Continuous Probability Distributions

Normal or Gaussian Distribution


PAY
ATTENTION

Normal or Gaussian Distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

$-\infty < x < \infty$

is a normal random variable

with mean μ,

and standard dewviation σ

sometimes denoted as $N(\mu, \sigma)$

Carl Friedrich Gauss (1777 –1855)
German mathematician

Normal Distribution

- The location and spread of the normal are independently determined by mean (μ) and standard deviation (σ)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

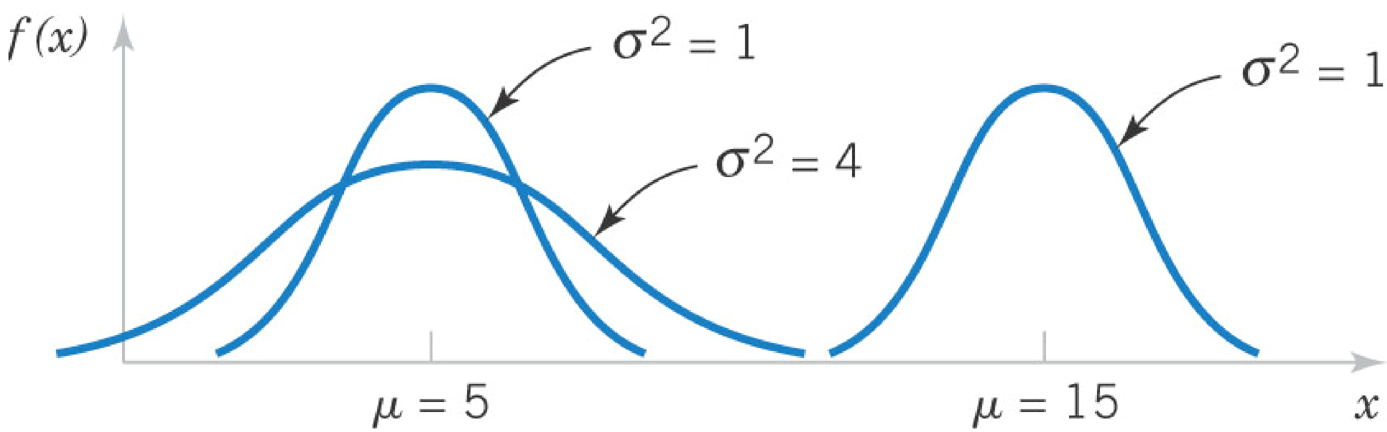
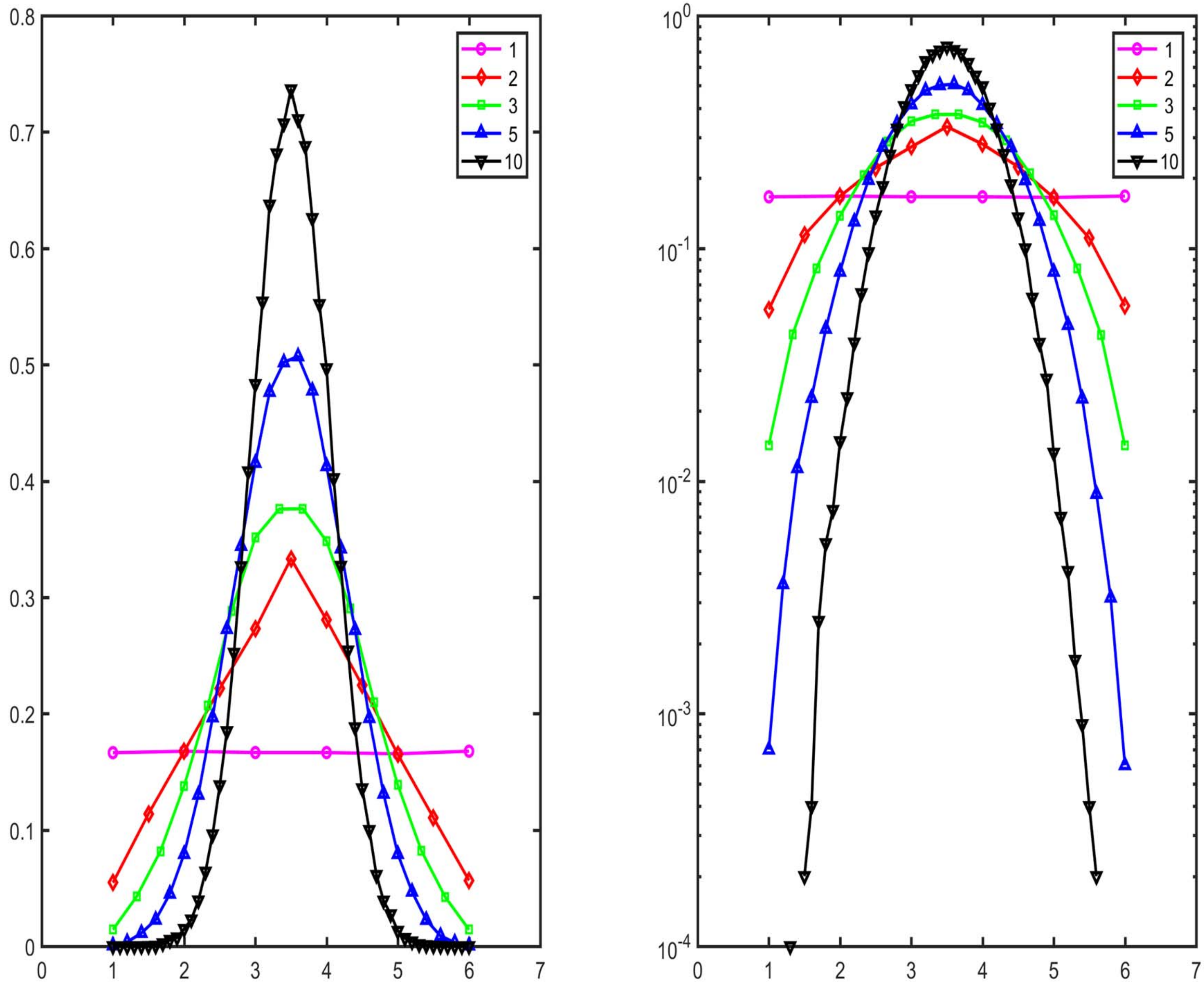


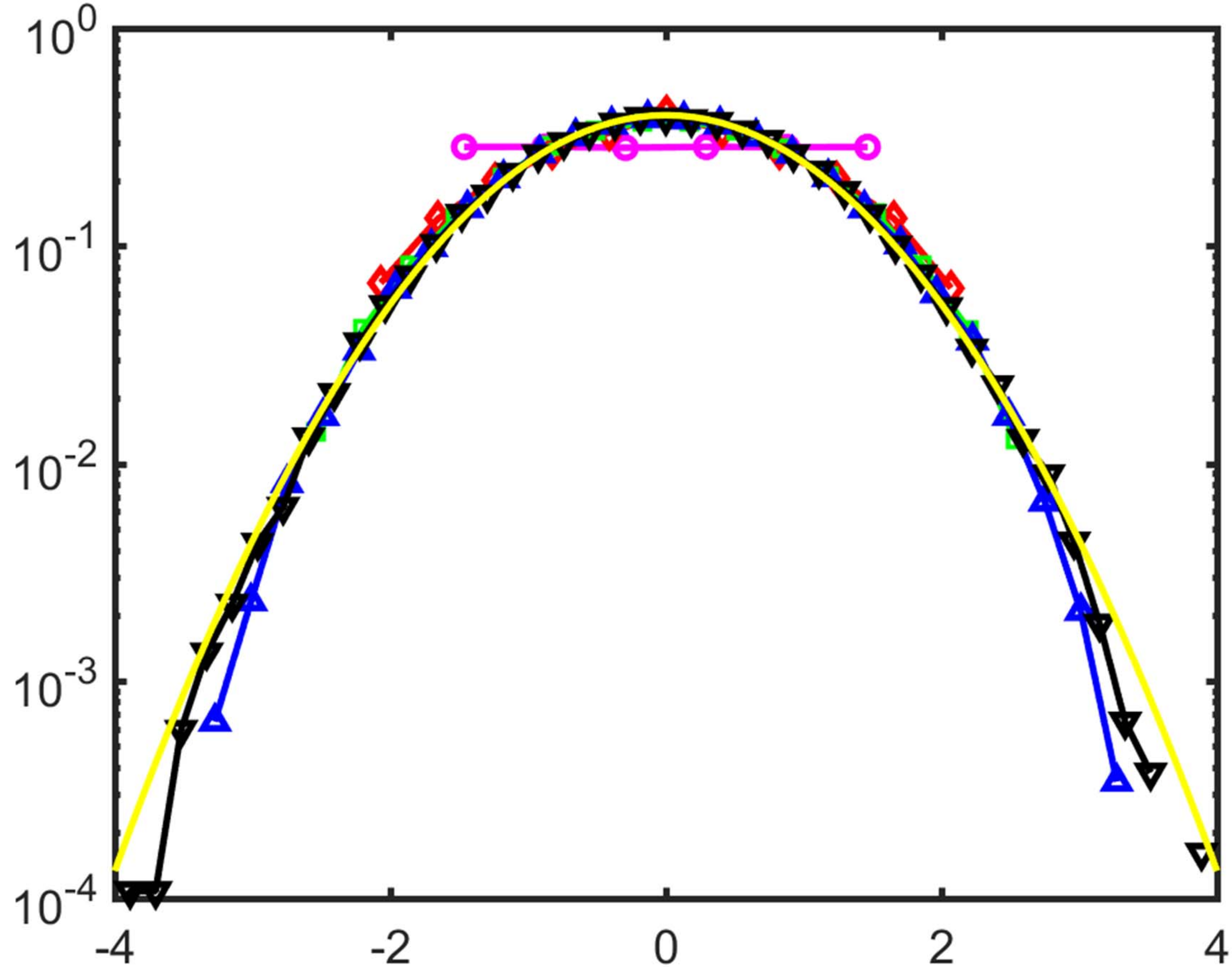
Figure 4-10 Normal probability density functions

Why Gaussian is so important?

- Central Limit Theorem states that a sum (or mean) of a large number of independent random numbers would approximately follow Gaussian distribution
- Let’s check it using a dice experiment in Matlab
- The outcome of each dice throw is a uniformly distributed random variable with average value $\mu=(1+6)/2=3.5$
- How about the mean of 2, 3, 5, 10 throws?

- Stats=100000; N=10; r_table=floor(6.*rand(Stats,N))+1;
- %%
- r1=r_table(:,1);
- step=1; [a,b1]=hist(r1,1:step:6); pdf_r1=a./sum(a)./step;
- figure; hold on; subplot(1,2,1); plot(b1,pdf_r1,'mo-'); hold on; axis([0 7 0 0.2]);
- subplot(1,2,2); semilogy(b1,pdf_r1,'mo-'); hold on; axis([0 7 1e-3 1]);
- %%
- r2=(r_table(:,1)+r_table(:,2))./2;
- step=0.5; [a,b2]=hist(r2,1:step:6); pdf_r2=a./sum(a)./step;
- subplot(1,2,1); plot(b2,pdf_r2,'rd-'); axis([0 7 0 0.4]);
- subplot(1,2,2); semilogy(b2,pdf_r2,'rd-');
- %%
- r5=sum(r_table(:,1:5),2)./5;
- step=1./5; [a,b5]=hist(r5,1:step:6); pdf_r5=a./sum(a)./step;
- subplot(1,2,1); plot(b5,pdf_r5,'b^-'); axis([0 7 0 0.6])
- subplot(1,2,2); semilogy(b5,pdf_r5,'b^-'); axis([0 7 1e-4 1]);
- %%
- r10=sum(r_table(:,1:10),2)./10;
  step=1./10; [a,b10]=hist(r10,1:step:6); pdf_r10=a./sum(a)./step;
- subplot(1,2,1); plot(b10,pdf_r10, 'kv-'); axis([0 7 0 0.8])
- legend(num2str([1,2,5,10]'));
- subplot(1,2,2); semilogy(b10,pdf_r10, 'kv-'); legend(num2str([1,2,5,10]'));


0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
1
2
3
4
5
6
7
1
2
3
5
10
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
0
1
2
3
4
5
6
7
1
2
3
5
10



10^{0}
10^{-1}
10^{-2}
10^{-3}
10^{-4}
-4
-2
0
2
4

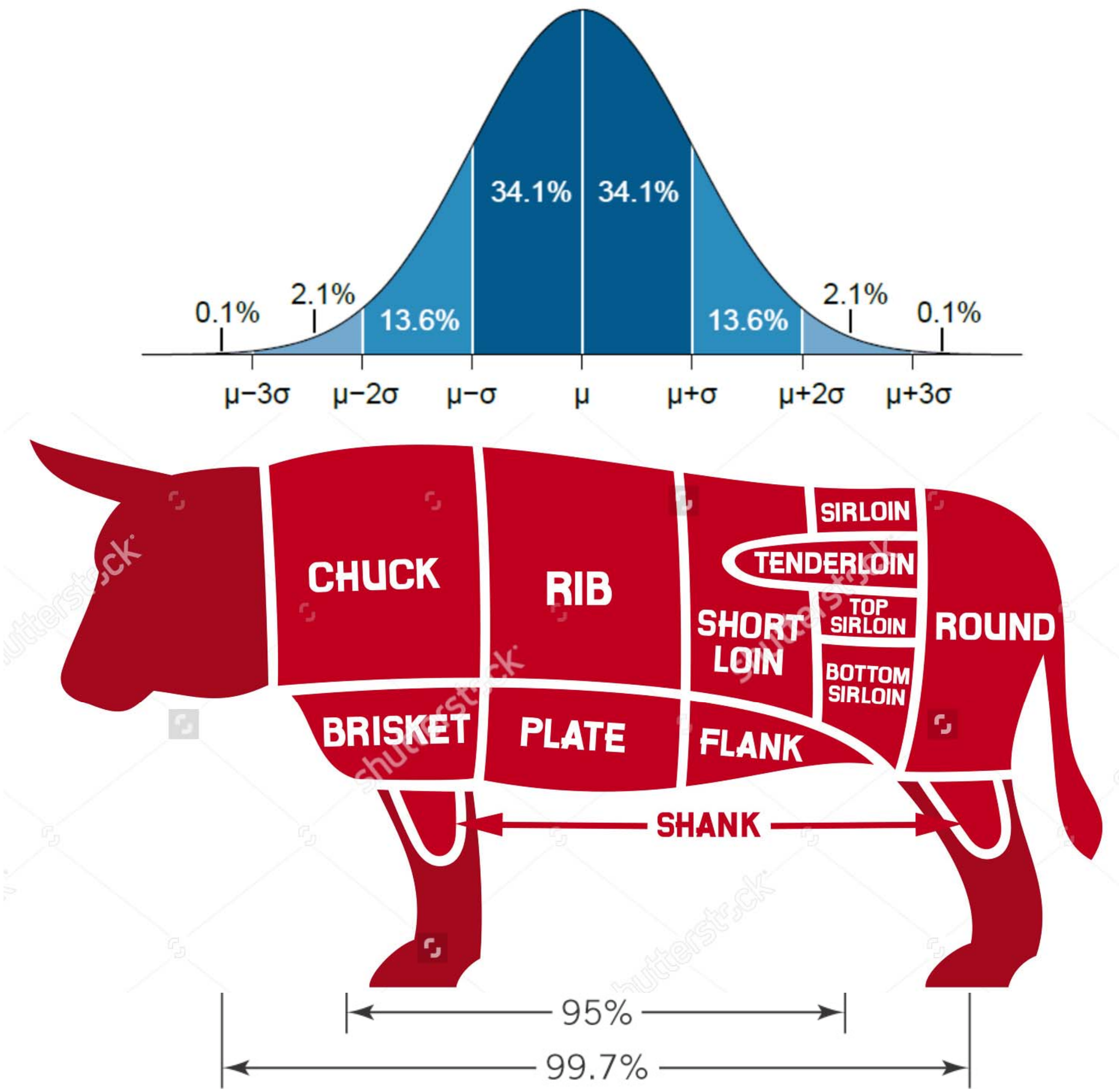
Standard Normal Distribution

- A normal (Gaussian) random variable with

  $$\mu = 0 \text{ and } \sigma^2 = 1$$

  is called a standard normal random variable and is denoted as Z.
- Thed cumulative distribution function of a standard normal random variable is denoted as:

  $$\Phi(z) = P(Z \leq z)$$
- Values are found in Appendix A Table III to Montgomery and Runger textbook


0.1%
2.1%
13.6%
34.1%
34.1%
13.6%
2.1%
0.1%
μ−3σ
μ−2σ
μ−σ
μ
μ+σ
μ+2σ
μ+3σ
CHUCK
RIB
SHORT LOIN
SIRLOIN
TENDERLOIN
TOP SIRLOIN
BOTTOM SIRLOIN
ROUND
BRISKET
PLATE
FLANK
SHANK
95%
99.7%

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

Standardizing

If X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, the random variable

$$Z = \frac{X - \mu}{\sigma} \qquad (4\text{-}10)$$

is a normal random variable with $E(Z) = 0$ and $V(Z) = 1$. That is, Z is a standard normal random variable.

Suppose X is a normal random variable with mean μ and variance σ^2.

Then, $P(X \le x) = P\left(\frac{X - \mu}{\sigma} \le \frac{x - \mu}{\sigma}\right) = P(Z \le z)$ (4-11)

where Z is a standard normal random variable, and

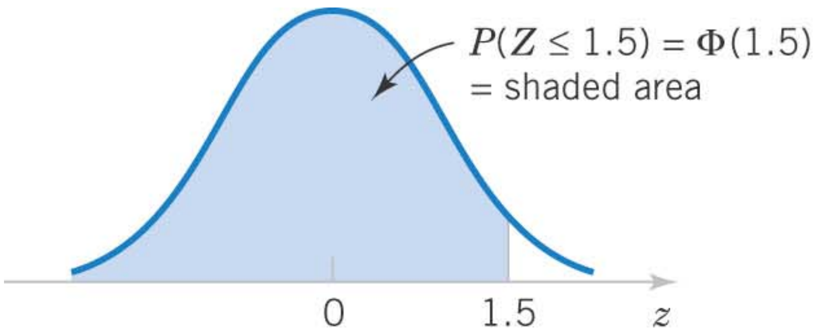
$z = \frac{(x - \mu)}{\sigma}$ is the z-value obtainedby standardizing x.

The probability is obtained by using Appendix Table III

Standard Normal Distribution Tables

Assume *Z* is a standard normal random variable.

Find $P(Z \leq 1.50)$. Answer: 0.93319



z	0.00	0.01	0.02	0.03
0	0.50000	0.50399	0.50398	0.51197
⋮		⋮		
1.5	0.93319	0.93448	0.93574	0.93699

Figure 4-13 Standard normal PDF

Table III from, Appendix A in Montgomery & Runger

Find $P(Z \leq 1.53)$. Answer: 0.93699

Find $P(Z \leq 0.02)$. Answer: 0.50398

Example 4-12: Standard Normal Exercises

1. P(Z > 1.26) = 0.1038
2. P(Z < -0.86) = 0.195
3. P(Z > -1.37) = 0.915
4. P(-1.25 < Z < 0.37) = 0.5387
5. P(Z ≤ -4.6) ≈ 0
6. Find z for P(Z ≤ z) = 0.05, z = -1.65
7. Find z for (-z < Z < z) = 0.99, z = 2.58

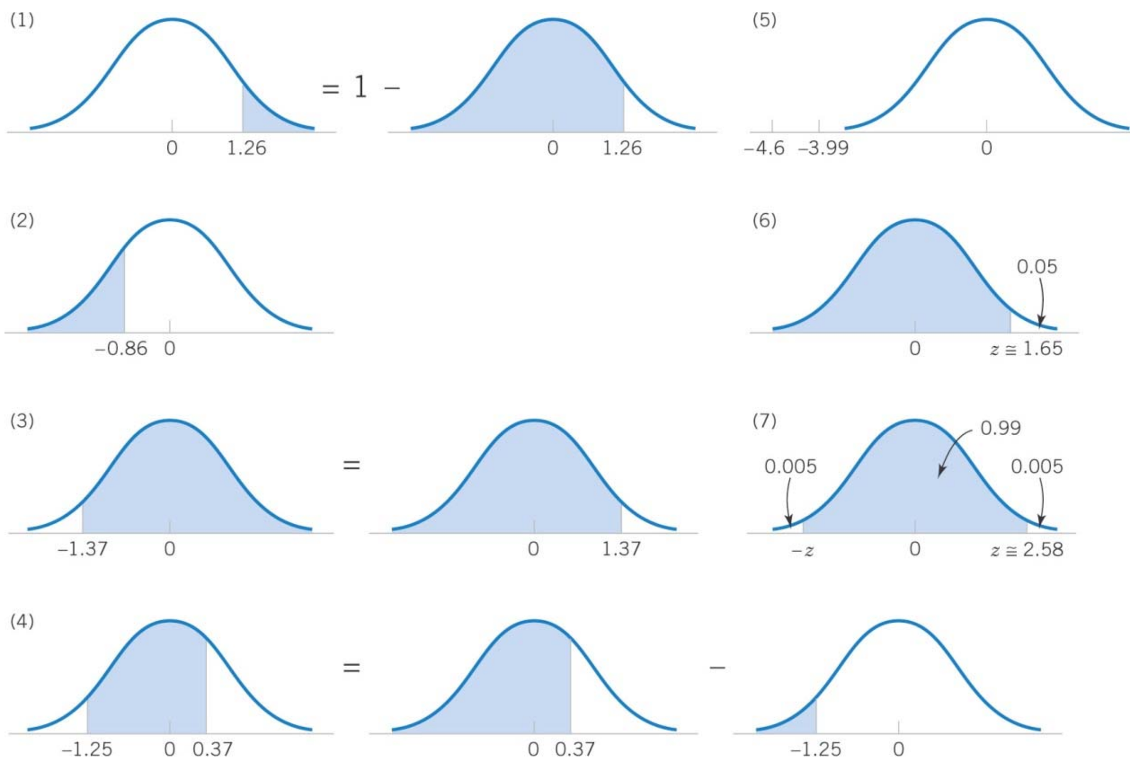


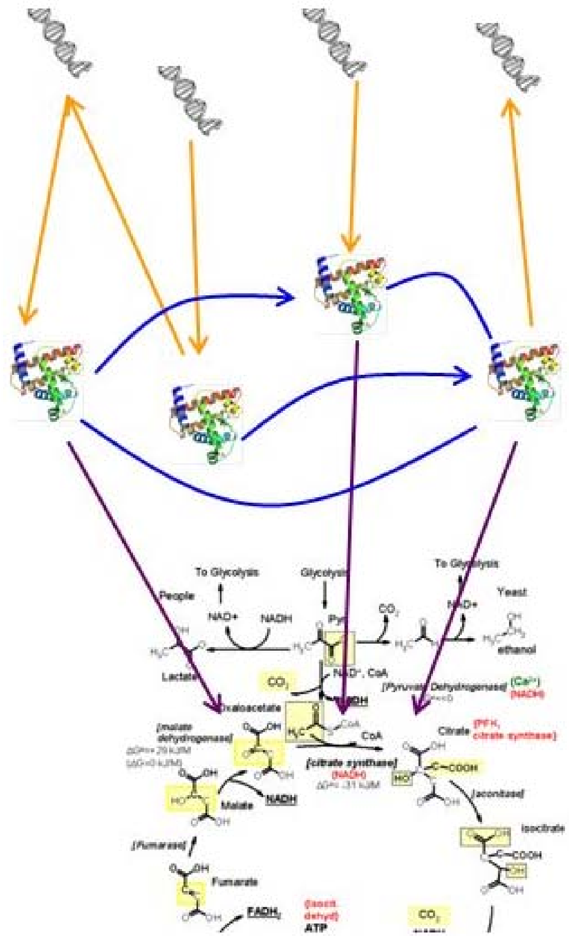
Figure 4-14 Graphical displays for standard normal distributions.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.532922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555760	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705401	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823815	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1.0	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914657	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971283	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2.0	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995604	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3.0	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999065	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999931	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967

Gaussian in systems biology (example)

Molecular binding is used at multiple levels

Networks use binding for specific interactions



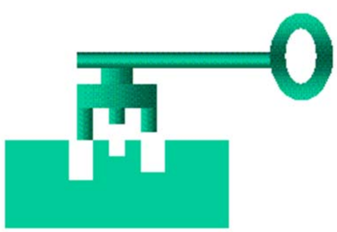
Regulatory network:
RNA-level regulation
By DNA-binding Proteins

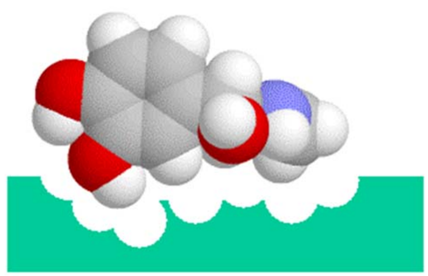
Protein-Protein
(binding)
Interaction Network

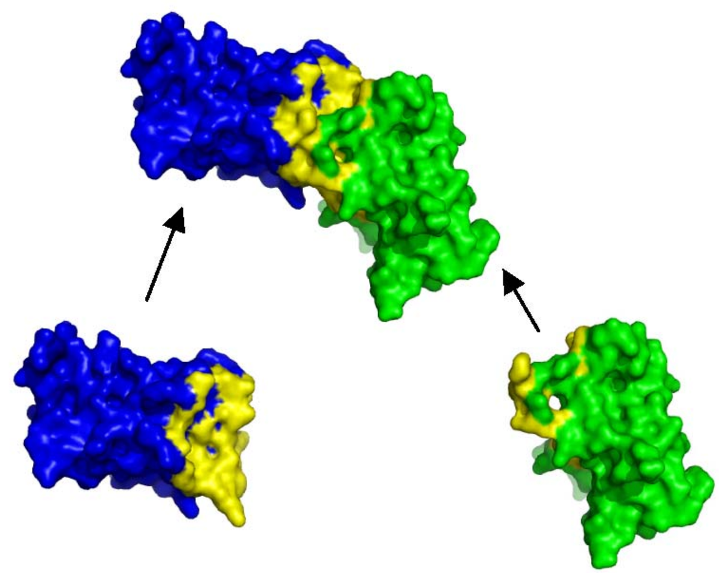
Protein-Metabolite Interactions:
Metabolic network

Binding Energy of Protein-Protein Interactions

- Proteins and other biomolecules (metabolites, drugs, DNA) specifically (and non-specifically) bind each other
- For specific bindings: Lock-and-Key theory
- For non-specific bindings: random contacts







A simple physical model for scaling in protein–protein interaction networks

Eric J. Deeds*, Orr Ashenberg†, and Eugene I. Shakhnovich‡§

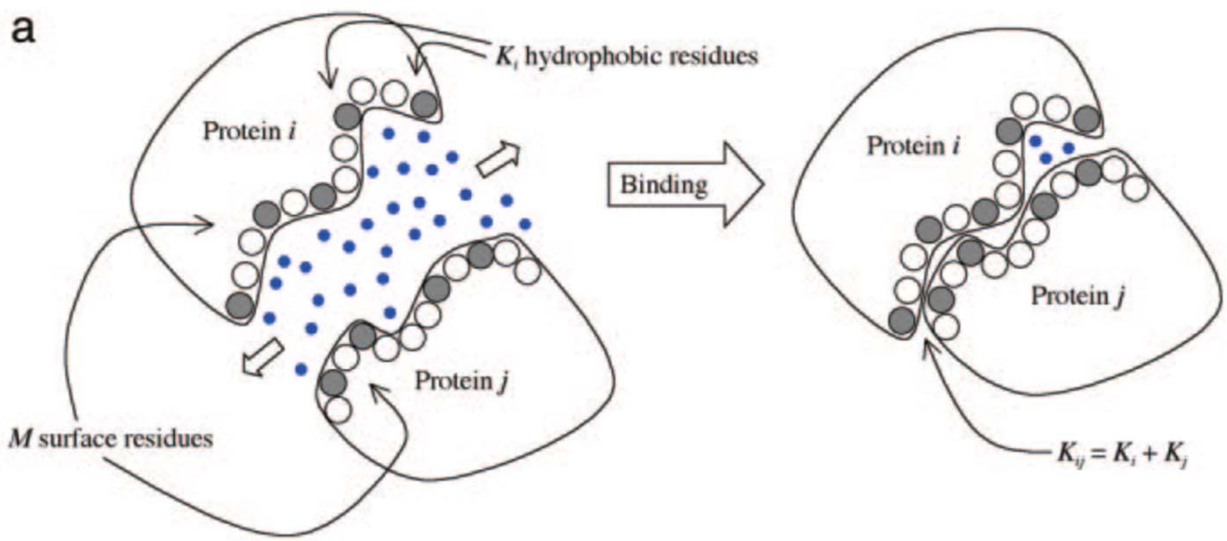
*Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, MA 02138; †Harvard College, 12 Oxford Street, Cambridge, MA 02138; and ‡Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138

Communicated by David Chandler, University of California, Berkeley, CA, November 10, 2005 (received for review September 23, 2005)

It has recently been demonstrated that many biological networks exhibit a "scale-free" topology, for which the probability of observing a node with a certain number of edges (*k*) follows a power law: i.e., $p(k) \sim k^{-\gamma}$. This observation has been reproduced by

(19–22). Indeed, when the two major *S. cerevisiae* p protein interaction (PPI) experiments are compared w another, one finds that only ≈150 of the thousands of tions identified in each experiment are recovered in th

Most Binding energy is due to hydrophobic amino-acid residues being screened from water



Predicted Gaussian distribution: PDF(E_{ij}=E)– because E_{ij} – sum of hydrophobicities of many independent residues

Matlab exercise

- Use Matlab to open file PINT_binding_energy.mat

with binding energy E_{ij} (in units of kT at room temperature) for 430 pairs of interacting proteins from human, yeast, etc.

- Data collected in 2007 from the PINT database http://www.bioinfodatabase.com/pint/
  and analyzed in J. Zhang, **S. Maslov**, E. Shakhnovich, Molecular Systems Biology (2008)

- Fit Gaussian to the distribution of E_{ij} using dfittool
  (read Help page on how to use)
  Find the right threshold to exclude outliers

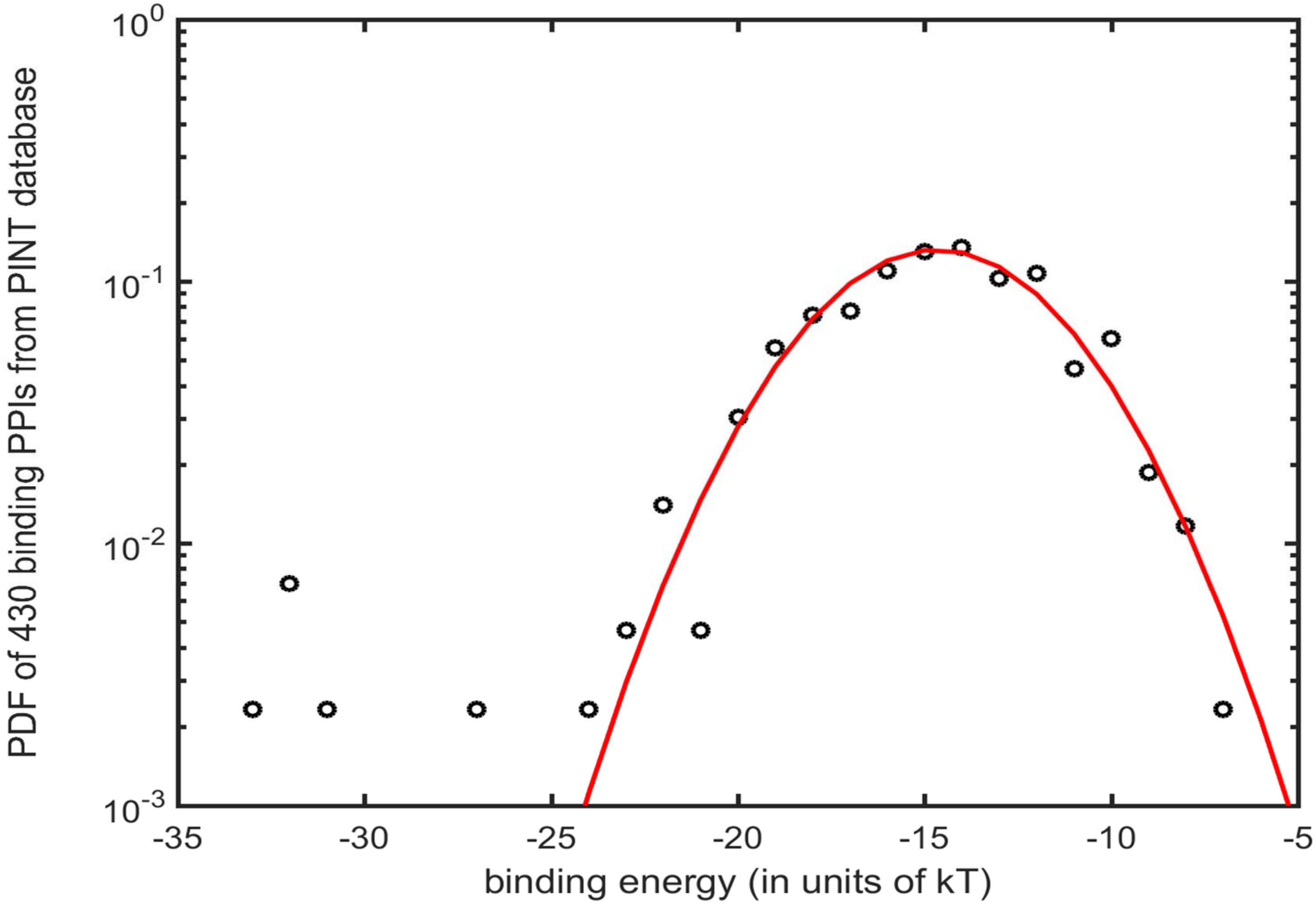
How does it compare with the experimental data? Let’s do Matlab exercise

- **load PINT_binding_energy.mat;**
- **dfittool(binding_energy);**
- **% use “exclude” to exclude X<-23 from the fit**
- **mu=-14.6685**; **sigma=3.01762;**
- **[a,b]=hist(binding_energy,-34:1:0)**
- **figure; semilogy(b,a./sum(a),'ko')**
- **hold on;**
- **plot(b,1/sqrt(2*pi)/sigma.*exp(-(b-mu).^2/2/sigma.^2), 'r-');**

J. Zhang, **S. Maslov**, E. Shakhnovich,
Nature/EMBO Molecular Systems Biology (2008)

Data on binding interactions
from PINT database

How does it compare with the experimental data ?



J. Zhang, **S. Maslov**, E. Shakhnovich,
Nature/EMBO Molecular Systems Biology (2008)

Data on binding interactions
from PINT database

Dissociation constant

- Interaction between two molecules (say, proteins) is usually described in terms of dissociation constant
  K_{ij}=1M $\exp(-E_{ij}/kT)$
- Law of Mass Action: the concentration D_{ij} of a dimer formed out of two proteins with free (monomer) concentrations C_i and C_j : $D_{ij}=C_iC_j/K_{ij}$
- What is the distribution of K_{ij}?
- Answer: it is called log-normal since the logarithm of K_{ij} is the binding energy $-E_{ij}/kT$ which is normally distributed

Lognormal Graphs

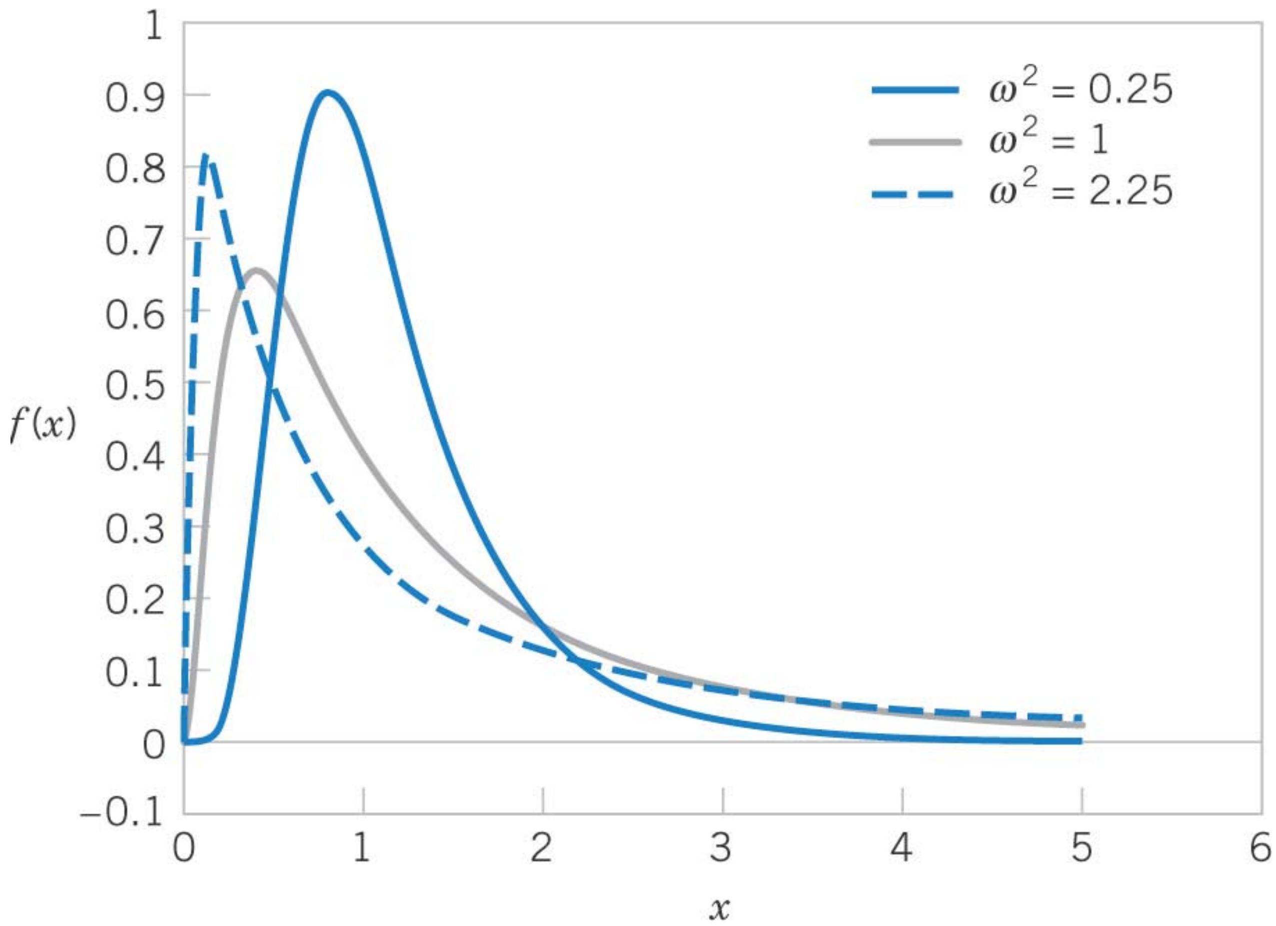
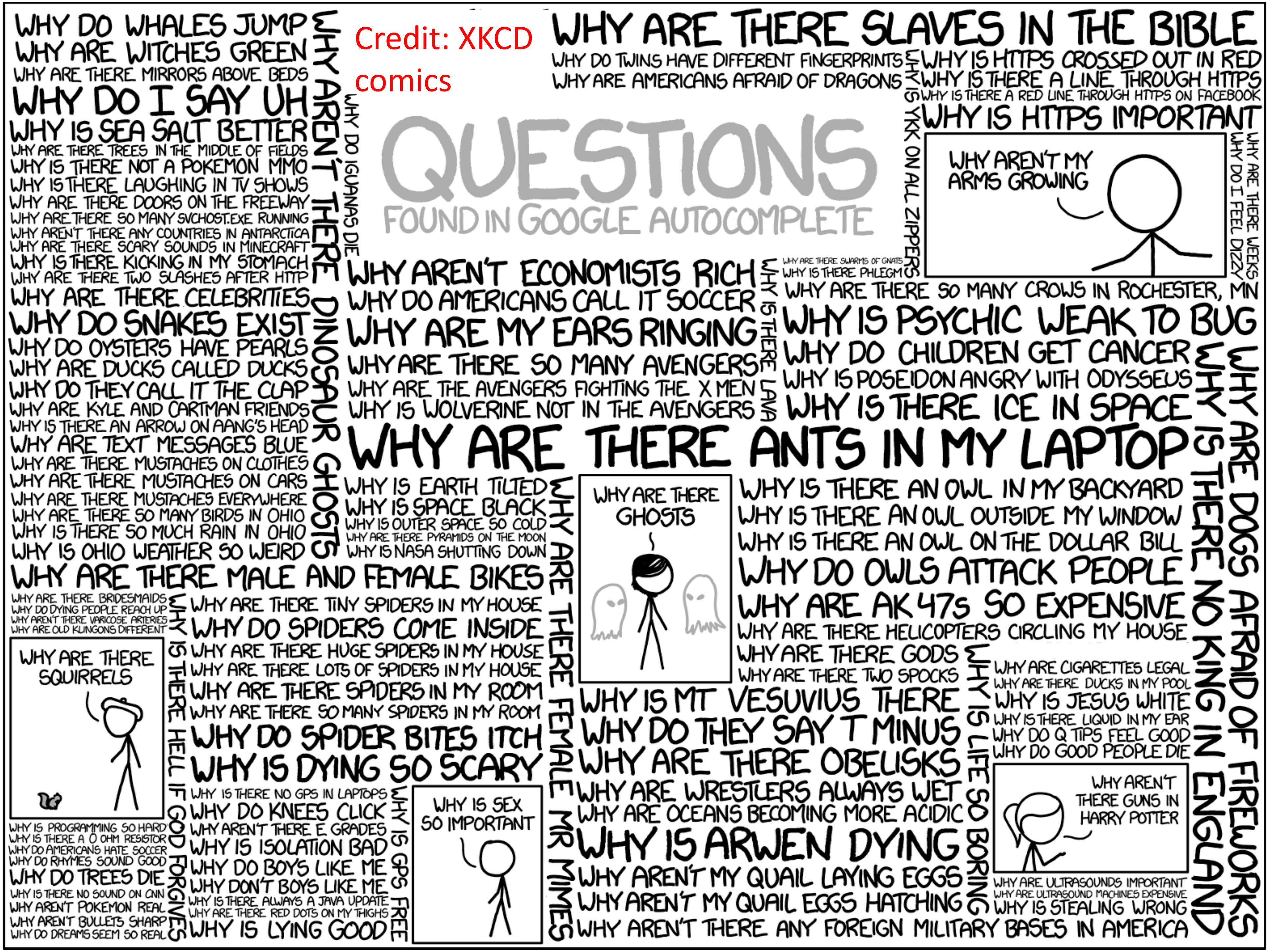


Figure 4-27 Lognormal probability density functions with $\theta = 0$ for selected values of ω^2.



Credit: XKCD comics