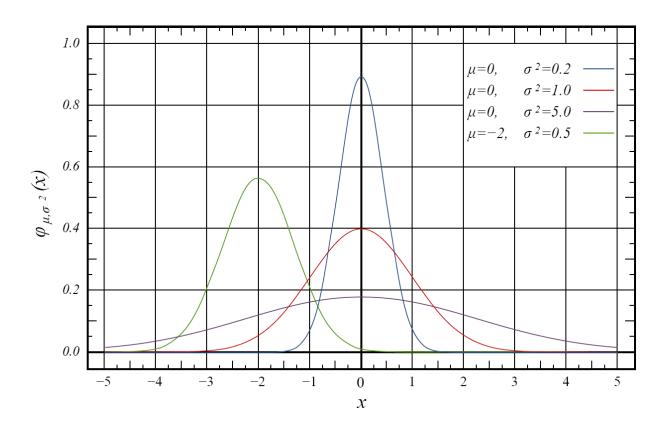
The Normal Distribution

- It's probably the most important distribution in the world!
- Many numerical populations have distributions that can be fit very closely by an appropriate normal curve. (people's height/weight; testing scores; etc.) Even when the underlying distribution is discrete, (yearly number of customers to Wal-Mart; etc.) the normal curve often gives an excellent approximation.
- A continuous rv is said to have a normal (Gaussian) distribution with parameters μ and σ , where $-\infty < \mu < \infty$, and $0 < \sigma$, if the pdf of X is

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} - \infty < x < \infty$$

The Normal pdf

• Normal distribution is a bell-shaped, single peaked and symmetric distribution.



Parameters

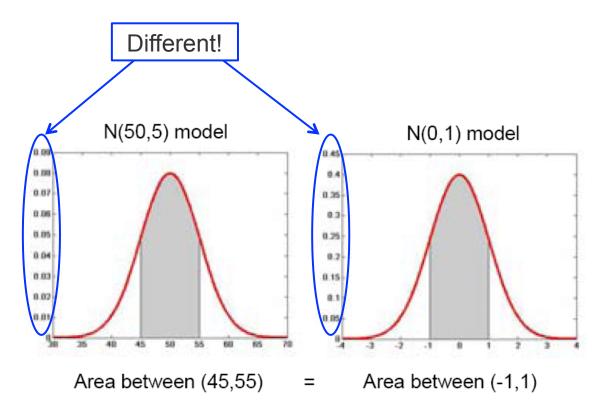
• Clearly $f(x; \mu, \sigma) \ge 0$, but a somewhat complicated calculus argument must be used to verify that

$$\int_{-\infty}^{\infty} f(x; \mu, \sigma) dx = 1.$$

- Parameter μ , stands for the expected value of the normal distribution. Exercise: show that if $X \sim N(\mu, \sigma^2)$, then $E(X) = \mu$.
- Parameter σ , stands for the standard deviation of the normal distribution. Exercise: show that if $X \sim N(\mu, \sigma^2)$, then $Var(X) = \sigma^2$.

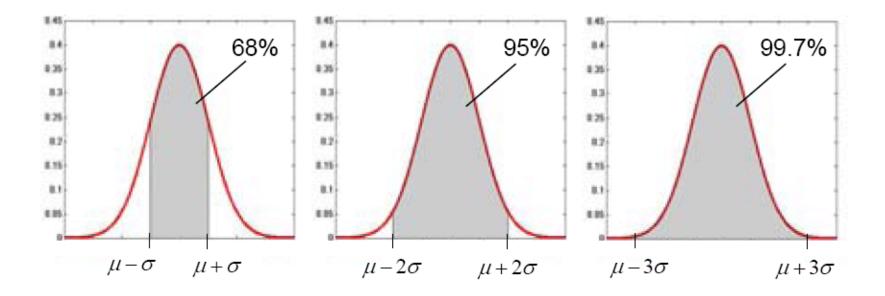
Basic Properties

 All normal models have the same shape and the same area within x standard deviations of its mean.



The 68-95-99.7 Rule

For any normal distribution, we have the following result:

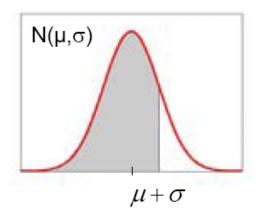


Example

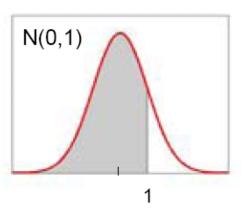
Ex. On an exam the scores followed an approximate normal model with μ = 72 and σ = 8.

- 68% of the students scored between 72±8 or (64, 80).
- 95% of the scores were between 72±2*8 or (56, 88).
- 99.7% of the scores were between 72±3*8 or (48, 96).
- What proportion scored below 84?

Key Result



$$area\{y < \mu + \sigma\}$$



$$area\{z < 1\}$$

Standard Normal

- If $Z \sim N(0, 1)$, i.e., if Z is a normal random variable with μ =0, σ =1. Then Z is said to have a standard normal distribution.
- Any normally distributed rv's could be obtained by using standard normal rv's. To put it more mathematically, if $X \sim N(\mu, \sigma^2)$, then X could be written as

$$X = \mu + \sigma \cdot Z$$

where Z is a standard normal rv.

• Conversely, if $X \sim N(\mu, \sigma^2)$, then

$$Z = (X - \mu) / \sigma$$

has a standard normal distribution. And Z is often called the "z-score" of X.

Example cont.

Ex. The exam scores followed a N(72,8) model.

What proportion of the students scored below 84?

$$z = \frac{y - \mu}{\sigma} = \frac{84 - 72}{8} = 1.5$$

Answer: 93.32%

1	.00	.81	.02	,63	.04	.05	.56	.07	.04	.09
0.0	.5000	.3040	.5080	.5120	3160	5199	5239	3279	.3319	.335
0.1	.5398	.5438	.5428	5817	.5557	5596	5636	5675	.5714	375
0.2	.5793	.5832	501	.5910	.5948	.5967	.6026	.6064	.A103	.614
0.3	.6179	A217	.6255	8293	.6331	4368	.6406	.6443	.6490	991
0.4	.6554	4591	.6628	6664	.6700	.6736	.6772	.6908	.6544	.687
0.5	LARLS S	8950	26985	2019	.7054	.7088	3123	7137	7196	.722
用在:	7257	.3291	.7324	2357	.7389	7422	3454	,7486	2517	.754
0.7	7580	Jan	.7642	2(7)	.7794	7734	3764	.7794	3823	.785
O.A.	7881	2910	,2999	2967	2995	.0023	.9051	.9079	.A106	303
0.9	.8159	3006	3212	3218	.8264	1299	AH5	,5340	.8965	Als
1.0	.6413	.8438	.5465	.1485	.8506	.8531	.8554	.6577	A599	.862
I.I.	.8643	.8665	3686	.8708	.8729	3749	3770	.8790	.8810	383
1.2	.8949	3869	.3585	.8907	3925	.8944	.8962	.5980	.8997	.900
13	.9032	.9049	.9066	.9082	,9099	3115	8934	.9147	.5162	.917
1.4	9197	,9297	.9222	.9236	.9251	.9245	.9279	.9292	3906	.931
15:	.9532	.9945	.9357	9970	.9382	3194	.9406	.9418	.9429	.944
15	17976	,9463	.9474	3484	.946	.9505	3515	.9525	.4515	354
1.7	.9534	,9564	3573	3582	,9591	3599	.9608	.9636	.5625	.963
1.8	.9641	3649	39656	5664	.9471	.96TH	.968b	.9649	56/99	.970
1.9	.9713	9719	.972h	9752	3038	.9744	9750	.9756	3761	.90%
2.0	.8172	.9778	.9793	3788	.//293	.5799	.9903	,1605	.5812	- 581
2.1	.9821	.9626	.9636	3834	.9838	.9842	.9946	.9939	.9654	.985
2.2	.5941	.9664	.9668	9871	.9675	3878	9881	3684	5667	,949
23	.997	.9896	.9656	.9901	.9904	.9906	.9909	.9911	.9913	3990
2.4	.9918	.9920	.9922	.9925	/9927	.9929	.9931	.9932	.9934	399
25	.9938	.9940	.9941	3941	.9945	.9946	.9948	.9949	.9951	.995
24	.9953	.965	.9956	.9957	.9959	.5965	.9961	.9962	.9963	,996
27	3965	.9966	.5947	.9968	.9969	.9979	.9971	.9977	9973	.997
2.8	.9974	,1975	.9976	9977	,9977	.9978	.9979	.1979	,9990	,998
2.9	.9981	,9982	.9962	9981	,9984	.9984	.9985	.9965	,9986	,998
1.0	.9967	.9987	.9987	.9988	,9968	.9989	,9989	.9999	,9990	.999
3,1	.9990	,9991	3996	9991	.9992	.9997	.9992	,9992	.9993	,999
32	.9993	.9993	.9994	9994	.9994	.9994	,9994	.5995	.996	.999
33	.9995	.3995	3995	3996	.3996	.39%	.999b	,9996	.59%	.999
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.999

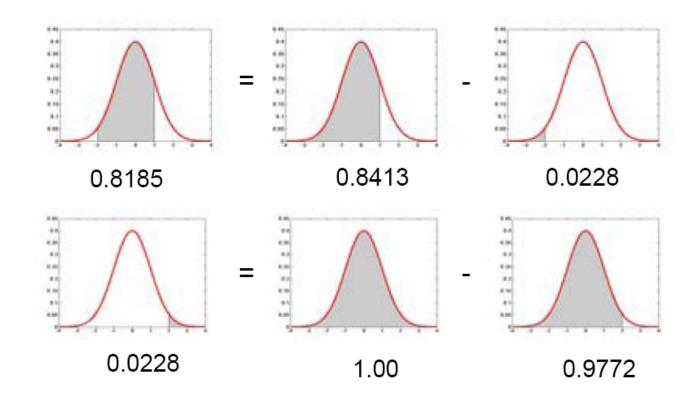
Simplification

• Thus, any problem about any normal rv X ~ N(μ , σ^2), can be translated to a problem about a standard normal rv Z.

$$\underline{\mathsf{Ex.}}\ \mathsf{P}(a \le \mathsf{X} \le b) = \mathsf{P}[(a - \mu)/\sigma \le (\mathsf{X} - \mu)/\sigma] = \mathsf{P}[(a - \mu)/\sigma \le \mathsf{Z} \le (b - \mu)/\sigma].$$

- The cumulative distribution function of standard normal distribution, that is $\Phi(z) = P(Z \le z)$, is already known! (Appendix Table.)
- Check Table A.3 to determine $P(Z \le 0.76)$; P(Z > 0.76); $P(-1.32 \le Z \le 0.76)$.
- Question: How to get the p-th percentile of the standard normal from A.3?

Using the Normal Table



R instead of tables

```
R command: dnorm(x, mean = 0, sd = 1),
pnorm(q, mean = 0, sd = 1),
qnorm(p, mean = 0, sd = 1).
```

Example

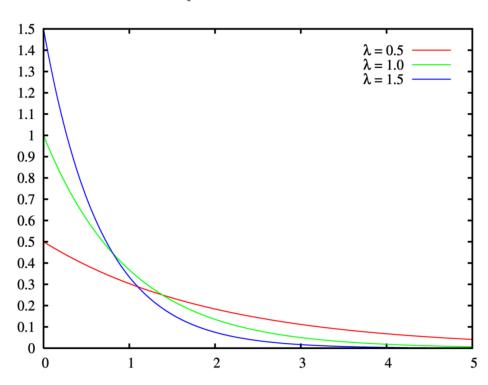
Ex. Suppose the height of all Columbia students can be described by a N(68, 4) model.

- 1. What proportion of students is shorter than 74 inches?
- 2. What proportion of students is taller than 74 inches?
- 3. How tall does a student have to be to be among the 10% tallest students?

The Exponential Distribution

• X is said to have an exponential distribution with parameter λ (λ >0) if the pdf of X is

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0\\ 0 & \text{otherwise} \end{cases}$$



More on Exponential

Note that an exponential rv X can only take positive values. And the cdf of X is

$$F(x;\lambda) = \begin{cases} \int_0^x \lambda e^{-\lambda y} dy = 1 - e^{-\lambda x} & x \ge 0\\ 0 & x < 0 \end{cases}$$

- Thus $P(X>x) = 1 F(x; \lambda) = e^{-\lambda x}$
- Proposition: (proof?)

If X is an exponential rv with parameter
$$\lambda$$
, then $E(X) = 1/\lambda$, $Var(X) = 1/\lambda^2$

R command: dexp(x, lamda=1),
pexp(q, lamda=1),
qexp(p, lamda=1).

Exponential Distribution and Poisson Distribution

Suppose the number of events occurring in a time interval of length t has Poisson Distribution with parameter αt , and the numbers of occurrences in non-overlapping intervals are independent of one another. Then the distribution of elapsed time between the occurrence of two successive events is exponential with parameter $\lambda = \alpha$.

Example

Ex. Suppose you are waiting for a bus at a bus station. And the distribution of the length of the time you have to wait to get on the bus after you arrive at the bus station is exponentially distributed with parameter λ. Assume you have already waited for s minutes, how much longer do you expect to wait?

First, we have to figure out the conditional probability distribution of the additional waiting time given we have waited for s minutes. For any t > 0

$$P(X \ge s + t \mid X \ge s) = P[(X \ge s + t) \cap (X \ge s)] / P(X \ge s)$$
$$= P(X \ge s + t) / P(X \ge s)$$
$$= e^{-\lambda t}$$

which is again an exponential distribution! Thus the expected additional waiting time is $1/\lambda$.

Memoryless Property

- From the previous example, we know that if a waiting time (or lifetime of something) follows an exponential distribution, the distribution of additional waiting time (lifetime) is exactly the same as the original distribution of waiting time (lifetime). In other words, the exponential distributed waiting time does NOT remember how much time you have waited, it starts afresh at any time!
- It is popular to model the distribution of component lifetime using the exponential distribution. However, the memoryless property may not be realistic in may applied problems. More general lifetime models can be furnished by the gamma, Weibull, and lognormal distributions. (Book: p159 – p168).

Joint Distribution

- How can we model two rv's using probability models? For example, if we are interested in both weight and height.
- Is it enough if we just use a normal model for weight and another normal model for height?
- We need to introduce joint probability distribution in order to model multiple rv's.

Joint PMF

- Let X and Y be two discrete rv's defined on the sample space. The joint probability mass function p(x, y) is defined for each pair of numbers (x, y) by p(x, y) = P(X=x, Y=y).
- As in the single rv case, we must have $p(x, y) \ge 0$ and $\sum_{x} \sum_{y} p(x, y) = 1$.

Example

Ex. We randomly put two different balls into 3 numbered (numbered as {1,2,3}) boxes. Let X be the number of empty boxes left; let Y be the minimum of the box number that has balls in it. What is the joint distribution of (X, Y)?

X can take values from {1, 2};

Y can take values from {1, 2, 3};

It's not hard to see we have the following (why?):

$$p(2, j) = P(X=2, Y=j) = 1/9$$
, for $j = 1, 2, 3$.

$$p(1, 3) = P(X=1, Y=3) = 0.$$

$$p(1, 1) = P(X=1, Y=1) = 4/9.$$

$$p(1, 2) = P(X=1, Y=2) = 2/9.$$

p_{ij}	1	2	3
1	4/9	2/9	0
2	1/9	1/9	1/9

Marginal PMF

• The marginal probability mass functions of X and Y, denoted by $p_X(x)$ and $p_Y(y)$, respectively, are given by

$$p_{\mathbf{X}}(x) = \sum_{y} p(x, y) \quad p_{\mathbf{Y}}(y) = \sum_{x} p(x, y)$$

Ex.

 Notice that the marginal probability mass functions are automatically proper pmf's. (why?)