Chapter 5. PROBABILITY DENSITIES (A)

February 15, 2011

1 Probability Density Functions

Continuous Random Variables

- Continuous random variable: a random variable that can assume any values within an interval or collection of intervals.
- Examples: heights; length; speed of a car; the amount of alcohol in a person's blood; the tensile strength of a new alloy.

- Probability mass function is not applicable: If X is a continuous random variable, then for any specific real value x, P(X=x) $=$ 0!

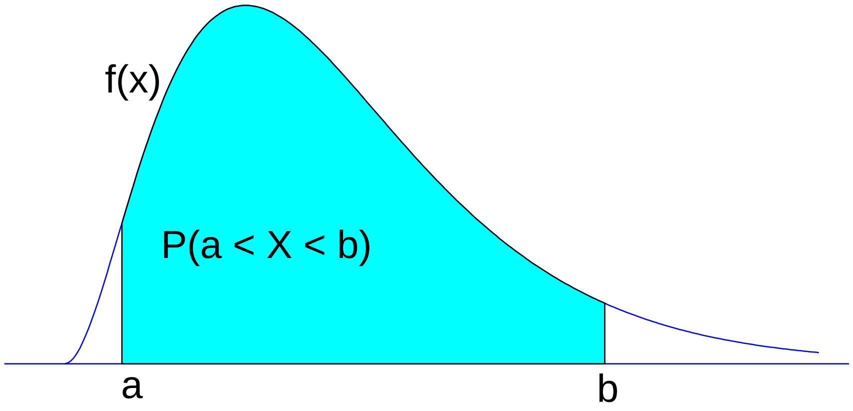
- It is meaningful only to characterize the probability distribution of a continuous random variable, X, as P($a < X < b$) for any interval (a,b)

- Mathematically, for continuous random variable, X,

$$P(a < X < b) = \int_a^b f(x)dx.$$

or $P(a < X \leq b) = \int_a^b f(x)dx$. or $P(a \leq X \leq b) = \int_a^b f(x)dx$.

Meaning of f(x) $f(x) \geq 0$ is called the probability density function (p.d.f., or density function). It is a continuous function on the interval or collection of intervals, on which X assumes values



A function f(x) is called a probability density function if it meets the following requirements:

- $f(x) \geq 0$ for all x
- The total area under the curve is 1. i.e.

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

- On the other hand, any function $f(x)$ satisfies the above two conditions simultaneously can be a density function of a random variable.

Example Find k so that the following can serve as the probability density of a random variable:

$$f(x) = \begin{cases} 0, & \text{for } x \leq 0 \\ kxe^{-4x^2}, & \text{for } x > 0 \end{cases}$$

find k such that it is a density function.

Solution: since $f(x) \geq 0$, we only need to make

$$\int_{-\infty}^{\infty} f(x)dx = \int_{0}^{\infty} kxe^{-4x^2}dx = \int_{0}^{\infty} \frac{k}{8}e^{-u}du = \frac{k}{8} = 1$$

so that $k = 8$.

Example If a random variable has the probability density

$$f(x) = \begin{cases} 2e - 2x & \text{for } x > 0 \\ 0, & \text{for } x \leq 0 \end{cases}$$

find the probabilities that it will take on a value

(a) between 1 and 3 ;

(b) greater than 0.5 .

Evaluating the necessary integrals, we get

(a) $\int_1^3 2e^{-2x} dx = e^{-2} - e^{-6} = 0.133$

(b) $\int_{0.5}^{\infty} 2e^{-2x} dx = e^{-1} = 0.368$

Cumulative distribution

- $F(x) = P(X < x)$ is called Cumulative Distribution Function (C.D.F.), i.e.

$$F(x) = \int_{-\infty}^{x} f(t)dt$$

- Therefore, we have the following relationship between $f(x)$ and $F(x)$,

$$f(x) = \frac{dF(x)}{dx}, \qquad \text{and} \qquad F(x) = \int_{-\infty}^{x} f(t)dt$$

- if X has C.D.F. F(x), then

$$P(a < X < b) = F(b) - F(a)$$

- F(x) must satisfy the following conditions:
  - F(x) is an non-decreasing function of x.
  - $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$
- If any function F(x) satisfies the above two conditions simultaneously, then it can be a C.D.F. of a random variable.
- **Example** For the above X in the previous Example, its C.D.F. is

$$\begin{aligned} F(x) &= P(X \le x) = \int_{-\infty}^{x} f(t)dt \\ &= \begin{cases} \int_0^x 2e^{-2t}dt, \text{ if } x > 0 \\ 0 \text{ if } x \le 0 \end{cases} \\ &= \begin{cases} 1 - e^{-2x}, \text{ if } x > 0 \\ 0 \text{ if } x \le 0 \end{cases} \end{aligned}$$

2 Expectation

- Definition: for X has pdf $f(x)$, then its expectation is defined as

$$EX = \int_{-\infty}^{\infty} x f(x) dx$$

In calculus, it is known

$$\int_{-\infty}^{\infty} x f(x) dx \approx \sum_i x_i f(x_i)(x_{i+1} - x_i) \approx \sum_i x_i p_i$$

which is roughly the expectation for discrete variables.

- **kth moment about the origin**

$$\mu_k' = \int_{-\infty}^{\infty} x^k f(x) dx$$

- **kth moment about the mean**

$$\mu_k = \int_{-\infty}^{\infty} (x-\mu)^k f(x)dx$$

μ_4 can tell us the Kurtosis of a distribution, and μ_3 gives us information about skewness (not discussed in this module).

- **Variance and standard deviation**

  – Variance

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2.$$

In calculus, we have again

$$\int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \approx \sum_{\ell} (x_\ell - \mu)^2 f(x_\ell)(x_{\ell+1} - x_\ell) \approx \sum_{\ell} (x_\ell - \mu)^2 p_\ell$$

which is roughly the variance for discrete variables.

– standard deviation

$$\sigma = \sqrt{\sigma^2}$$

– interpretation of σ^2 and σ: the dispersion of the values of the random variable according to the probability; unstableness of X; ...

Example for the distribution above

$$\mu = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{\infty} x \cdot 2e^{-2x} dx = 0.5$$

and

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \int_{0}^{\infty} (x-\mu)^2 \cdot 2e^{-2x} dx = 0.25$$

3 Normal Distribution

Under certain conditions the mean of a number of random variables with finite means and variances approaches a normal distribution as the number of variables increases.

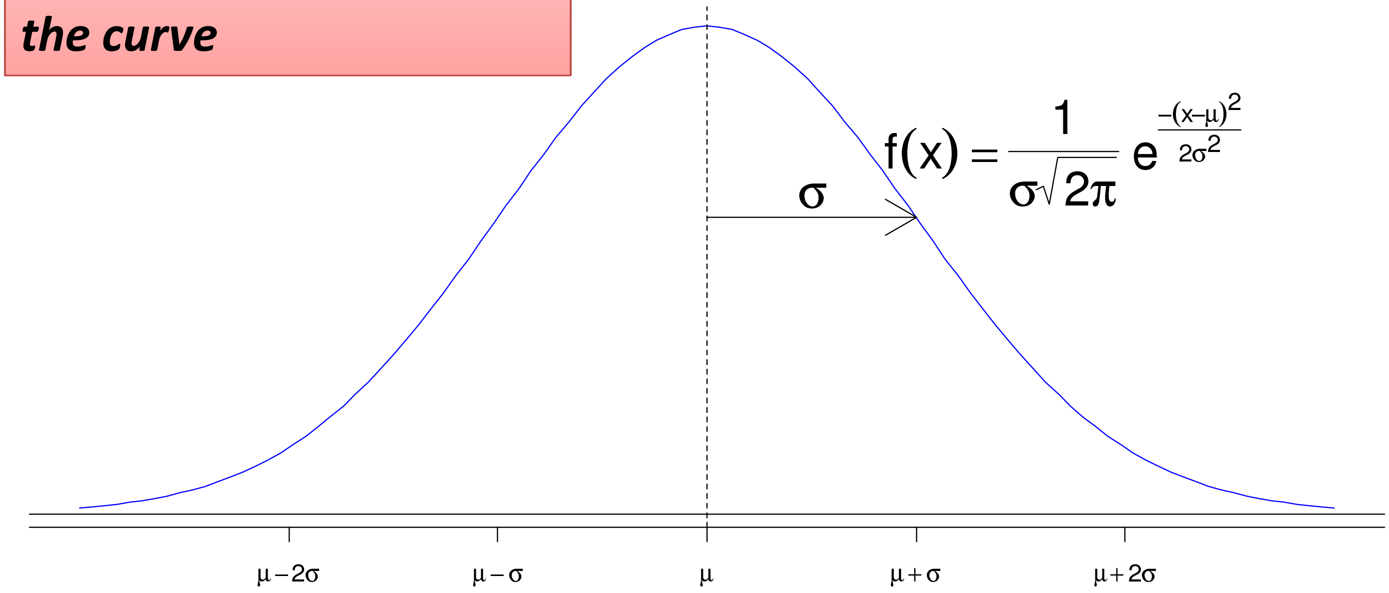
- The most important distribution in real application – normal distribution.
- A random variable, X, is normally distributed, if its density function is

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

- The graph of the above density function is bell shaped (refer to the next slide).

curve of the normal density function

$e = 2.718$ and $\pi = 3.142$; μ and σ ($\sigma > 0$) are the parameters that affects the center and spread of the curve



Cumulative Distribution Function (CDF) of Normal distribution

- The CDF of normal distribution is

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

- No closed form for the C.D.F. of normal distribution. Numerical method is required to evaluate P(X $\leq x$) for some specific x.

Expectation and Variation

- If a random variable, X, follows a normal distribution with density function,

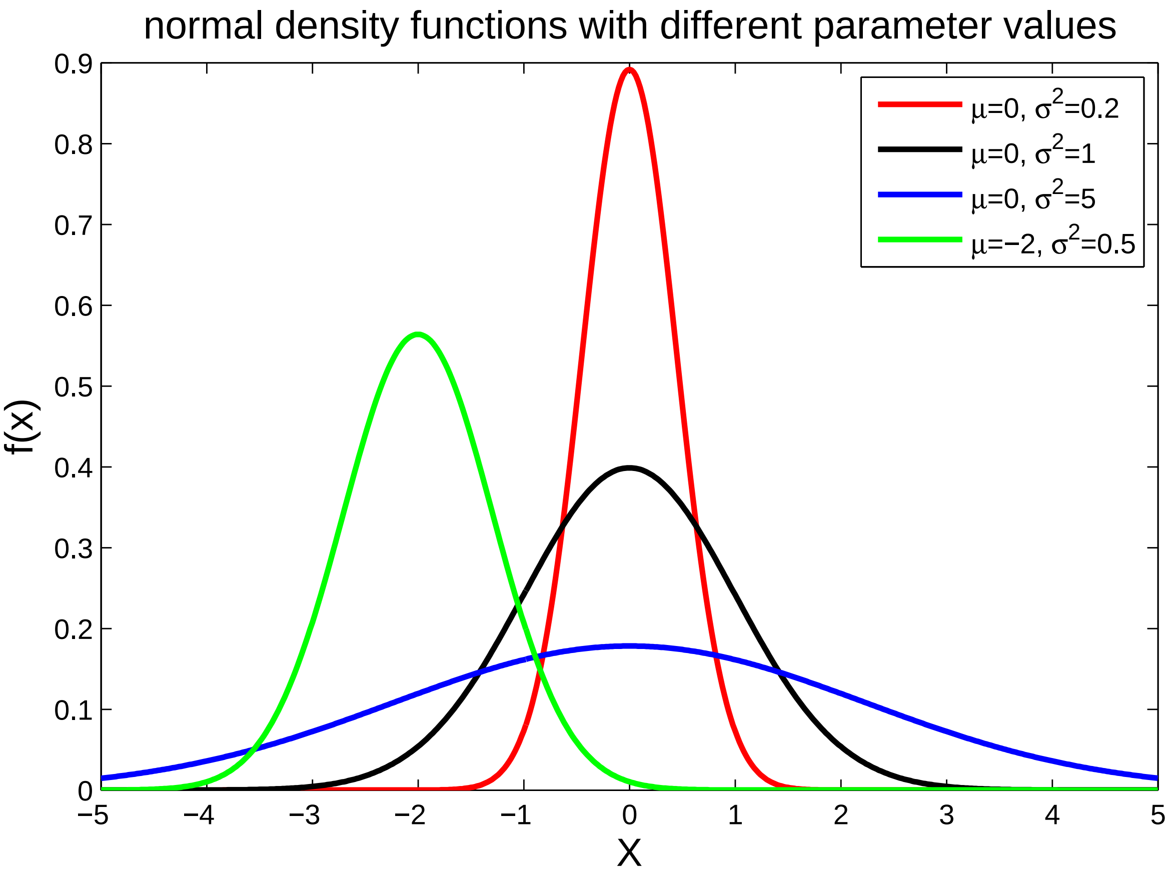
$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

then

$$E(X) = \mu \qquad \& \qquad var(X) = \sigma^2$$

μ and σ^2 are those encountered in $f(x;\mu,\sigma^2)$

- Usually, we use $X \sim N(\mu,\sigma^2)$ to denote a random variable, X, following normal distribution with mean μ , variance σ^2.


normal density functions with different parameter values
🟥 μ=0, σ^2=0.2
⬛ μ=0, σ^2=1
🟦 μ=0, σ^2=5
🟩 μ=−2, σ^2=0.5
f(x)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
−5
−4
−3
−2
−1
0
1
2
3
4
5
X

Standard Normal Distribution

- When $\mu = 0$ & $\sigma^2 = 1$, the resulted normal distribution is called standard normal distribution.

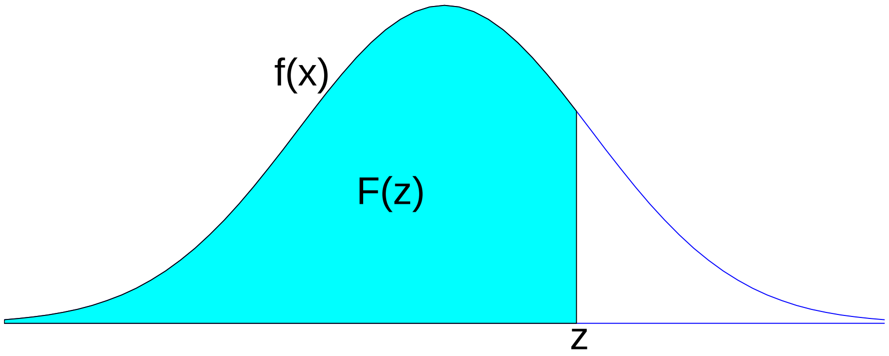
- A random variable follows a standard normal distribution is denoted as Z, whose C.D.F is given by

$$F(z) = P(Z \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$$

  Most commonly, it is denoted by $\Phi(z)$.

- We denote $Z \sim N(0, 1)$

- The value of F(z) is only numerically accessible, and can be read directly from Table 3 at the end of the textbook (or below).



- Note that because the distribution of Z is symmetric about 0, we have

$$P(Z < -z) = P(Z > z)$$

Other facts

$$\Phi(z) = 1 - \Phi(-z)$$

and if $z > 0$

$$\Phi(|Z| > z) = 2\Phi(-z)$$

- Example:

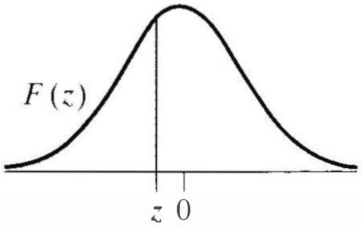
1. $P(0.87 < Z \leq 1.28) =$

2. $P(Z > 2) =$

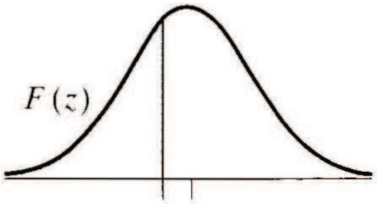
3. $P(|Z| < 1.96) =$

Standard Normal Distribution Function

$$F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt$$



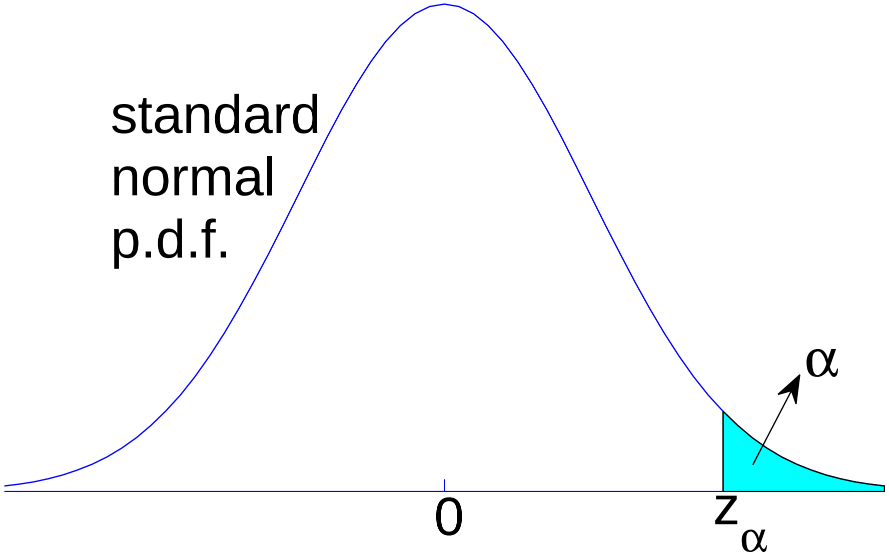
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−5.0	0.0000003									
−4.0	0.00003									
−3.5	0.0002									
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0006	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

- What if the probabilities relating to standard normal distribution are given and need to find the corresponding values of z, denoted as Z_α :

$$\alpha = P(Z > Z_\alpha)$$



- Important values for z_α:

$$z_{0.01} \text{ and } z_{0.05} \text{ and } z_{0.025}$$

- Note the fact that $F(z_{0.01}) = 0.99$, read from the statistical table

$$z_{0.01} = 2.33$$

- Note the fact that $F(z_{0.025}) = 0.975$, read

$$z_{0.0975} = 1.96$$

- Note the fact that $F(z_{0.05}) = 0.95$, read

$$z_{0.05} = 1.64$$

- **Example** find z such that

1. $P(|Z| > z) = 0.05$

   $$P(|Z| > z) = P(Z > z) + P(Z < -z) = 2P(Z > z) = 0.05$$

   Thus

   $$P(Z > z) = 0.025$$

   we have $z = 1.96$

2. $P(|Z| > z) = 0.01$

   $$P(|Z| > z) = P(Z > z) + P(Z < -z) = 2P(Z > z) = 0.01$$

   Thus $P(Z > z) = 0.005$, we have $z = 2.58$

Connect $N(\mu, \sigma^2)$ to $N(0,1)$

– Property: if $X \sim N(\mu, \sigma^2)$, let

$$Z = \frac{X - \mu}{\sigma}$$

then $Z \sim N(0,1)$ (note that $EZ = 0$ and $Var(Z) = 1$)

– Note that

$$\begin{aligned} P(X \leq x) &= P(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}) \\ &= P(Z \leq \frac{x-\mu}{\sigma}) \\ &= F(\frac{x-\mu}{\sigma}) \end{aligned}$$

we can look up $F(\frac{x-\mu}{\sigma})$ from the statistical Table

- When X has the normal distribution with mean μ and standard deviation σ

  then

$$P(a < X \leq b) = F(\frac{b-\mu}{\sigma}) - F(\frac{a-\mu}{\sigma})$$

two special cases

$$P(X > a) = 1 - F(\frac{a-\mu}{\sigma})$$

and

$$P(X \leq b) = F(\frac{b-\mu}{\sigma})$$

Example Let X be the number of power outrage per month of a city. Assume approximately

$$X \sim N(11.6,\ 3.3^2)$$

Find the probability of at least 8 outrages in any one month

$$P(X \geq 8) = 1 - F(\frac{8 - 11.6}{3.3}) = 1 - F(-1.24) =$$

Example The actual amount of instant coffee that filling a machine puts into "4-ounce" may be looked upon as a random variable having a normal distribution with $\sigma = 0.04$ ounce. If only 2% of the jars are to contain less than 4 ounces, what should be the mean fill of these jars?

To find μ such that

$$F(\frac{4-\mu}{0.04}) = 0.02$$

The value in the Table closest to 0.02 is 0.0202, corresponding to Z = -2.05.

$$\frac{4-\mu}{0.04} = -2.05$$

we find that $\mu = 4.082$ ounces.

The Normal Approximation to the Binomial Distribution

For a binomial random variable, X, when n is large, then,

$$X \approx N(np, np(1-p))$$

or

$$P(X \leq x) \approx F(\frac{x - np}{\sqrt{np(1-p)}})$$

Usually, when np and n(1-p) are both greater than 15 then the approximation is reasonably accurate

Example If 20% of the memory chips made in a plant are defective, what is the probability of at most 15 defectives among 100 randomly chosen memory chips. Rule of thumb: np = 100(20%) = 20 >15, n(1-p)=80>15. Therefore,

Since $np = 100(20\%) = 20 > 15, n(1-p) = 80 > 15$. Therefore, $X \sim N(20, 16)$

Using normal: $X \approx N(20, 16)$

$$P(X \leq 15) \approx F(\frac{15-20}{4}) = 0.1292$$

Using binomial: B(100, 0.2)

$$P(X \leq 15) = C_{100}^{0} p^0 (1-p)^{100} + ... + C_{100}^{15} p^{15} (1-p)^{85} = 0.1285$$

Poisson Approximation Versus Normal Approximation

- Possion Approximation requires n large and p small. In fact, providing large n, its main concern is “fixed (or small) np”! It is more applicable for the cases of “when $n \to \infty$, $p \to 0$ and in the end, n p is fixing.

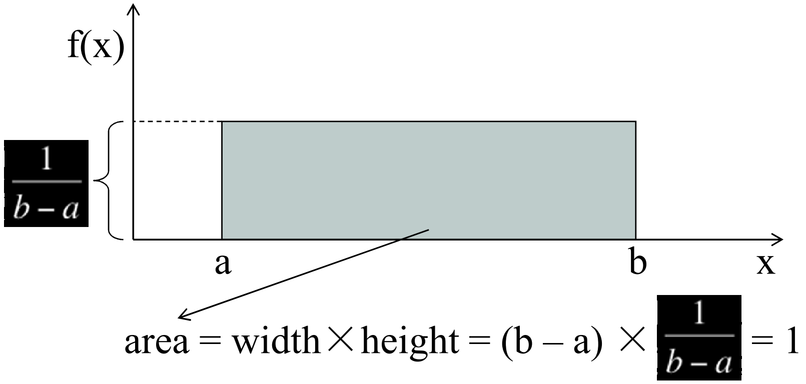
- Normal approximation requires n large. Its application domain contains the cases that p fix and $n \to \infty$. Since now, $np \to \infty$

Uniform Distribution

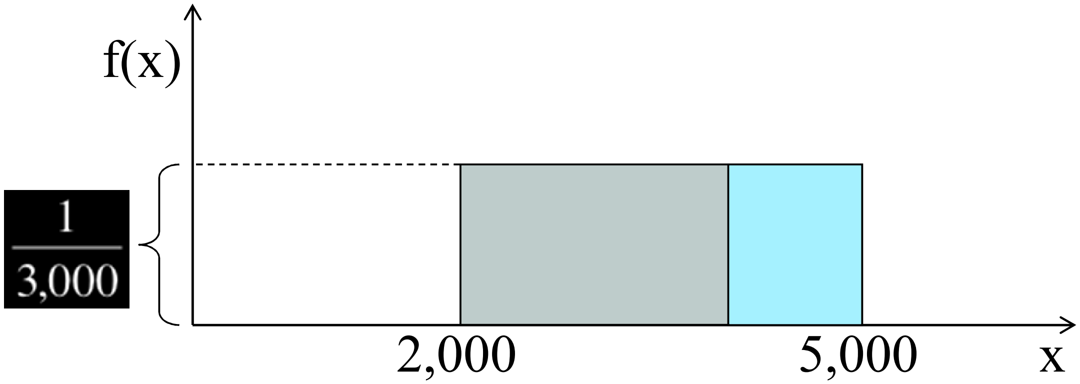
Consider the uniform probability distribution (sometimes called the rectangular probability distribution).

Suppose X is distributed "equally likely" for all values in [a, b]. It is described by the function:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$



Example The amount of gasoline sold daily at a service station is uniformly distributed with a minimum of 2,000 gallons and a maximum of 5,000 gallons.



What is the probability that the service station will sell at least 4,000 gallons?

Algebraically: what is P(X $\geq$ 4,000)?

P(X $\geq$ 4,000) = (5,000 - 4,000) $\times$ (1/3000) = .3333

- Expectation:

$$EX = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b \frac{1}{b-a} x dx = \frac{1}{b-a} \frac{x^2}{2}\Big|_a^b = \frac{a+b}{2}.$$

- Variance:

$$Var(X) = \int_{-\infty}^{\infty} (x - E(X))^2 f(x) dx = \frac{1}{12}(b-a)^2.$$

- C.D.F.

$$F(x) = \int_{-\infty}^{x} f(t) dt = \begin{cases} 0, & \text{if } x < a \\ \frac{x-a}{b-a}, & \text{if } a \leq x \leq b \\ 1, & \text{if } x > b \end{cases}$$

- **Example** A student attending ST2334 travel to LT25 by bus 95, which arrives at the bus stop near his home every 15 mins. It takes 20 mins for him to arrive at LT25 after he gets on the bus. (1) The time, X, he needs to arrive at LT25 from the bus stop is uniform distribution with $a = 20, b = 35$ (2) Assume that this morning, he arrives at the bus stop at 9:35.

  The mean time for him go to LT25: $\mu = (a + b)/2$

  The variance of his time on the road: $\sigma^2 = (b - a)^2/12$

  The probability that he will arrive before the lecture starts (i.e. in less then 25 mins)

$$P(X < 25)$$

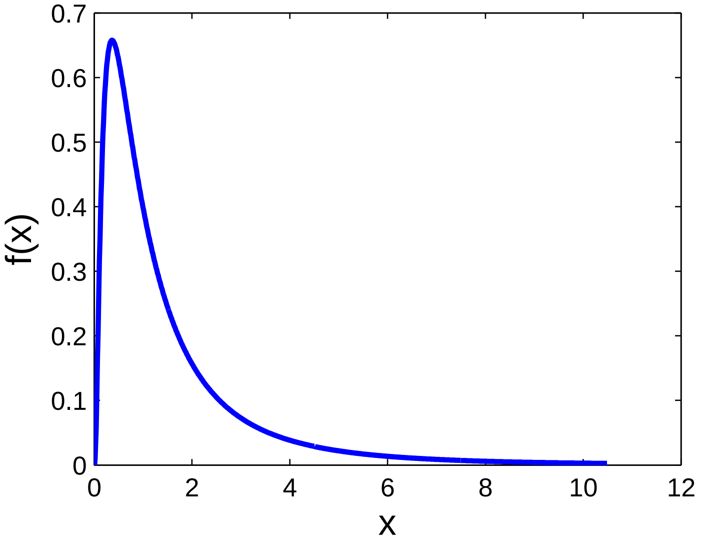
log-normal distribution

- X follows log-normal distribution, if $\ln(X)$ follows a normal distribution N(μ, σ^2)

- p.d.f.

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} x^{-1} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, & \text{if } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

where $\mu, \sigma^2 > 0$ are parameters.



- Expectation

$$EX = e^{\mu+\sigma^2/2}$$

Variance

$$Var(X) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}$$

- For and $a > 0$ and $b > 0$,

$$P(a < X \leq b) = F(\frac{\ln b - \mu}{\sigma}) - F(\frac{\ln a - \mu}{\sigma})$$

and F is the C.D.F. of the standard normal distribution.

- **Example** The current gain of certain transistors equal to $\ln(I_0/Ii)$, which is the log ratio of the output to the input current. It follows N(2, 0.01). Find $P(6.1 < I_0/I_i \leq 8.2)$.

  Answer : $\alpha = 2, \beta = 0.1$, use the property :

$$\begin{aligned} P(6.1 < I_0/I_i \leq 8.2) &= F(\frac{\ln 8.2 - 2}{0.1}) - F(\frac{\ln 6.1 - 2}{0.1}) \\ &= F(1.0) - F(-1.92) = 0.8139 \end{aligned}$$

4 Checking If the Data Are Normal

Why Checking Normality

- On one hand, the distribution of many real data can be well approximated by normal distribution.
- On the other hand, many statistical methods are based on the assumption that the collected data are normally distributed.
- Normality is widely used as basic, preliminary conditions in many statistical methods. But mis-specification can result serious errors in statistical inference.

Methods of Checking Normality

- Use histograms to check symmetry: if the data follows the normal distribution, its histogram must at least bell shaped. Skewed histogram implies the data are not normally distributed.

- Normal scores plot (normal quantile plot, QQ-plot) provides effect way of checking normality.

Normal Score

- Normal Scores $(m_i, i = 1, ..., n)$: values of z that separate the region under the curve of standard normal density into n+1 parts of equal areas.

- n can be 2, 3, 4, ...

- Example, n = 4, then,

$$m_1 = -z_{0.20} = -0.84$$

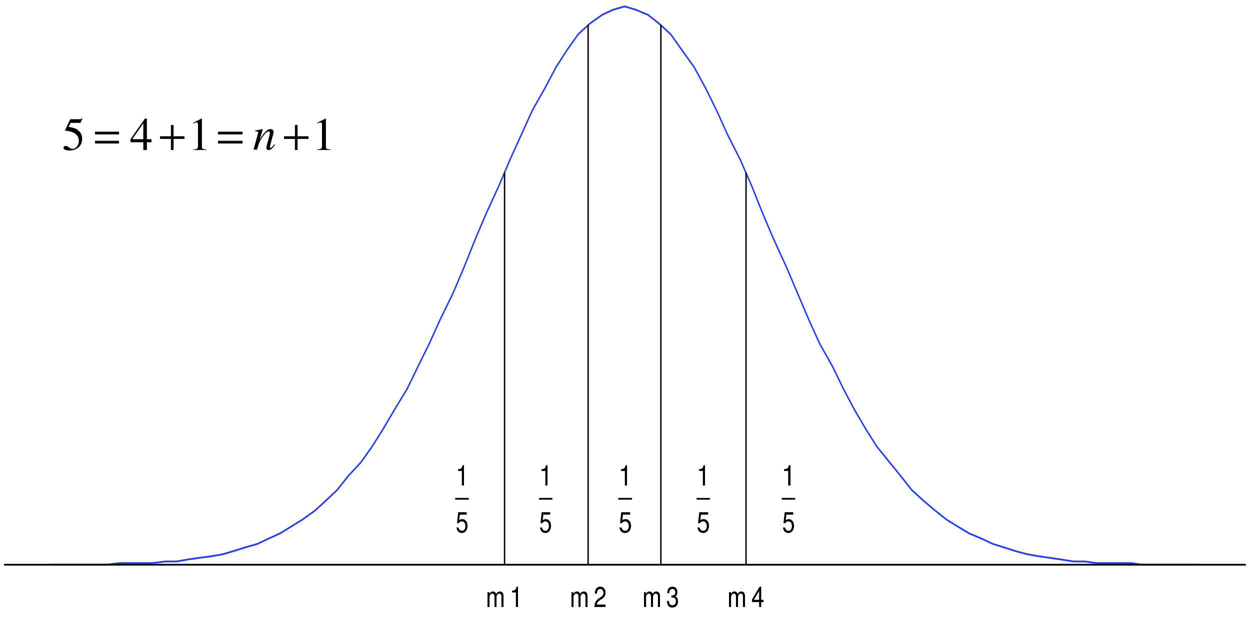
$$m_2 = -z_{0.40} = -0.25$$

$$m_3 = -z_{0.40} = 0.25$$

$$m_4 = -z_{0.20} = 0.84$$

Recall: for any $0 < \alpha < 1$, is the one satisfying $P(Z > z_\alpha) = \alpha$.

- More intuitively



- In general, for any n, normal scores consist of n values and

  1. The n values are

  $$Z_{1-1/(n+1)}, Z_{1-2/(n+1)}, , \ldots, Z_{1-n/(n+1)}$$

  2. These n values separate the region under the curve of standard normal density into n+1 parts of equal areas.

- Normal scores refers to an ideal/"perfect" sample from the standard normal distribution.

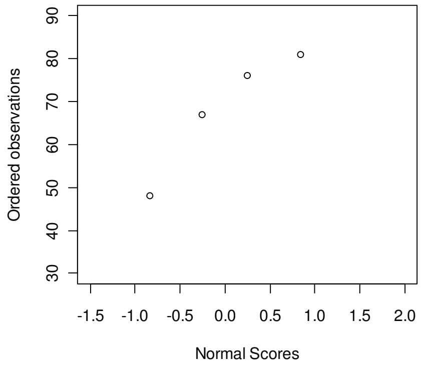
- Follow the steps below to construct a normal scores plot, (assume the data size is n)

  1. Order the data (with sample size n) from smallest to largest.
  2. Obtain the normal scores
  3. Plot the ith smallest observation versus the ith normal score m_i for all i $=$ 1, ...,n.

- Suppose four observations are 67, 48, 76, 81. Construct a normal scores plot. Three steps

  1. Order data are: 48, 67, 76, 81
  2. Normal scores: -0.84, -0.25, 0.25, 0.84
  3. Plot (48, -0.84), (67, -0.25), (76, 0.25) and (81, 0.84)

  The constructed normal scores plot:



Checking the Normality

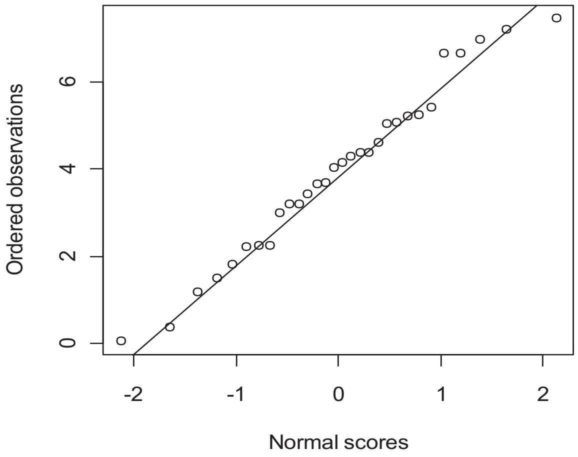
- If the data follow the standard normal distribution, the points of the plot will be roughly on the 45^o line.

- If the data follow a normal distribution, the points of the plot will be roughly follow a line.

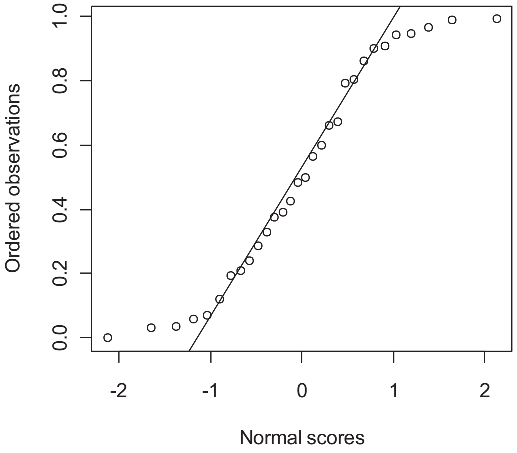
- By observing whether points of normal scores plot are roughly on a line, we can conclude whether normality is roughly ok to be assumed.

- **Example**

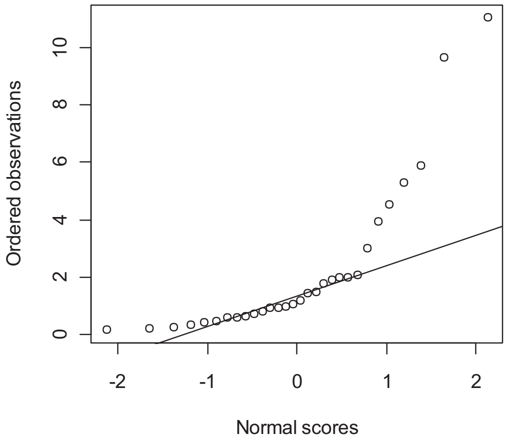
  - When the data are normally distributed



  - When the data are uniformly distributed



– When the data are log-normally distributed



- Usually we need a minimum 15 observations (i.e. $n \geq 15$) in order to evaluate its agreement with normality.

- Normal Scores Plot have been implemented and easy to be used in most statistical software.

  In R,

  ```
  qqnorm(data)
  ```

Transforming Observations to Near Normality

When the histogram or normal scores plot indicate that the assumption of a normal distribution is invalid, transformations of the data can often improve the agreement with normality.

- To make large values smaller (when the distribution skewed to the right), we can use

$$-\frac{1}{x}, \quad \ln(x), \quad x^{1/4}, \quad x^{1/2}$$

- To make large values larger (when the distribution skewed to the left), we can use

$$x^2, \quad x^3$$

Example A computer scientist, trying to optimize system performance, collected data on the time, in microseconds, between requests for a particular process service, 2808, 4201, 3848, 9112, 2082, 5913, 1620, 67192, 1657, 3072, 2949, 11768, 4731, 14211, 1583, 9853, 78811, 6655, 1803, 7012, 1892, 4227, 6583, 15147, 4740, 8528, 10563, 43003, 16723, 2613, 26463, 34867, 4191, 4030, 2472, 28840, 24487, 14001, 15241, 1643, 5732, 5419, 28608, 2487, 995, 3116, 29508, 11440, 28336, 3440.

```
hist(c(2808, 4201, 3848, 9112, 2082, 5913, 1620, 67192, 1657, 3072, 2949, 11768, 4731, 14211, 1583, 9853,
78811, 6655, 1803, 7012, 1892, 4227, 6583, 15147, 4740, 8528, 10563, 43003, 16723, 2613, 26463, 34867, 4191, 4030,
2472, 28840, 24487, 14001, 15241, 1643, 5732, 5419, 28608, 2487, 995, 3116, 29508, 11440, 28336, 3440))
```

```
qqnorm(c(2808, 4201, 3848, 9112, 2082, 5913, 1620, 67192, 1657, 3072, 2949, 11768, 4731, 14211, 1583, 9853,
78811, 6655, 1803, 7012, 1892, 4227, 6583, 15147, 4740, 8528, 10563, 43003, 16723, 2613, 26463, 34867, 4191, 4030,
2472, 28840, 24487, 14001, 15241, 1643, 5732, 5419, 28608, 2487, 995, 3116, 29508, 11440, 28336, 3440))
```

```
qqnorm(log(c(2808, 4201, 3848, 9112, 2082, 5913, 1620, 67192, 1657, 3072, 2949, 11768, 4731, 14211,
1583, 9853, 78811, 6655, 1803, 7012, 1892, 4227, 6583, 15147, 4740, 8528, 10563, 43003, 16723, 2613, 26463, 34867,
4191, 4030, 2472, 28840, 24487, 14001, 15241, 1643, 5732, 5419, 28608, 2487, 995, 3116, 29508, 11440, 28336, 3440)))
```

Please copy and paste to R of each command, and check the normality assumption