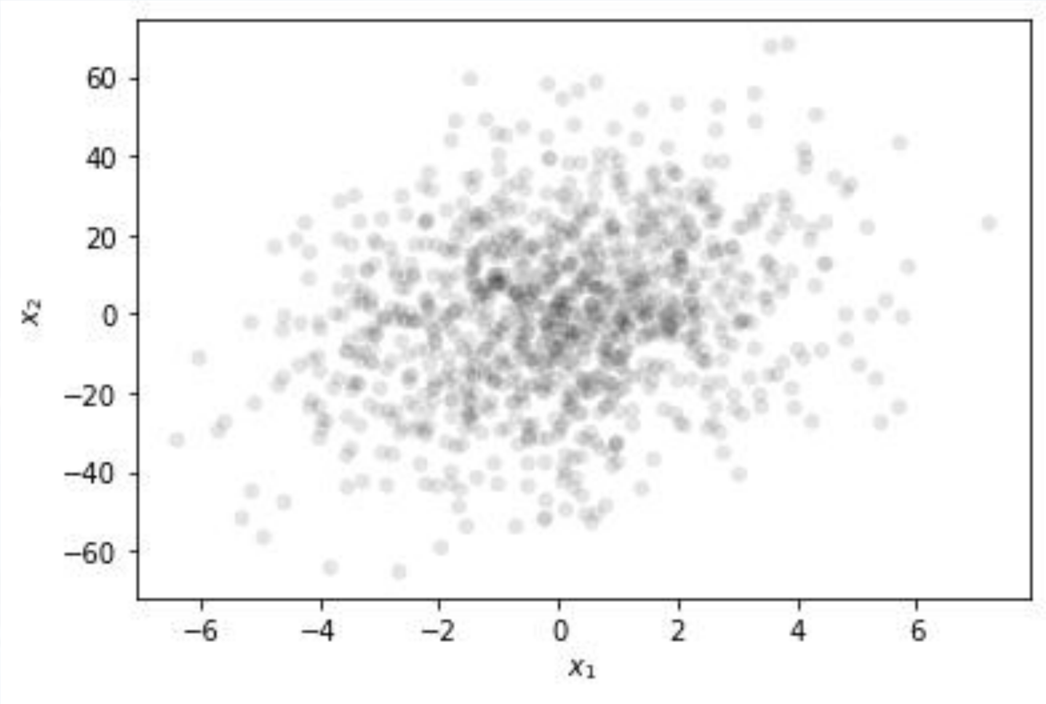
$$cov(x_1, x_2) = cov(x_2, x_1) = E(x_1 x_2) - E(x_1)E(x_2)$$

covariance

```
x1 = np.random.normal(0, 2, (1000,))
x2 = 2*x1 + np.random.normal(0, 2, (1000,))
```



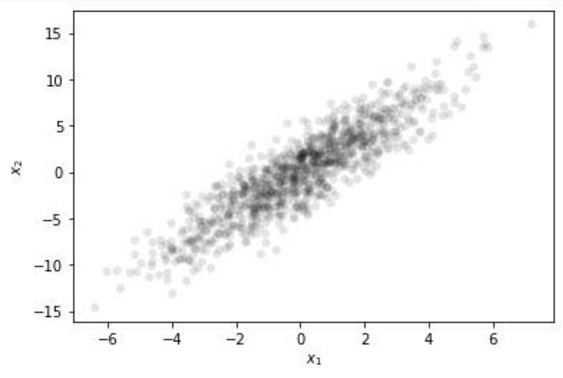

```
x1 = np.random.normal(0, 2, (1000,))
x2 = 2*x1 + np.random.normal(0, 20, (1000,))
```

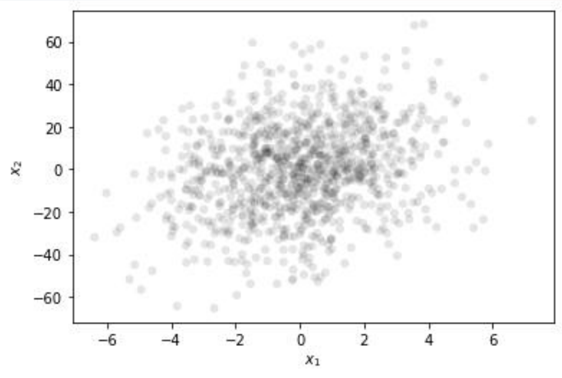


Same dependency, but different variance!

$$cov(x_1, x_2) = cov(x_2, x_1) = E(x_1 x_2) - E(x_1)E(x_2)$$

covariance





Same dependency, but different variance!

Need to scale for the variance!

Pearson's correlation coefficient

$$\rho(x_1, x_2) = \frac{cov(x_1, x_2)}{\sqrt{\sigma_1^2 \sigma_2^2}}$$

$\rho(x_1, x_2)$:

- ranges from -1 to +1
- zero: no correlation
  (completely independent)
- -1: max anti correlation
- +1: max correlation

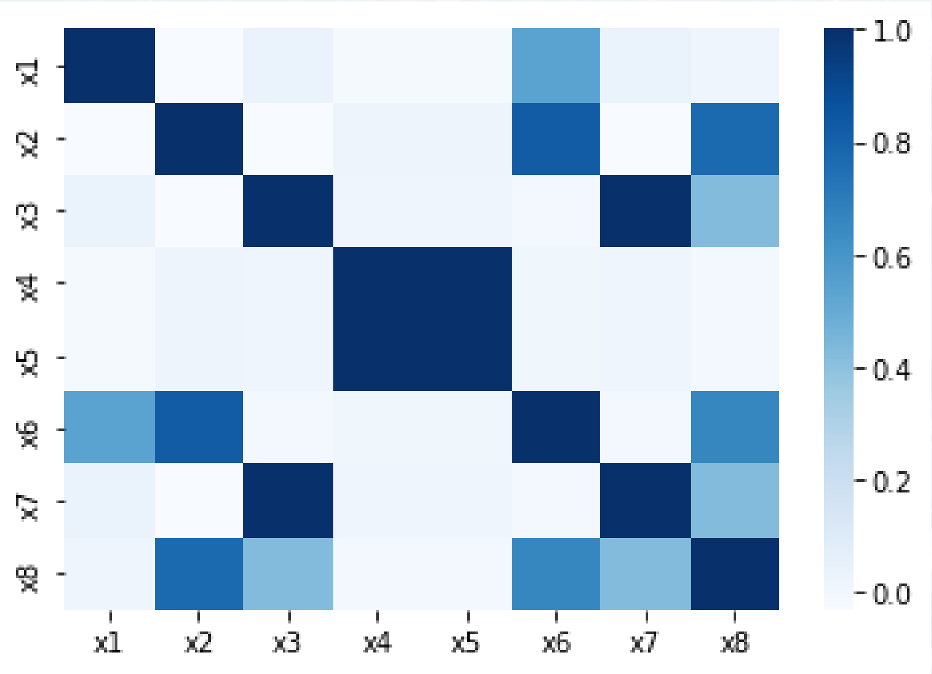
$$cov(x_1, x_2) = cov(x_2, x_1) = E(x_1 x_2) - E(x_1)E(x_2)$$

covariance

$$\rho(x_1, x_2) = \frac{cov(x_1, x_2)}{\sqrt{\sigma_1^2 \sigma_2^2}}$$

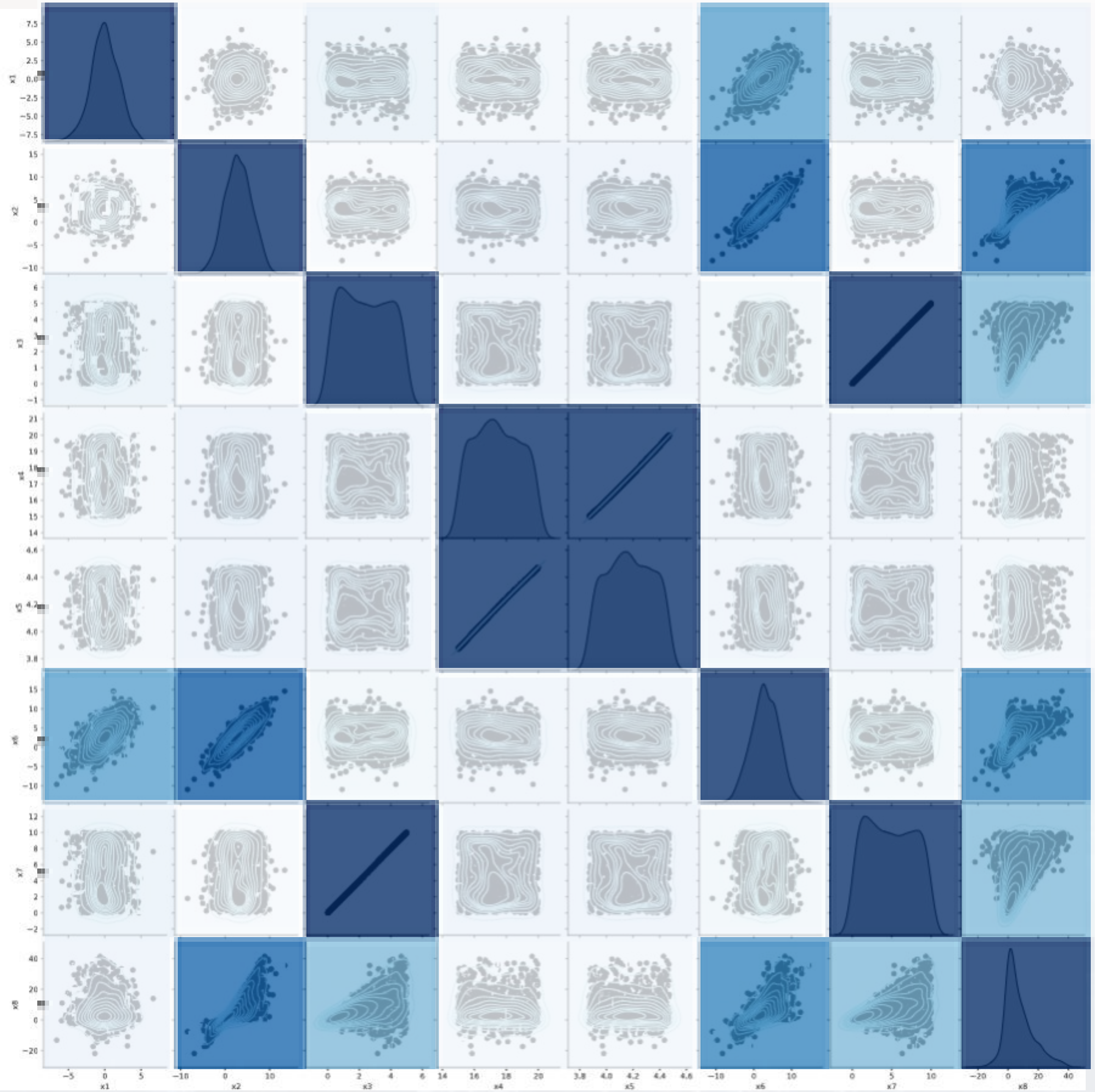
Pearson's correlation coefficient

```
sns.heatmap(data.corr(), cmap = "Blues")
```



$\rho(x_1, x_2)$:

- ranges from -1 to +1
- zero: no correlation
  (completely independent)
- -1: max anti correlation
- +1: max correlation



$\rho(x_1, x_2)$:

- ranges from -1 to +1
- zero: no correlation

  (completely independent)
- -1: max anti correlation
- +1: max correlation

Important quantities you should know:

mean $$\mu = E(x) = \int x\, p(x)\, dx$$

median *m* $$\int_a^m p(x)\, dx = \frac{1}{2}$$

variance $$\sigma^2 = var(x) = \int (x-\mu)^2\, p(x)\, dx$$

$$\sigma^2 = E(x^2) - E(x)^2$$

$$\sigma_{tot}^2 = \sigma_1^2 + \sigma_2^2 + 2\, cov(x_1, x_2)$$

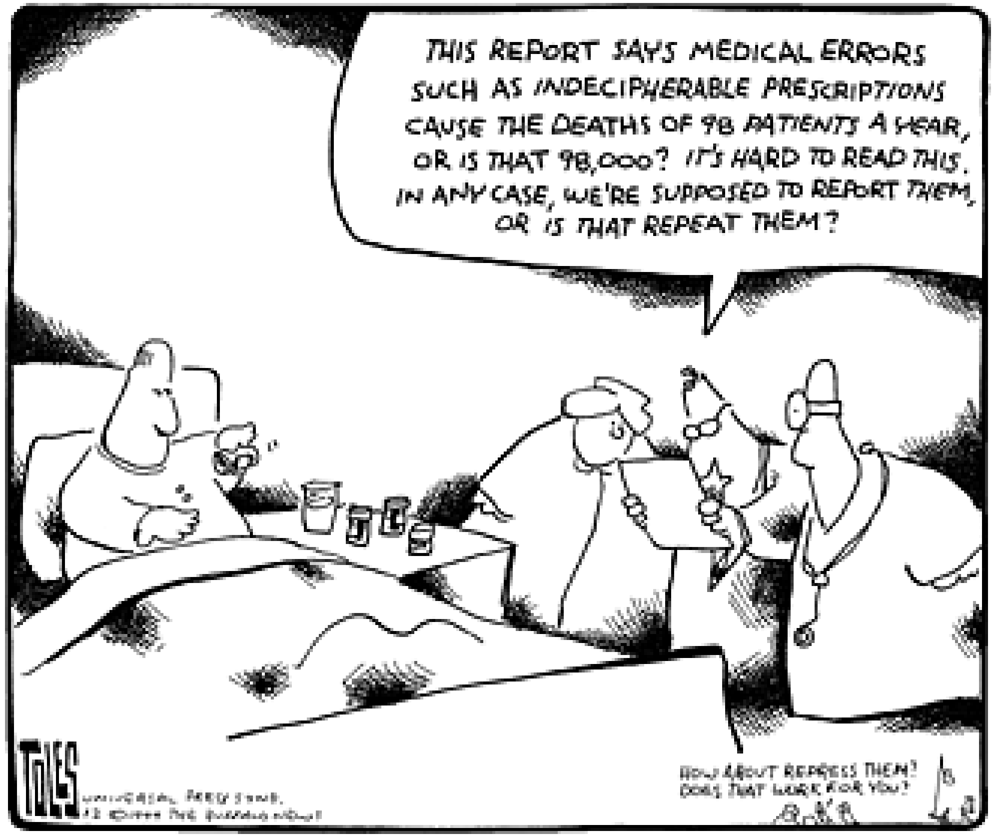
covariance $$cov(x_1, x_2) = E(x_1 x_2) - E(x_1)E(x_2)$$

correlation coefficient $$\rho(x_1, x_2) = \frac{cov(x_1, x_2)}{\sqrt{\sigma_1^2 \sigma_2^2}}$$

note:

$$\int (x-\mu)^n\, p(x)\, dx$$

called n-th *moment* of a pdf



Outline:

Basics

Most Common PDFs

- **uniform**
- binomial
- Poissonian
- Normal/Gaussian

Error Estimation

Bayesian Statistics

finding those $p(x)$ that maximize the **entropy S**, given constrains **C**

$$S = -\int p(x)\ln[p(x)]\,dx$$

C: $\int p(x)\,dx = 1$ $\longrightarrow$ $p(x) = const$

$$\mu = \int_a^b x\,p(x)\,dx = const\int_a^b x\,dx = const\,\frac{1}{2}(b^2 - a^2)$$

$$\int_a^b p(x)\,dx = 1 \qquad const\int_a^b dx = 1 \qquad \rightarrow const = \frac{1}{b-a}$$

$$\boxed{\mu = \frac{1}{2}\frac{b^2 - a^2}{b-a}}$$

$$\boxed{\sigma^2 = \frac{1}{12}(b-a)^2}$$

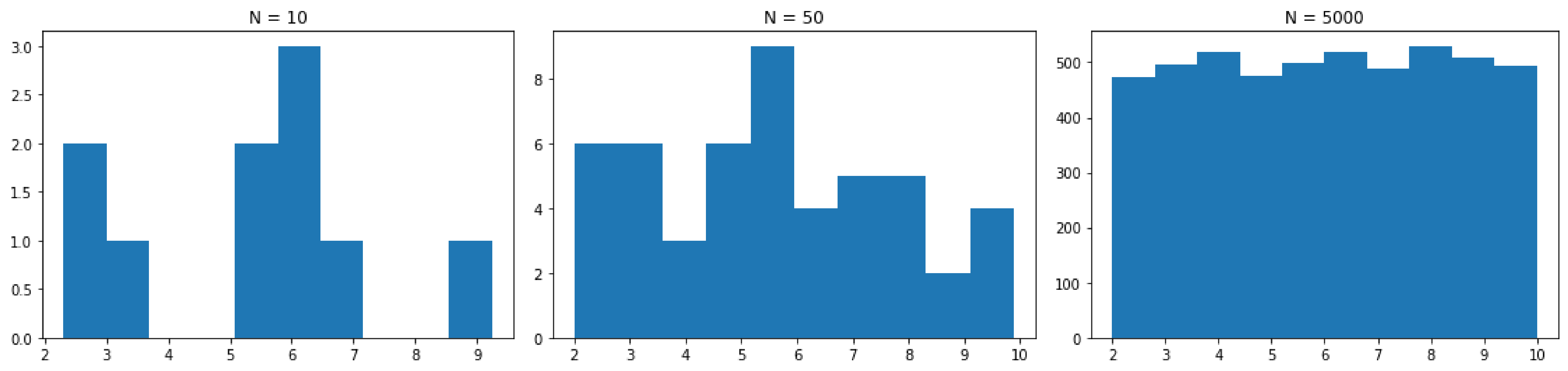
Note: the uniform distribution has the largest entropy
→ maximum ignorance = no prior information = unbiased

$p(x) = const$

plotting the pdf

```
U = np.random.uniform(low, high, shape)
plt.hist(U)
```

continuous support

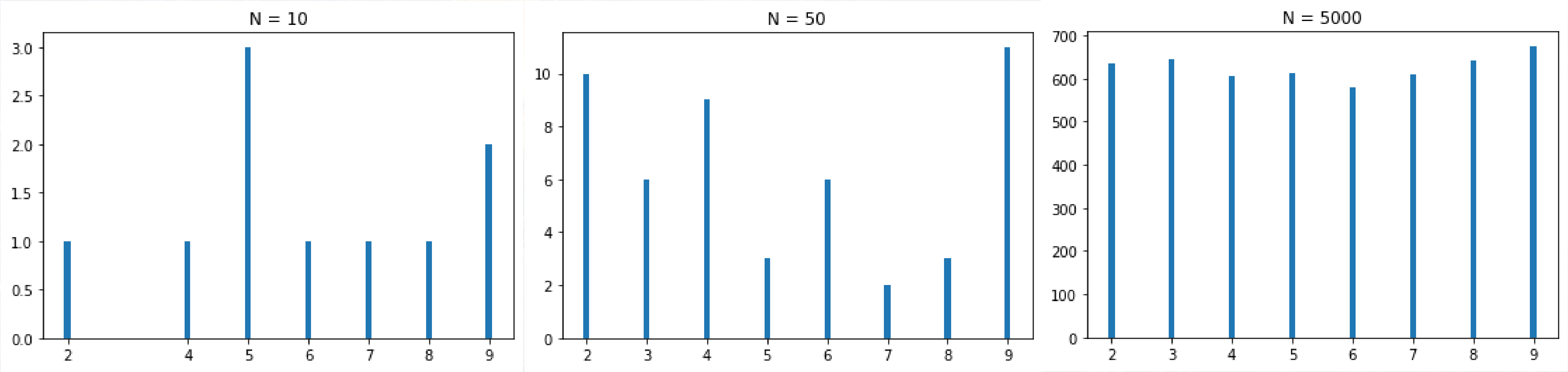


$p(x) = const$

plotting the pdf

```
U = np.random.randint(low, high, shape)

labels, counts = np.unique(U, return_counts = True)
plt.bar(labels, counts, align = 'center', width = 0.1)
plt.gca().set_xticks(labels)
plt.title('N = ' + str(N))
```

discrete support



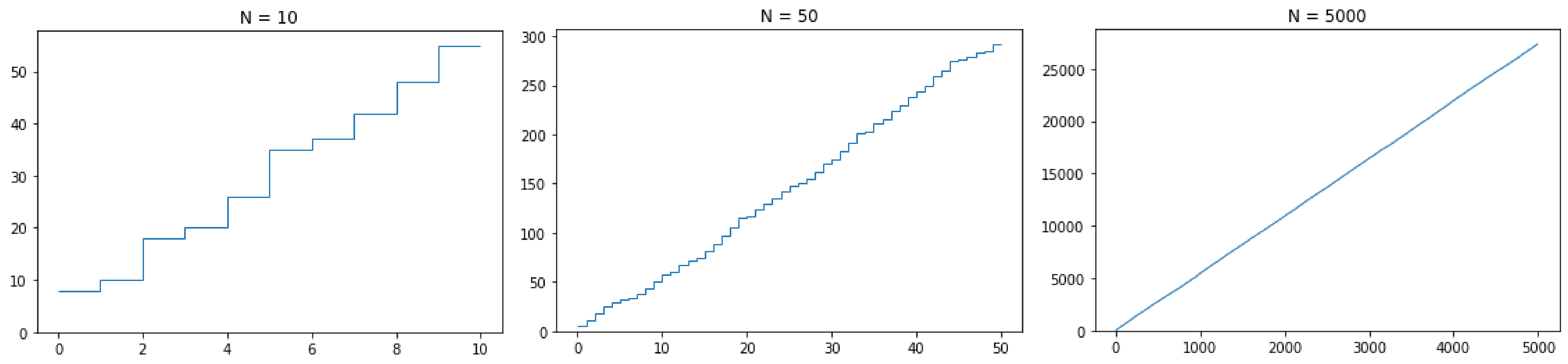
$p(x) = const$

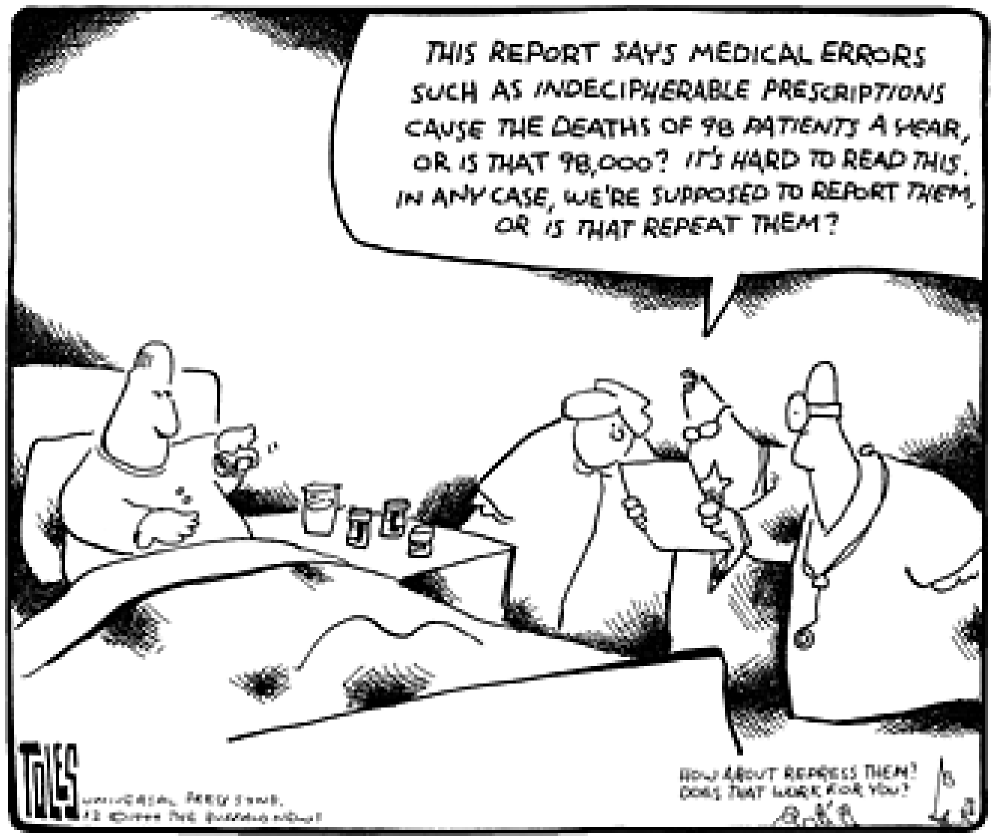
plotting the cdf

```
U = np.random.randint(low, high, shape)
```

discrete support

```
C = np.cumsum(U)
plt.stairs(C, baseline = None)
plt.title('N = ' + str(N))
```





Outline:

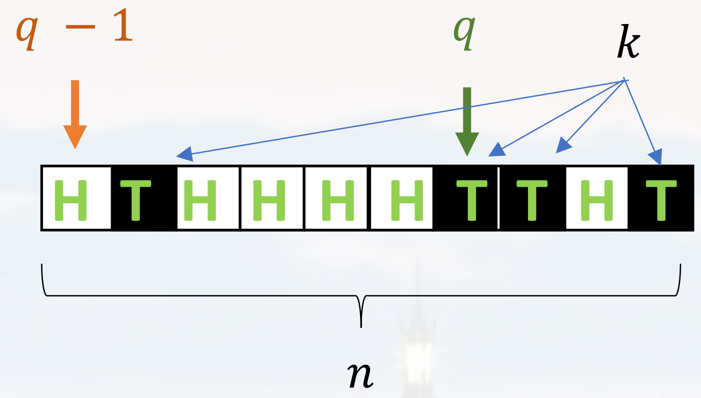
Basics

Most Common PDFs

- uniform
- **binomial**
- Poissonian
- Normal/Gaussian

Error Estimation

Bayesian Statistics



probability of having **a** sequence of **k tails** and **n-k heads**

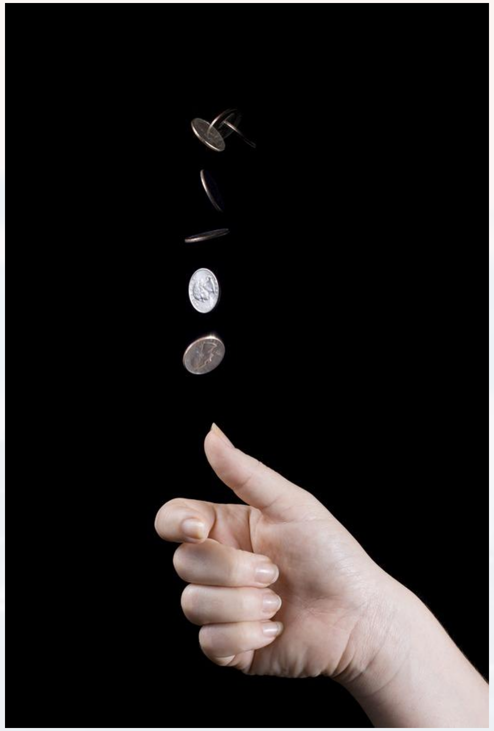
$$p_{tot} = \prod_i {q_i}^{n_i} = q^k \, (1-q)^{n-k}$$

probability of having **any** sequence of **k tails** and **n-k heads**

$$\boxed{P(k|q,n) = \binom{n}{k} \, q^k \, (1-q)^{n-k}}$$

binomial distribution

$$\frac{n!}{k!(n-k)!} =: \binom{n}{k}$$ **``n choose k''**



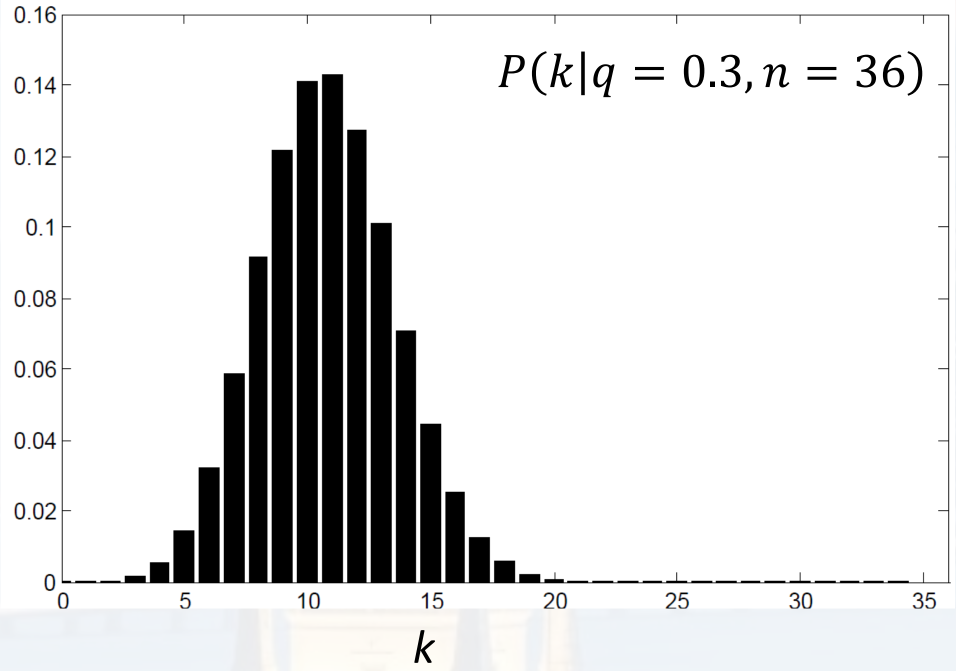
fair coin? q = 0.5 ???

$$\boxed{P(k|q,n) = \binom{n}{k}\, q^k\, (1-q)^{n-k}}$$

binomial distribution

$$\mu = \sum_i x_i\, p(x_i)$$

$$\mu = \int x\, p(x)\, dx$$



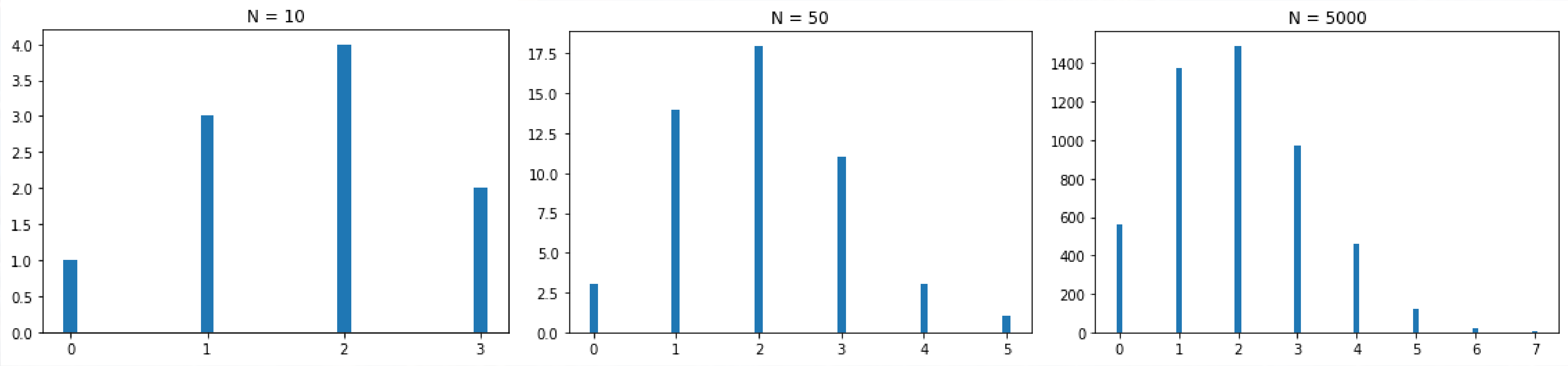
$$\mu = \sum_{k=0}^{n} k \binom{n}{k} q^k (1-q)^{n-k} = \boldsymbol{qn}$$

$$var(k) = \sum_{k=0}^{n} (k - qn)^2 \binom{n}{k}\, q^k\, (1-q)^{n-k} = \boldsymbol{qn(1-q)}$$

$$P(k|q,n) = \binom{n}{k} q^k (1-q)^{n-k}$$

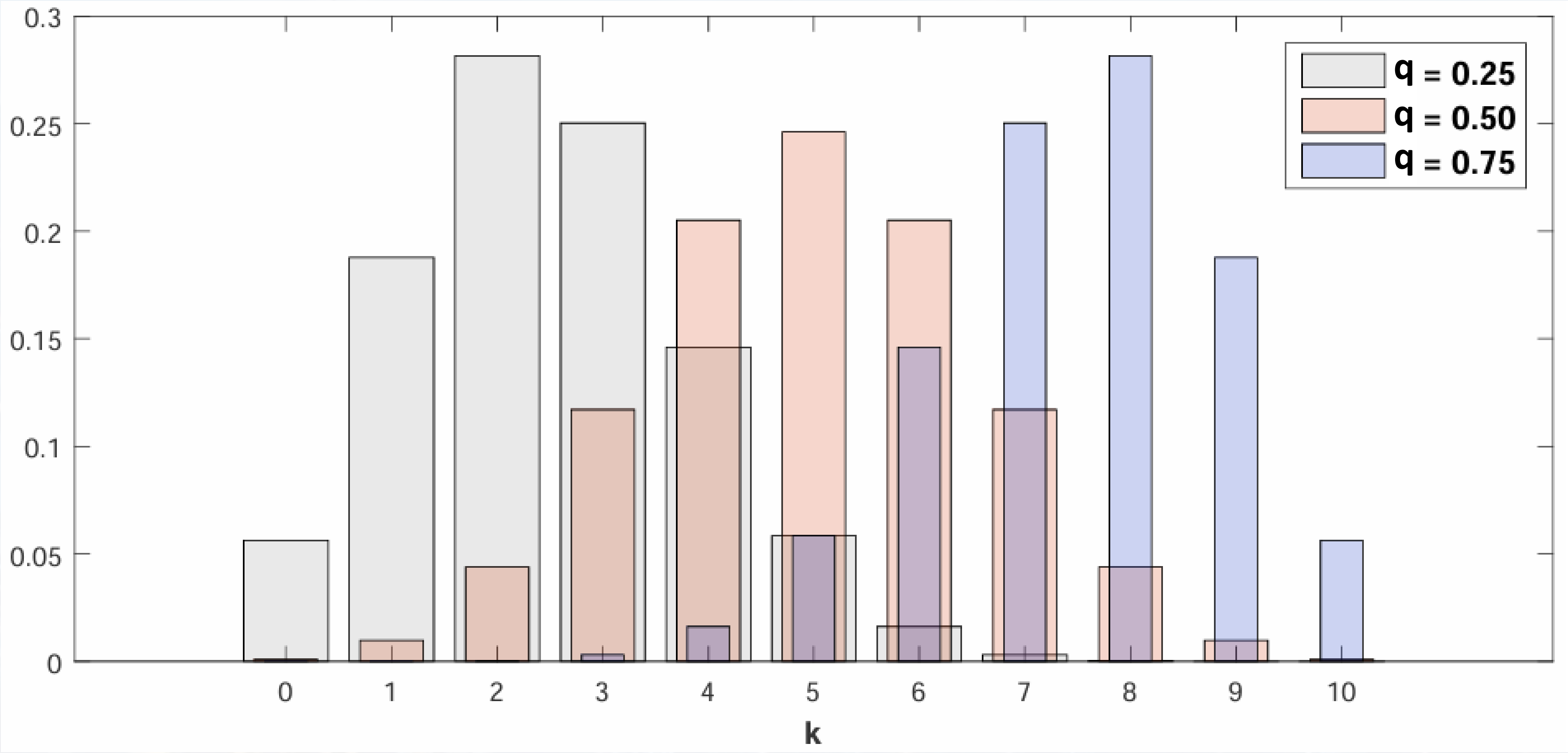
binomial distribution

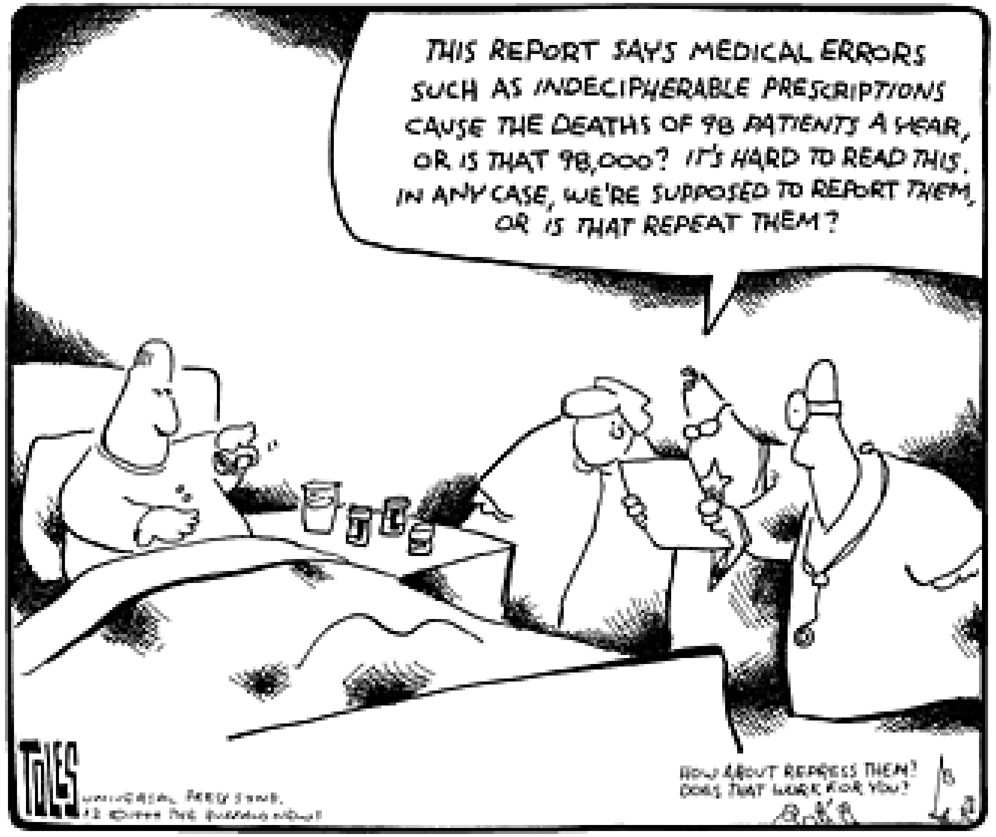
```
q = 0.2
n = 10
K = np.random.binomial(n, q, N)

labels, counts = np.unique(K, return_counts = True)
plt.bar(labels, counts, align = 'center', width = 0.1)
plt.gca().set_xticks(labels)
plt.title('N = ' + str(N))
```



$$\boxed{P(k|q,n) = \binom{n}{k}\, q^k\, (1-q)^{n-k}}$$

binomial distribution





Outline:

Basics

Most Common PDFs

- uniform
- binomial
- **Poissonian**
- Normal/Gaussian

Error Estimation

Bayesian Statistics

$$\boxed{P(k|q,n) = \binom{n}{k} q^k (1-q)^{n-k}}$$

binomial distribution

rare events → q << 1

Taylor expansion for $(1-q)^{n-k}$ around q = 0

$$(1-q)^{n-k} = 1 - nq + \frac{(nq)^2}{2} - \frac{(nq)^3}{6} + \ldots = e^{-nq}$$

→ n → ∞

Stirling's approximation for n!

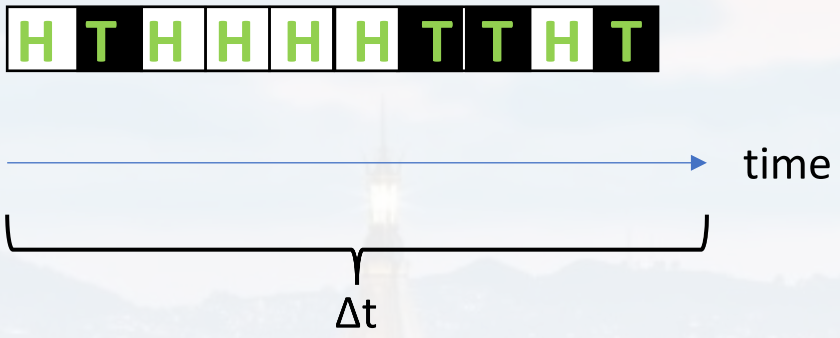
$$\frac{n!}{(n-k)!} \approx \sqrt{\frac{n}{n-k}} \frac{n^n e^{n-k}}{e^n (n-k)^{n-k}} \approx n^k$$

$$\binom{n}{k} q^k (1-q)^{n-k} \approx \frac{(nq)^k e^{-nq}}{k!}$$

$$\binom{n}{k} q^k (1-q)^{n-k} \approx \frac{(nq)^k e^{-nq}}{k!}$$

often: nq : = λ

events per time interval: λ = c Δt



rate c = 4 tails per Δt

$$\boxed{P(k|c) = \frac{(c\,\Delta t)^k e^{-c\,\Delta t}}{k!}}$$

Poisson distribution

$$\mu = qn \rightarrow \mathbf{qn = \lambda}$$

$$var(k) = qn(1-q) \rightarrow \mathbf{qn = \lambda}$$

$$P(k|c) = \frac{(c\,\Delta t)^k e^{-c\,\Delta t}}{k!}$$

Poisson distribution

$$\mu = \lambda$$
$$var(k) = \lambda$$

- **rare** events
- events are mutually **independent**
- events have **no duration**

<u>examples:</u>

- radioactive decay
- single photon detection
- lightning
- mutation of a gene
- receiving WhatsApp messages/SMS

rare: not that **a** atom decays,
→ that **this** atom decays within Δt

$$\boxed{P(k|c) = \frac{(c\,\Delta t)^k e^{-c\,\Delta t}}{k!}}$$

Poisson distribution

$$\mu = \lambda$$

$$var(k) = \lambda$$

```
c    = 5
delt = 10
lam  = c * delt

K    = np.random.poisson(lam, N)

labels, counts = np.unique(K, return_counts = True)
plt.bar(labels, counts, align = 'center', width = 0.1)
plt.gca().set_xticks(labels)
plt.title('N = ' + str(N))
```

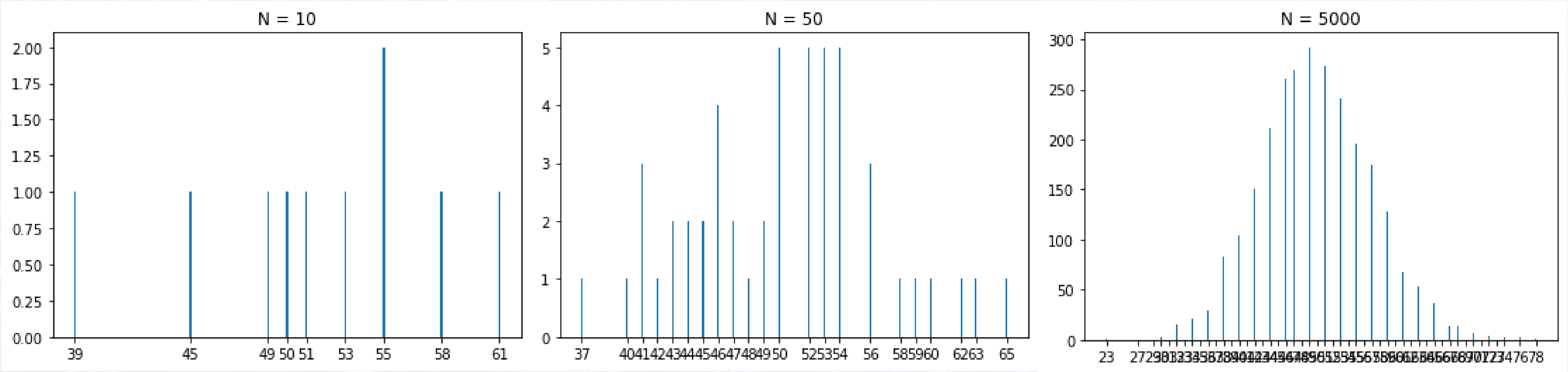
$$\boxed{P(k|c) = \frac{(c\,\Delta t)^k e^{-c\,\Delta t}}{k!}}$$

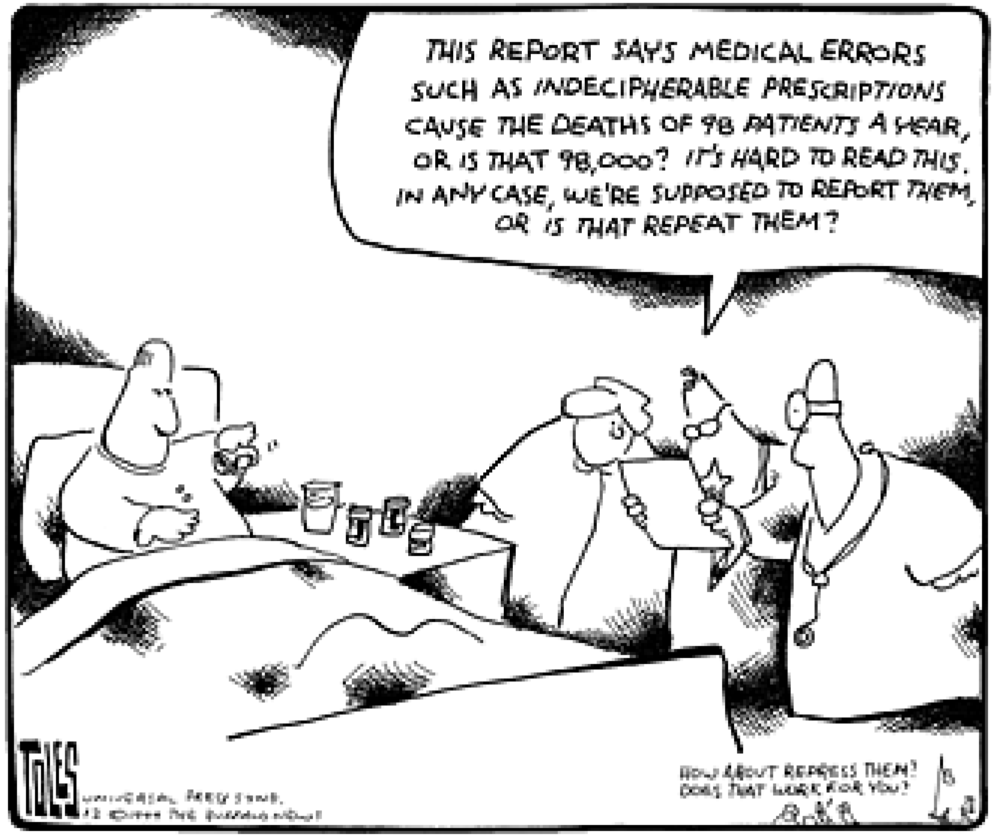
Poisson distribution

$$\mu = \lambda$$

$$var(k) = \lambda$$

```
c    = 5
delt = 10
lam  = c * delt
K    = np.random.poisson(lam, N)
labels, counts = np.unique(K, return_counts = True)
plt.bar(labels, counts, align = 'center', width = 0.1)
plt.gca().set_xticks(labels)
plt.title('N = ' + str(N))
```





Outline:

Basics

Most Common PDFs

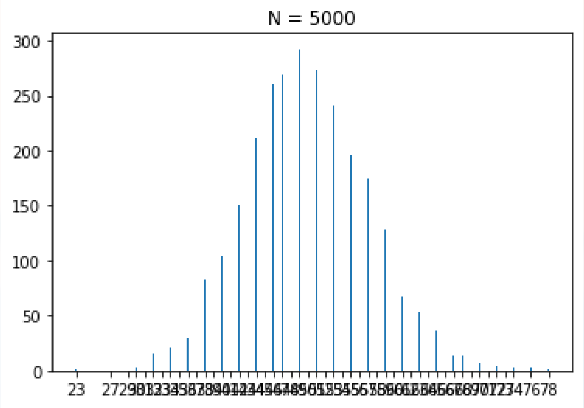
- uniform
- binomial
- Poissonian
- **Normal/Gaussian**

Error Estimation

Bayesian Statistics

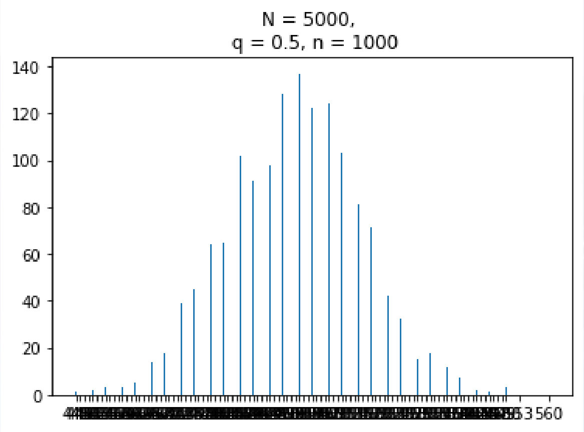
$$\boxed{P(k|c) = \frac{(c\,\Delta t)^k e^{-c\,\Delta t}}{k!}}$$

Poisson distribution



$$\boxed{P(k|q,n) = \binom{n}{k}\, q^k\, (1-q)^{n-k}}$$

binomial distribution



Stirling's approximation for even larger n

$$P(k|n,p) \approx \frac{1}{\sqrt{2\pi\, nq(1-q)}} \exp\left[-\frac{(k-nq)^2}{2nq(1-q)}\right]$$

Stirling's approximation for even larger n

$$P(k|n,p) \approx \frac{1}{\sqrt{2\pi\, nq(1-q)}} \exp\left[-\frac{(k-nq)^2}{2nq(1-q)}\right]$$

using $\sigma^2 = var(k) = qn(1-q)$

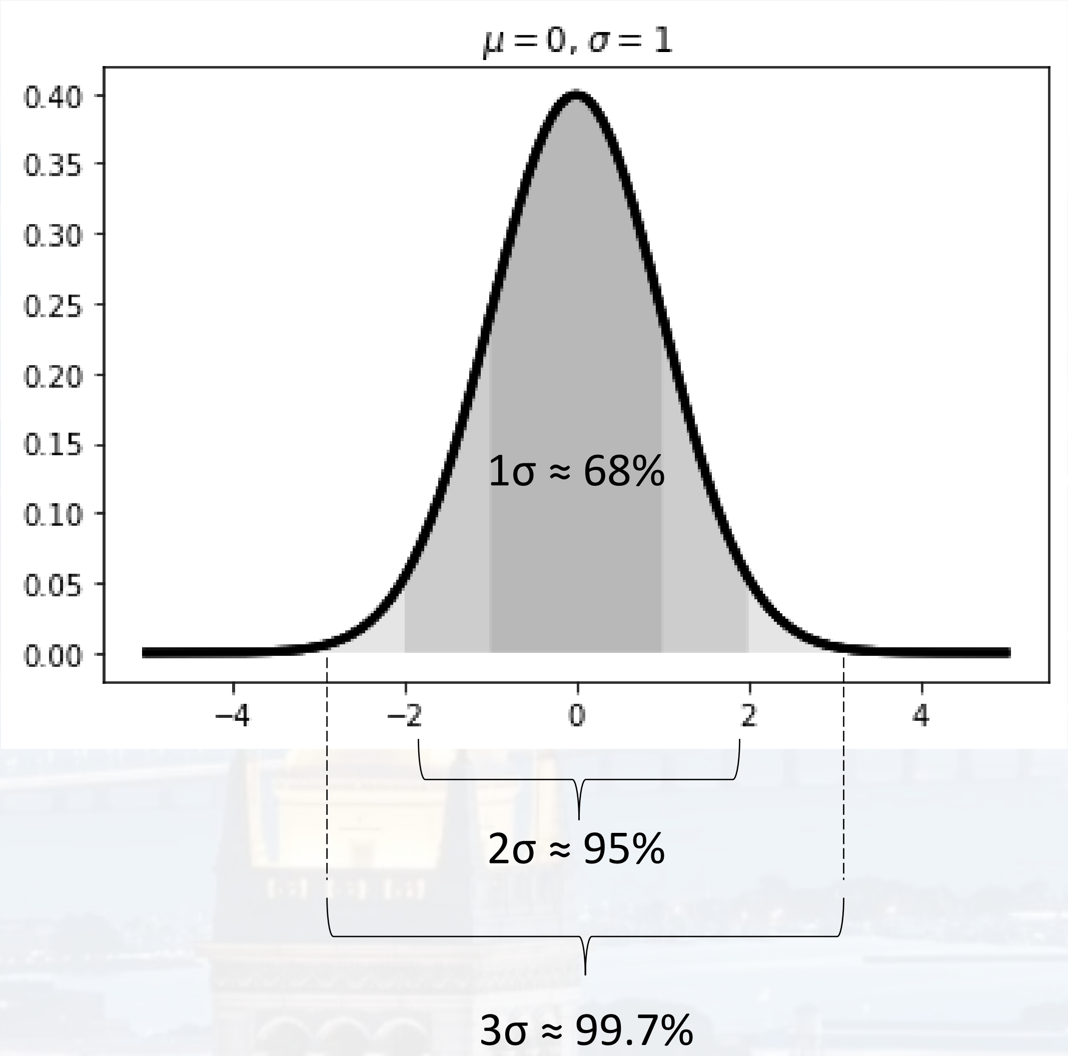
$\mu \;= qn$

and $k := x$

$$\boxed{P(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\, exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]}$$

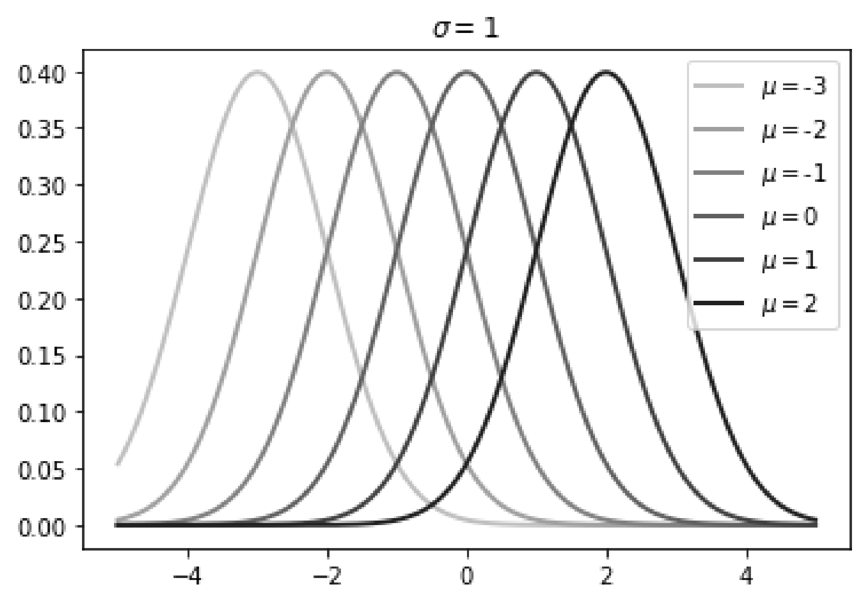
Normal/Gauss distribution

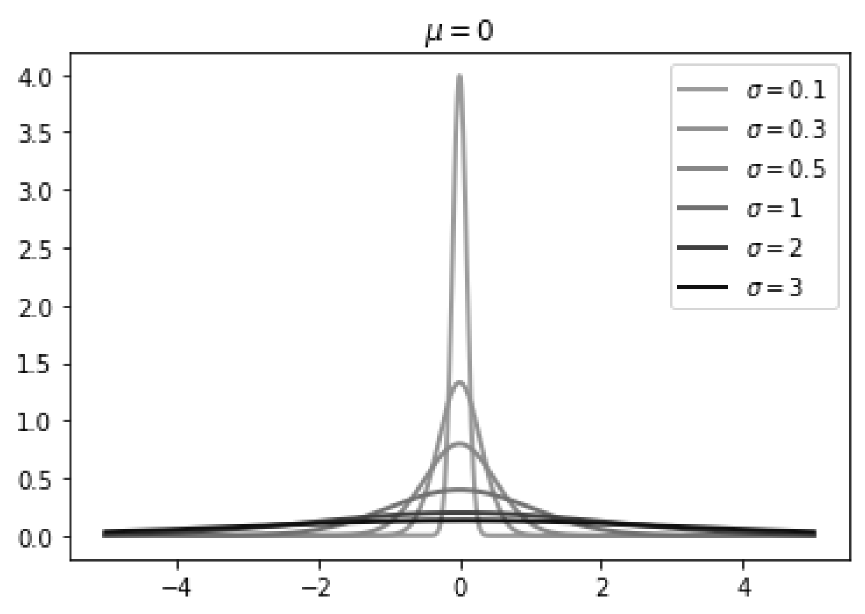
Note, that the **Poisson** and the **Binomial distribution** are *discrete*, whereas the **Normal distribution** is *continuous*!



$$P(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\,\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]$$

Normal/Gauss distribution





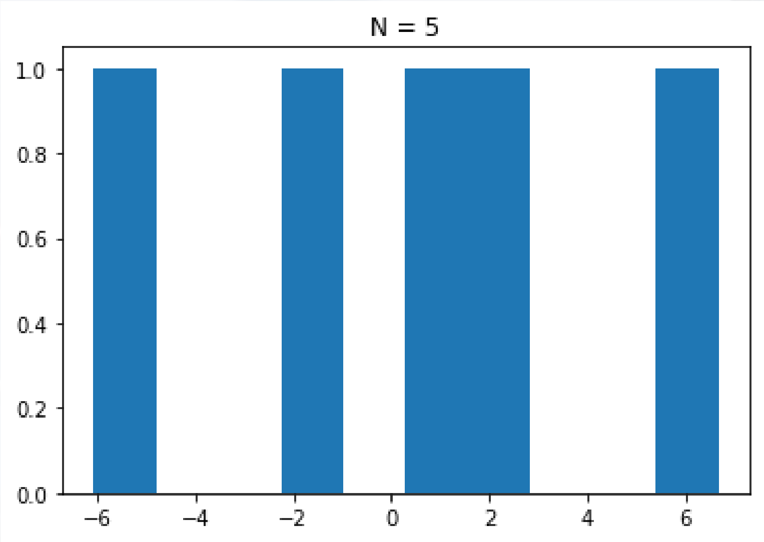
$$\boxed{P(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\,exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]}$$

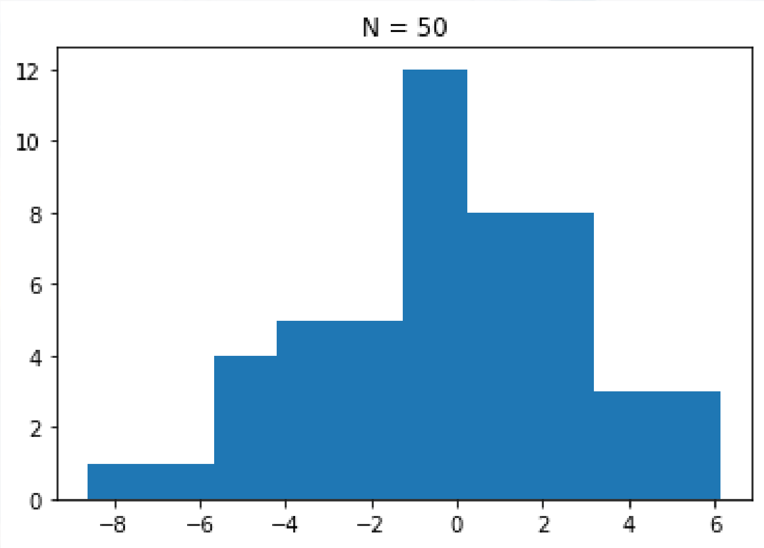
Normal/Gauss distribution

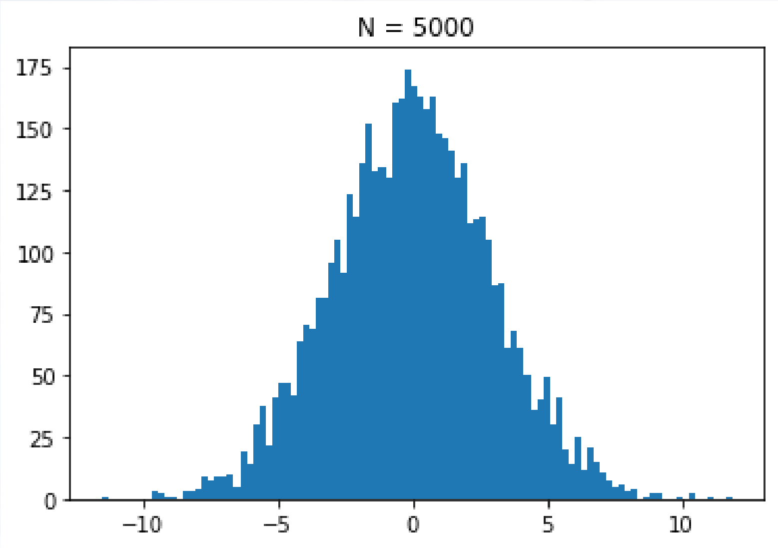
```
mu = 0
s  = 1
P  = np.random.normal(mu, s, N)
plt.hist(P)
plt.title('N = ' + str(N))
```

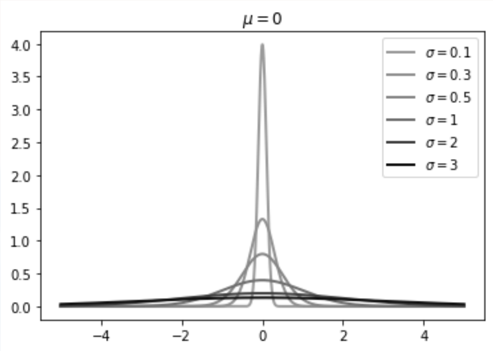
$$P(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\,\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]$$

Normal/Gauss distribution









$$\boxed{P(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\,\sigma^2}}\,exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]}$$

Normal/Gauss distribution

examples:
- diffusion processes
- approx. stat. error of data points
- approx. distribution of body height/shoe sizes/ weight, IQ
- approx. blood pressure, blood values
- approx. retirement age

….

applications:
- significance tests
- t-test
- ANOVA/MANOVA
- χ^2 - test
- χ^2 itself and students-t distribution

…

Why do so many quantities follow a normal distribution?

Why do so many quantities follow a normal distribution?

At the end… all probability distributions are **Maximum Entropy** Distributions, subject to a **set of constrains**

Distribution name	Probability density / mass function	Maximum Entropy constraint	Support
Uniform (discrete)	$f(k)=\frac{1}{b-a+1}$	None	$\{a,a+1,\ldots,b-1,b\}$
Uniform (continuous)	$f(x)=\frac{1}{b-a}$	None	$[a,b]$
Bernoulli	$f(k)=p^k(1-p)^{1-k}$	$\mathbb{E}[\,K\,]=p$	$\{0,1\}$
Geometric	$f(k)=(1-p)^{k-1}\,p$	$\mathbb{E}[\,K\,]=\frac{1}{p}$	$\mathbb{N}\smallsetminus\{0\}=\{1,2,3,\ldots\}$
Exponential	$f(x)=\lambda\exp(-\lambda x)$	$\mathbb{E}[\,X\,]=\frac{1}{\lambda}$	$[0,\infty)$
Laplace	$f(x)=\frac{1}{2b}\exp\left(-\frac{\lvert x-\mu\rvert}{b}\right)$	$\mathbb{E}[\,\lvert X-\mu\rvert\,]=b$	$(-\infty,\infty)$
Asymmetric Laplace	$f(x)=\frac{\lambda\,\exp\left(-(x-m)\,\lambda\,s\,\kappa^s\right)}{\left(\kappa+\frac{1}{\kappa}\right)}$ where $s\equiv\operatorname{sgn}(x-m)$	$\mathbb{E}[\,(X-m)\,s\,\kappa^s\,]=\frac{1}{\lambda}$	$(-\infty,\infty)$
Pareto	$f(x)=\frac{\alpha\,x_m^\alpha}{x^{\alpha+1}}$	$\mathbb{E}[\,\ln X\,]=\frac{1}{\alpha}+\ln(x_m)$	$[x_m,\infty)$
Normal	$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\mathbb{E}[\,X\,]=\mu\,,$ $\mathbb{E}[\,X^2\,]=\sigma^2+\mu^2$	$(-\infty,\infty)$

Why do so many quantities follow a normal distribution?

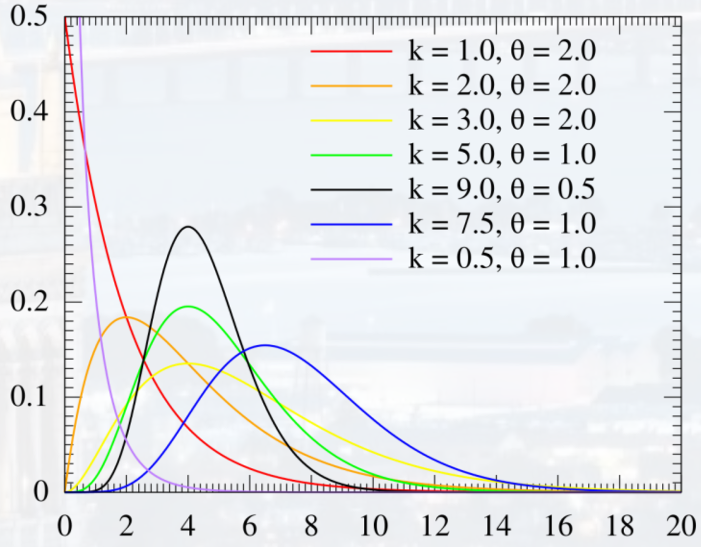
At the end… all probability distributions **are Maximum Entropy** Distributions, subject to a **set of constrains**

<u>examples:</u>

- approx. stat. error of data points
- approx. distribution of body height/shoe sizes/ weight, IQ
- approx. blood pressure, blood values
- approx. retirement age

….

Gamma	$f(x) = \dfrac{x^{k-1}\exp\left(-\frac{x}{\theta}\right)}{\theta^k\,\Gamma(k)}$	$\mathbb{E}[\,X\,] = k\,\theta\,,$ $\mathbb{E}[\,\ln X\,] = \psi(k) + \ln\theta$	$[0, \infty)$



binomial distribution

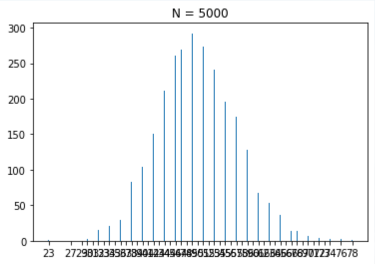
$$P(k|q,n) = \binom{n}{k} q^k (1-q)^{n-k}$$

$q \rightarrow 0$

Poisson distribution

$$P(k|c) = \frac{(c\,\Delta t)^k e^{-c\,\Delta t}}{k!}$$

$n \rightarrow \infty$



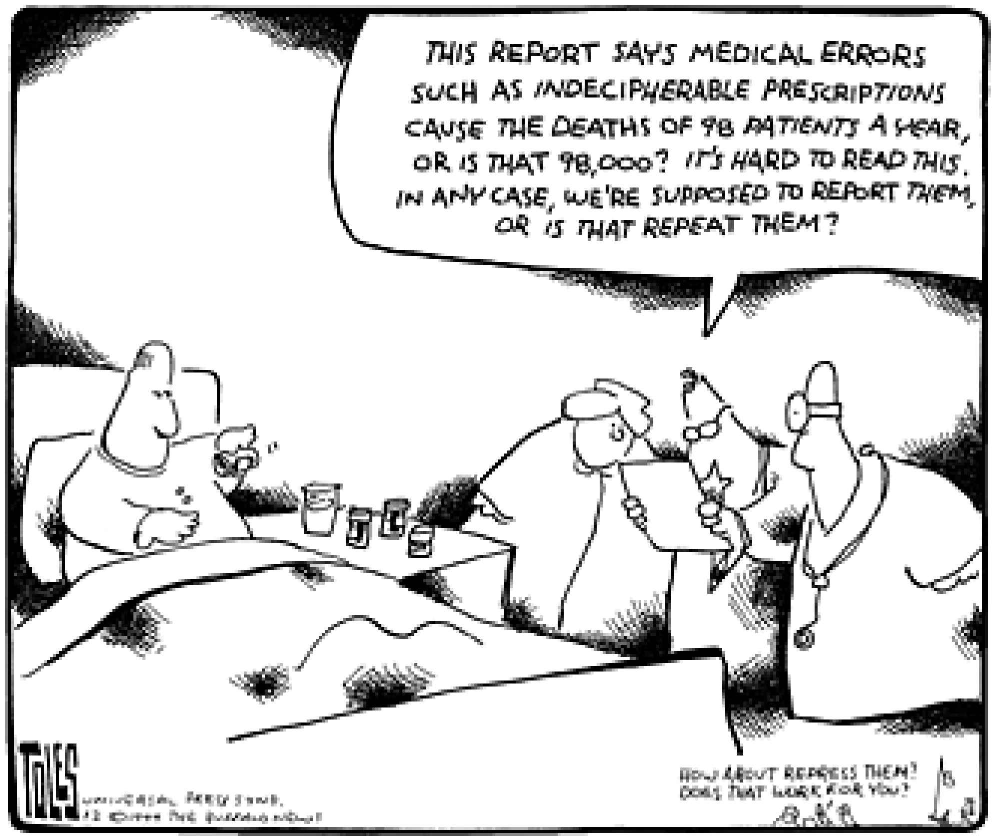
$n \rightarrow \infty$

Normal/Gauss distribution

$$P(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\,\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\,\sigma^2}\right]$$

The fact that that many datasets can be well approximated by a Normal distribution for $n \rightarrow \infty$ is called

Central Limit Theorem



Outline:

Basics

Most Common PDFs

- uniform
- binomial
- Poissonian
- Normal/Gaussian

Error Estimation

Bayesian Statistics