S1211Q Introduction to Statistics
Lecture 9

Wei Wang

July 16, 2012

Poisson Distribution

- Poisson Distribution is for describing outcomes that come in the form of count data, e.g., visits to a particular website during a time interval
- But unlike Binomial or Hypergeometric Distribution, there is no simple experiment that Poisson Distribution is based on.
- A random variable X is said to have Poisson Distribution with parameter $\mu(> 0)$ if the pmf of X is

$$p(x; \mu) = e^{-\mu}\frac{\mu^{x}}{x!}, x = 0, 1, 2, \ldots$$

Poisson Distribution PMF

- Verify the pmf is a valid pmf

$$p(x;\mu) = e^{-\mu}\frac{\mu^{x}}{x!}, x = 0, 1, 2, \ldots$$

Poisson Distribution PMF

- Verify the pmf is a valid pmf

$$p(x;\mu) = e^{-\mu}\frac{\mu^x}{x!}, x = 0, 1, 2, \ldots$$

- Recall from Calculus

$$e^{\mu} = 1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \frac{\mu^4}{4!} + \cdots$$

Poisson Distribution PMF

- Verify the pmf is a valid pmf

$$p(x;\mu) = e^{-\mu}\frac{\mu^x}{x!}, x = 0, 1, 2, \ldots$$

- Recall from Calculus

$$e^{\mu} = 1 + \mu + \frac{\mu^2}{2!} + \frac{\mu^3}{3!} + \frac{\mu^4}{4!} + \cdots$$

- So

$$p(0;\mu) + p(1;\mu) + p(2;\mu) + \cdots = e^{\mu} \times e^{-\mu} = 1$$

Example

- Let X denote the number of creatures of a particular type captured in a trap during a given time period. Suppose that X has a Poisson distribution with =4.5, so on average traps will contain 4.5 creatures. Then the probability that a trap contains exactly five creatures is

$$P(X = 5) = \frac{e^{-4.5}(4.5)^5}{5!} = 0.1708$$

Example

- Let X denote the number of creatures of a particular type captured in a trap during a given time period. Suppose that X has a Poisson distribution with =4.5, so on average traps will contain 4.5 creatures. Then the probability that a trap contains exactly five creatures is

$$P(X=5)=\frac{e^{-4.5}(4.5)^5}{5!}=0.1708$$

- The probability that the a trap has at most five creatures is

$$P(X\leq 5)=\sum_{x=0}^{5}\frac{e^{-4.5}(4.5)^x}{x!}=.7029$$

Poisson Distribution as a Limit

- Suppose that in the binomial pmf $b(x; n; p)$, we let $n \to \infty$ and $p \to 0$ in such a way that np approaches a value $\mu > 0$. Then $b(x; n; p) \to p(x; \mu)$.
- So in any binomial experiment in which n is large and p is small, , then Binomial can be approximated by Poisson Distribution with parameter $\mu = np$.

Example

- A typesetter, on the average makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be no more than two errors in five pages?

Example

- A typesetter, on the average makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be no more than two errors in five pages?
- X is the number of errors in five pages

$$X \sim Bin(1500, 1/500)$$

Example

- A typesetter, on the average makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be no more than two errors in five pages?
- X is the number of errors in five pages

$$X \sim Bin(1500, 1/500)$$

- Exact solution

$$P(X \leq 2) = \sum_{x=0}^{2} \binom{1500}{x} (\frac{1}{500})^{x} (\frac{499}{500})^{1500-x} = .4230$$

Example

- A typesetter, on the average makes one error in every 500 words typeset. A typical page contains 300 words. What is the probability that there will be no more than two errors in five pages?
- X is the number of errors in five pages

$$X \sim Bin(1500, 1/500)$$

- Exact solution

$$P(X \leq 2) = \sum_{x=0}^{2} \binom{1500}{x} (\frac{1}{500})^{x} (\frac{499}{500})^{1500-x} = .4230$$

- With Poisson Approximation $\mu = np = 3$

$$P(X \leq 2) \approx e^{-3} + 3e^{-3} + \frac{3^2 e^{-3}}{2} = .4232$$

Mean and Variance of Poisson Distribution

- If X has a Poisson Distribution with parameter μ, then $E(X) = Var(X) = \mu$.
- It can be derived directly from the pmf, or through the Binomial limit argument.
- If X is $b(x; n; p)$, then

$$E(X) = np \to \mu, Var(X) = np(1-p) \to \mu$$

Continuous Random Variables

Continuous RV

- Recall the definition of pmf for a discrete rv. P(X=*x*). Can we extend this definition to continuous rv's?

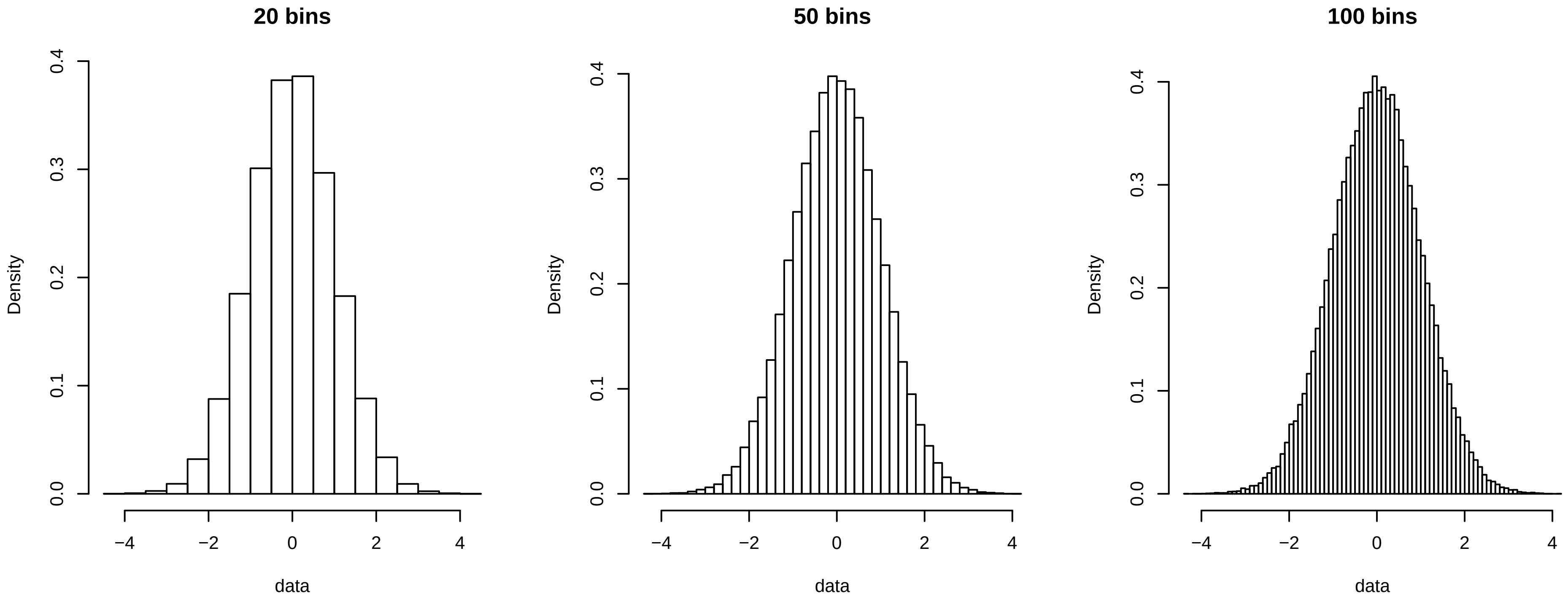
- Uniform random variable: X is equally likely to be any number on [0,1], what is the probability P(X=0.5)?

- The probability model for a continuous random variable assigns probabilities to intervals of outcomes rather than to individual outcomes.

- The probability model of X is often described by a smooth curve, which is the probability density function (pdf) of X.

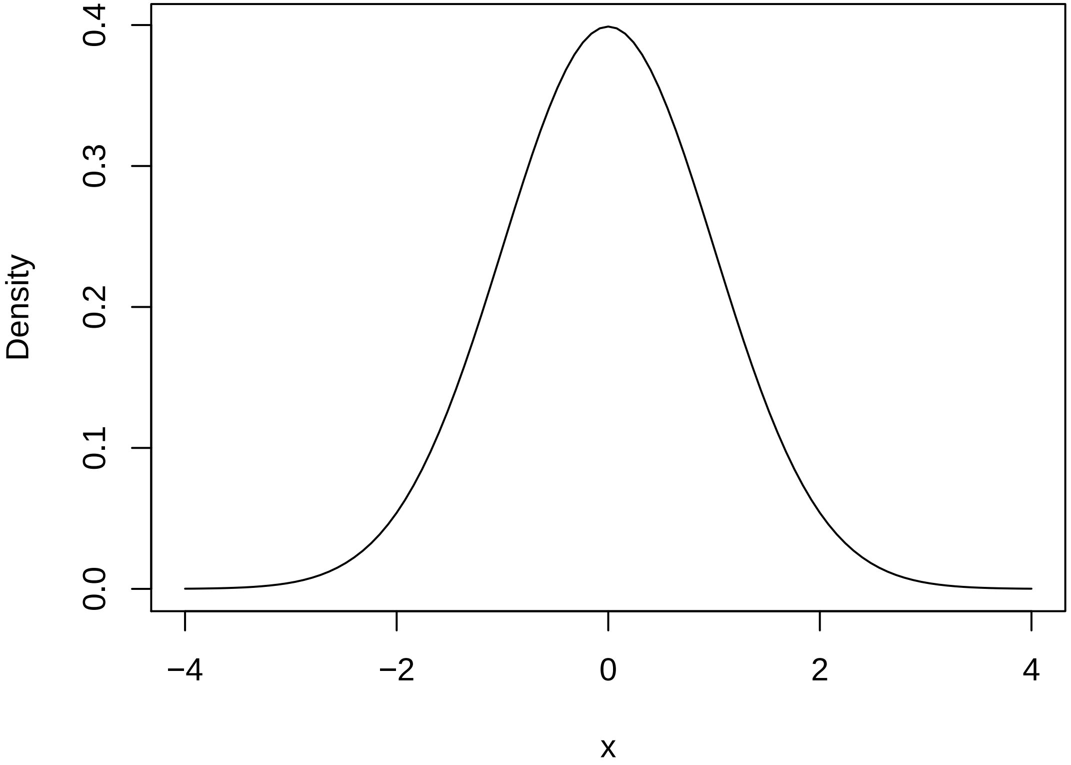
From Histogram to Density

- We have some data of sample size 100,000, if we draw Density Histogram and make the breakpoints finer and finer...



From Histogram to Density

- We will end up having the so-called density curve.



-

PDF

- The probability density function (pdf) of a continuous rv X is a function *f*(*x*) such that for any two numbers *a* and *b* with *a* ≤ *b*,

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

  The graph of *f*(*x*) is often referred to as the *density curve*.

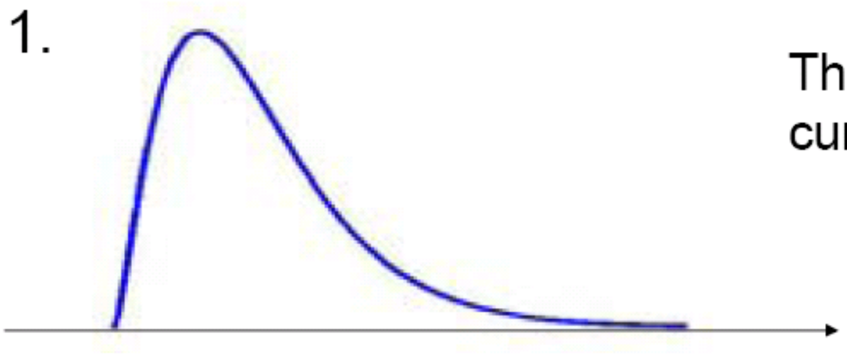
- This means the area under the density curve represents probability!

- Note that 0 ≤ *f*(*x*) for all *x*.

- *f*(*x*)d*x* can be treated as P(X=x)!

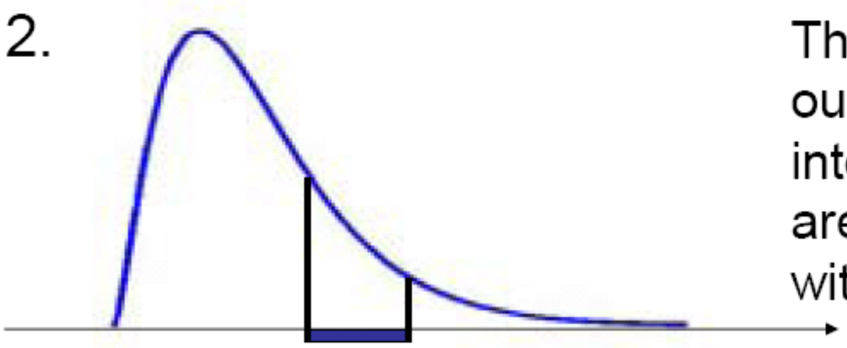
Properties of PDF

1.



The total area under the curve must equal 1.

2.



The probability that the outcome lies in a specific interval is given by the area under the curve within that interval.

Uniform Distribution

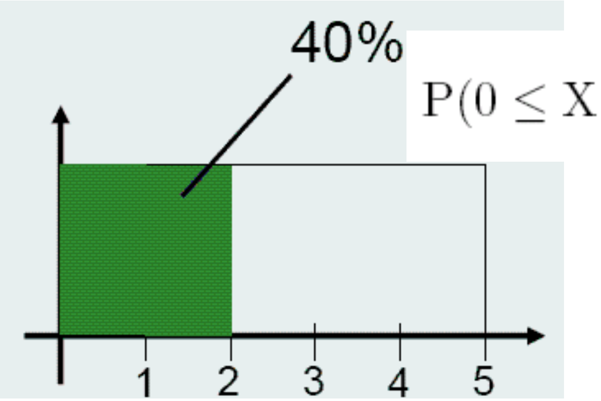
- A continuous rv X is said to have a uniform distribution on the interval [*A*, *B*] if the pdf of X is

$$f(x; A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & \text{otherwise} \end{cases}$$

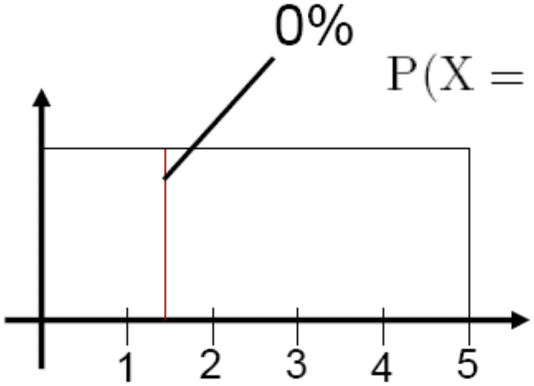
- Verify that this is a proper pdf.
  1. *f*(x) ≥ 0 for all *x*.
  2. Area under *f*(x) should be equal to 1.

Example

Ex. Suppose a bus arrives equally likely at any time between 7:00 – 7:05 AM. What is the probability it arrives sometime between 7:00 – 7:02 AM?



$$P(0 \le X \le 2) = \int_0^2 \frac{1}{5}dx = \frac{2}{5}$$



$$P(X = c) = \lim_{\epsilon \longrightarrow 0} P(c - \epsilon \le X \le c + \epsilon) = \lim_{\epsilon \longrightarrow 0} \int_{c-\epsilon}^{c+\epsilon} \frac{1}{B - A} dx = 0$$

The CDF

- Although the idea of pmd does not extend to the continuous rv's, the idea of cdf still works.

- The cumulative distribution function (cdf) *F*(*x*) for a continuous rv X is defined for every number *x* by

$$F(x) = \mathrm{P}(\mathrm{X} \leq x) = \int_{-\infty}^{x} f(y)dy$$

- *F*(*x*) is in fact the probability that a rv X is smaller than *x*. *F*(*x*) increases smoothly as *x* increases. *F*(-∞) = 0, *F*(+∞) = 1.

- It is easy to compute probabilities using *F*(*x*).
  - P(X > *a*) = 1 - *F*(*a*)
  - P(*a* ≤ X ≤ *b*) = *F*(*b*) - *F*(*a*)

pdf from cdf

- If X is a continuous rv with pdf *f*(*x*) and cdf *F*(*x*), then at every *x* at which the derivative *F*'(*x*) exists, *F*'(*x*) = *f*(*x*). *f*(*x*) is often a smooth curve, which is the probability density function (pdf) of X.

- Let *p* be a number between 0 and 1. The (100*p*)th percentile (quantile) of the distribution of a continuous rv X, denoted by *η*(*p*), is defined by

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)dy$$

- The median of a continuous distribution, denoted by $\tilde{\mu}$, is the 50th percentile, so $\tilde{\mu}$ satisfies .5 = F($\tilde{\mu}$). That is, half the area under the density curve is to the left of $\tilde{\mu}$ and half is to the right of $\tilde{\mu}$.

Expected Values

- Notice that the pdf *f*(*x*) of a continuous distribution is actually playing the role of pmf *p*(*x*) of a discrete distribution.

- Recall that the expected value of a discrete distribution is calculated by

$$\mu_X = \mathrm{E}(\mathrm{X}) = \sum_{x \in D} x \cdot p(x)$$

- Therefore, similarly we can define the expected value of a continuous distribution by

$$\mu_X = \mathrm{E}(\mathrm{X}) = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

- Take advantage of the *symmetry* of particular distributions, when calculating expectations.

Variance

- With a similar argument as in the discrete case, we can also define the expectation of a function of a continuous rv as well as the variance of a continuous rv.

- Proposition: if X is a continuous rv with pdf *f*(*x*) and *h*(X) is any function of X, then

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

- As a special case of the above proposition, the variance of X is defined by

$$\sigma_{\mathrm{X}}^2 = \mathrm{Var}(\mathrm{X}) = \mathrm{E}(\mathrm{X} - \mathrm{E}(X))^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f(x)dx$$

The standard deviation (SD) of X is $\sigma_X = \sqrt{\mathrm{Var}(\mathrm{X})}$.

Examples

Ex. Prove for continuous rv X, as in the discrete case, that $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$.

Ex. If a stick of length 1 is broken at random into two pieces. What is the expected length of the longer piece?

Properties

- Some properties of mean and variance hold in the continuous case in a similar way as in the discrete case.

- For example, under linear transformation of X, we have

1. $E(aX+b) = aE(X) + b$
2. $Var(aX+b) = a^2Var(X)$

- Exercise: prove the above formulas rigorously!

Uniform RV

- We call a uniform rv U a standard uniform, if and only if U ~ uniform on [0,1]

- For a standard uniform rv U, we can easily calculate,

$$\mathrm{E}(\mathrm{U}) = \int_0^1 x \cdot 1 dx = \frac{1}{2}$$

$$\mathrm{E}(\mathrm{U}^2) = \int_0^1 x^2 \cdot 1 dx = \frac{1}{3}$$

$$\mathrm{Var}(\mathrm{U}) = \mathrm{E}(\mathrm{U}^2) - [\mathrm{E}(\mathrm{U})]^2 = \frac{1}{12}$$

General Uniform

- Note that a general case of uniform distribution X on [*A, B*] can be treated as a linear transform of a standard uniform, i.e., X = (*B* – *A*)U + *A*.

- Proposition:

> If X is a continuous uniform rv on [*A, B*], then
> E(X) = (*B* + *A*)/2, Var(X) = (*B* - *A*)^2/12

- R command:

```
dunif(x, min=0, max=1),
punif(q, min=0, max=1),
qunif(p, min=0, max=1).
```

The Normal Distribution

- It’s probably the most important distribution in the world!

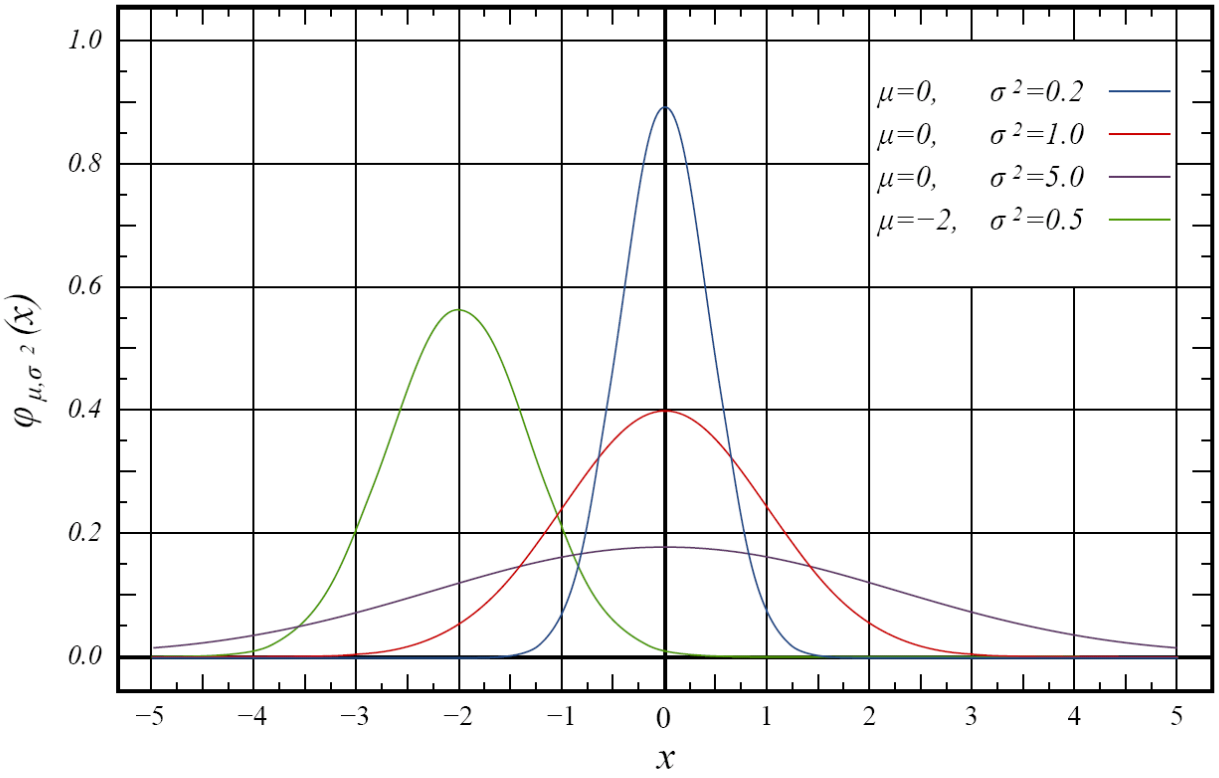
- Many numerical populations have distributions that can be fit very closely by an appropriate normal curve. (people’s height/weight; testing scores; etc.) Even when the underlying distribution is discrete, (yearly number of customers to Wal-Mart; etc.) the normal curve often gives an excellent approximation.

- 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

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

The Normal pdf

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



Parameters

- Clearly $f(x; \mu, \sigma) \geq 0$, 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.

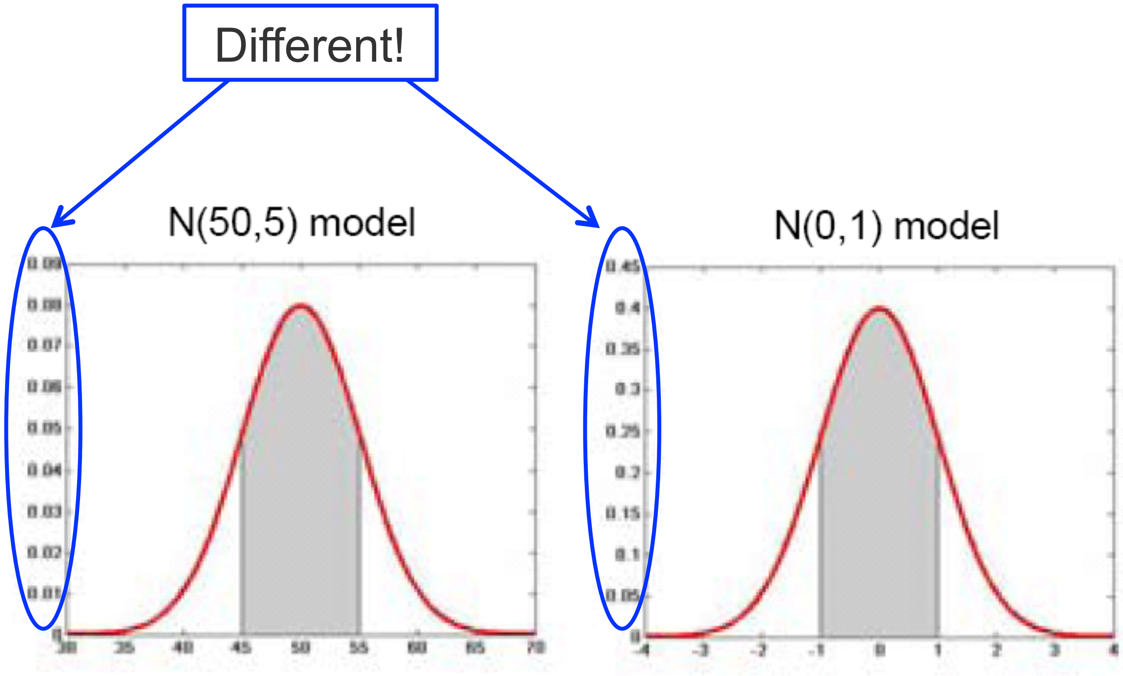
  Exercise: show that if X ~ N(μ, σ^2), then E(X) = μ.

- Parameter σ, stands for the standard deviation of the normal distribution.

  Exercise: show that if X ~ N(μ, σ^2), then Var(X) = σ^2.

Basic Properties

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



Different!

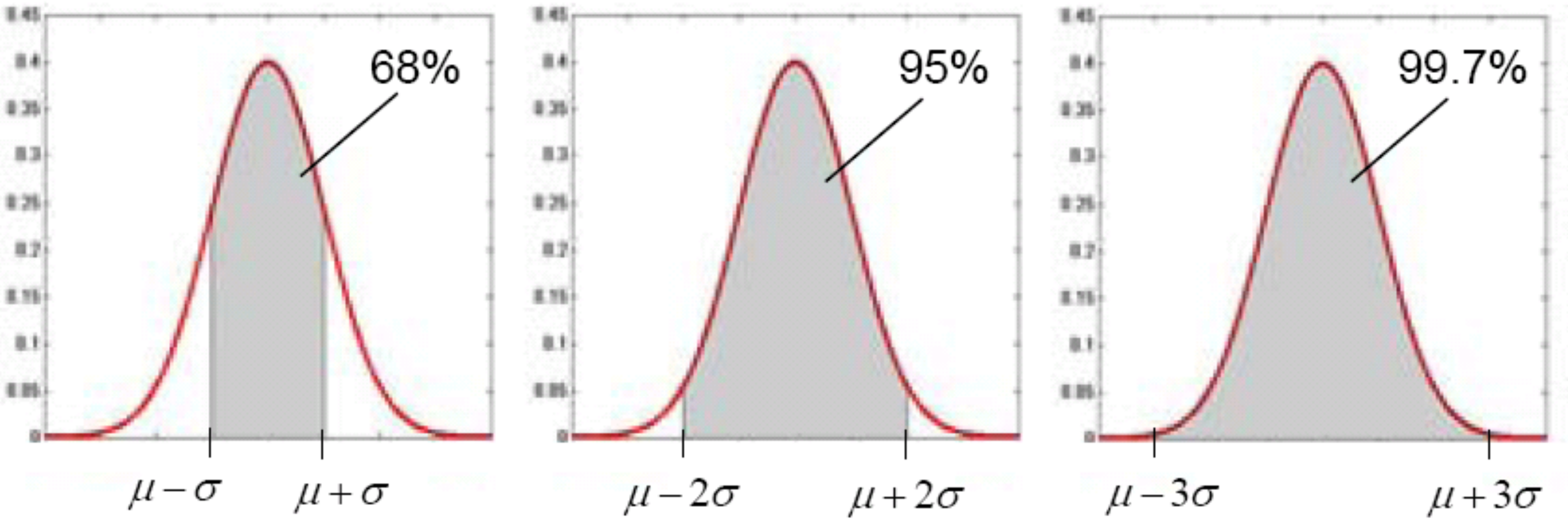
N(50,5) model

N(0,1) model

Area between (45,55) = Area between (-1,1)

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 $\mu = 72$ and $\sigma = 8$.

- 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?

Standard Normal

- If Z ~ N(0, 1), i.e., if Z is a normal random variable with $\mu=0$, $\sigma=1$. 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 ~ N(μ, σ^2), then X could be written as

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

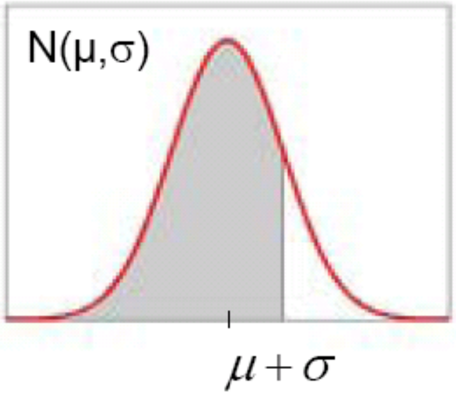
  where Z is a standard normal rv.

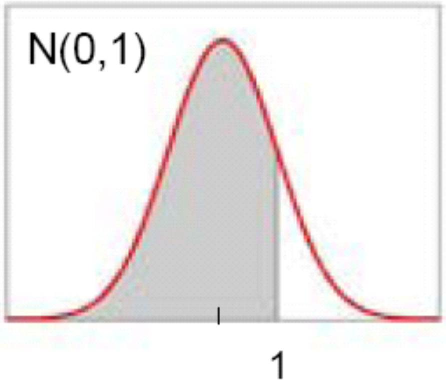
- Conversely, if X ~ N(μ, σ^2), then

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

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

Key Result





$$area\{y<\mu+\sigma\} \quad = \quad area\{z<1\}$$

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%

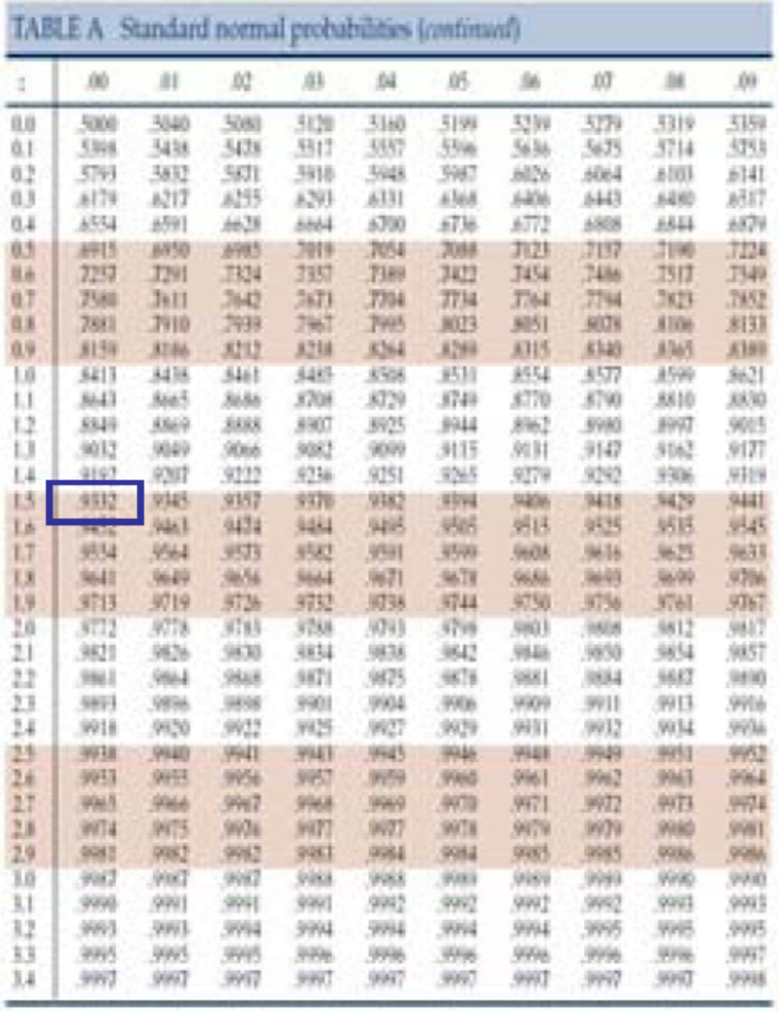


TABLE A Standard normal probabilities (*continued*)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

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.

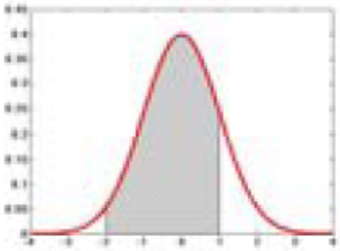
Ex. $P(a \le X \le b) = P[(a-\mu)/\sigma \le (X-\mu)/\sigma \le (b-\mu)/\sigma] = P[(a-\mu)/\sigma \le Z \le (b-\mu)/\sigma]$.

- 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 $P(Z \le 0.76)$; $P(Z > 0.76)$; $P(-1.32 \le Z \le 0.76)$.

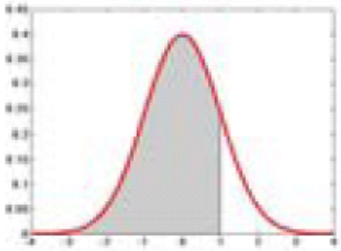
- Question: How to get the *p*-th percentile of the standard normal from A.3?

Using the Normal Table



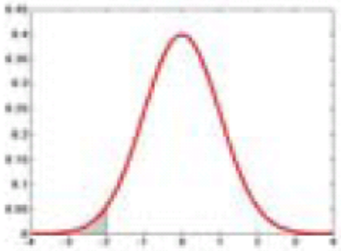
0.8185

=

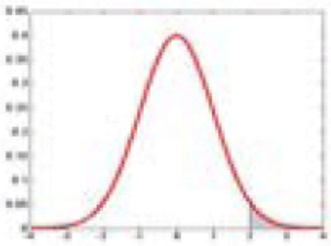


0.8413

-

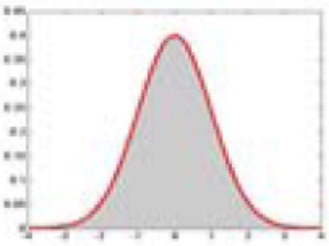


0.0228



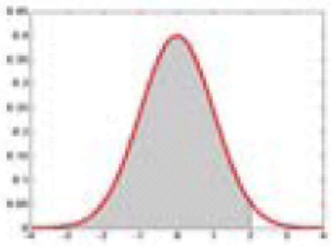
0.0228

=



1.00

-



0.9772

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?