Part 2 – Statistical Considerations
ME 3000

Dr. Shad Roundy

Spring 2024

Department of
MECHANICAL ENGINEERING
THE UNIVERSITY OF UTAH

Why is statistics important in design?

Why is statistics important in design?

- We must design elements such that
  - They fit with other elements during assembly
  - The part will perform its function given variation in design parameters and operating environment

Key Concepts

- Random variables & distributions
- Metrics to describe a distribution / data
- Linear combinations of variables
- Types of distributions
  - Uniform, Normal, Weibull

Random Variables

- Variable whose value depends on the outcome of a random event
- Discrete random variable

  – roll of a die, # of defects

- Continuous random variable

  length, stiffness, weight

- If I sample a part and measure a hole diameter, is diameter, D, a random variable? Why or why not?

  – Drilling of hole is random event
  – Measuring " " "
  – Selecting sample is a random event

Random Variables - Notation

- In statistics, we used UPPER CASE to denote a random variable

$$X$$

- And, we used lower case to denote some specific value of that variable

$$P(X < x) = \ \dots$$

- Finally, we used subscripts to refer to a particular item drawn from a population

$$x_i = 32.2 \text{ mm}$$

- But, the two textbooks are not consistent with this notation, so just make sure you understand the context

Probability Density Functions

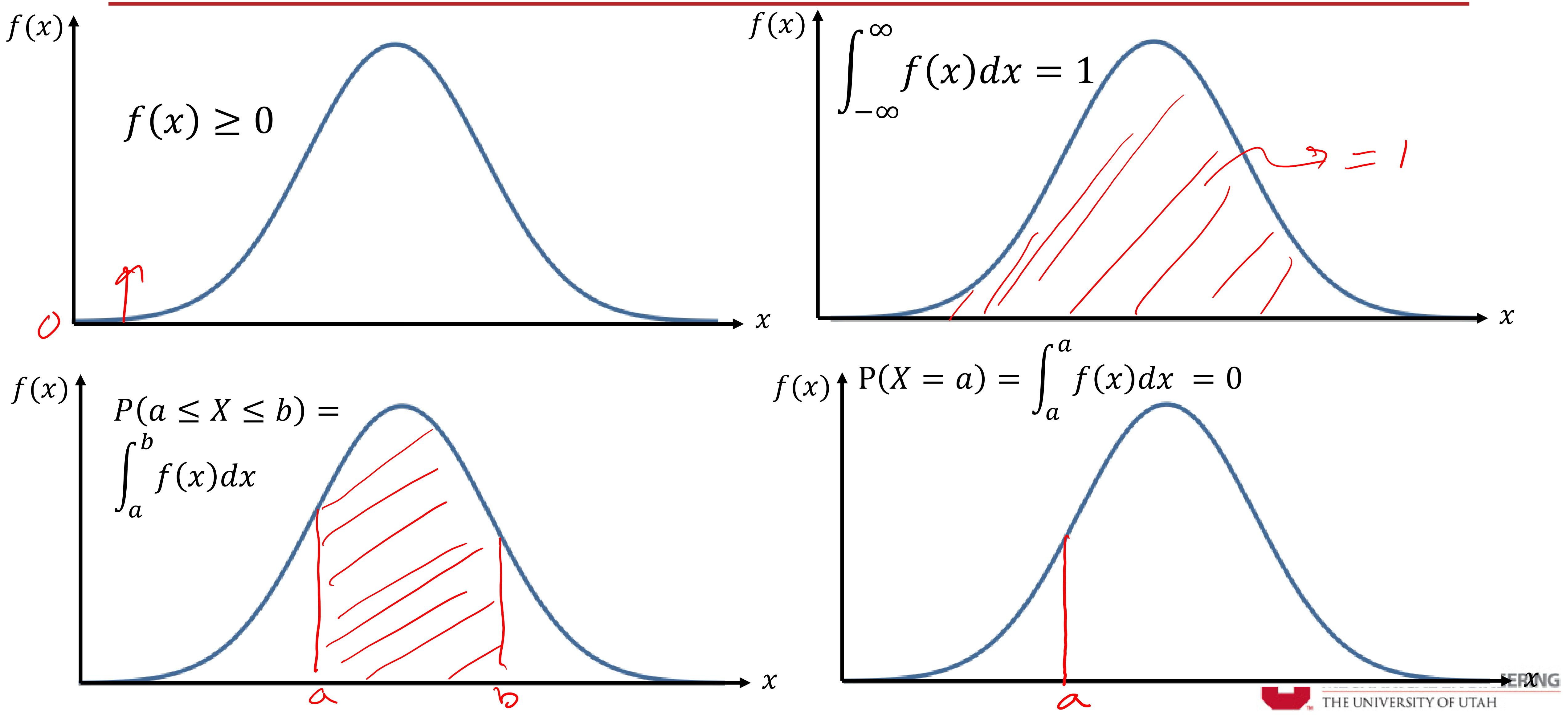
- We describe the distribution of a random variable with a **probability density function (PDF)**
- For a continuous random variable X, a probability density function is a function such that:

1. $f(x) \geq 0$ (function is non-negative)
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a \leq X \leq b) = \int_a^b f(x)dx$ = area under $f(x)$ from a to b
4. $\mathrm{P}(X = a) = \int_a^a f(x)dx = 0$

   (there is no area under the curve exactly at a)

Department of MECHANICAL ENGINEERING THE UNIVERSITY OF UTAH

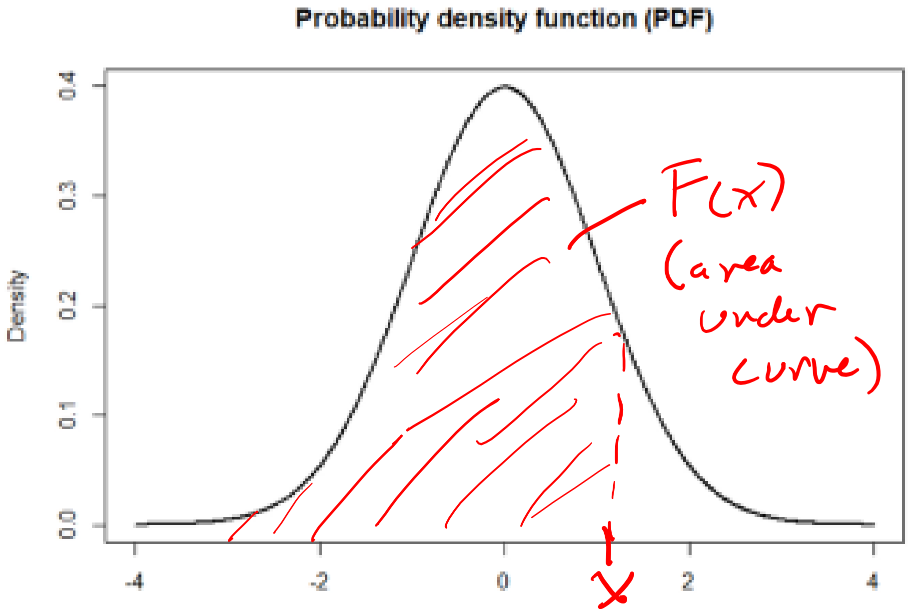
Probability Density Functions

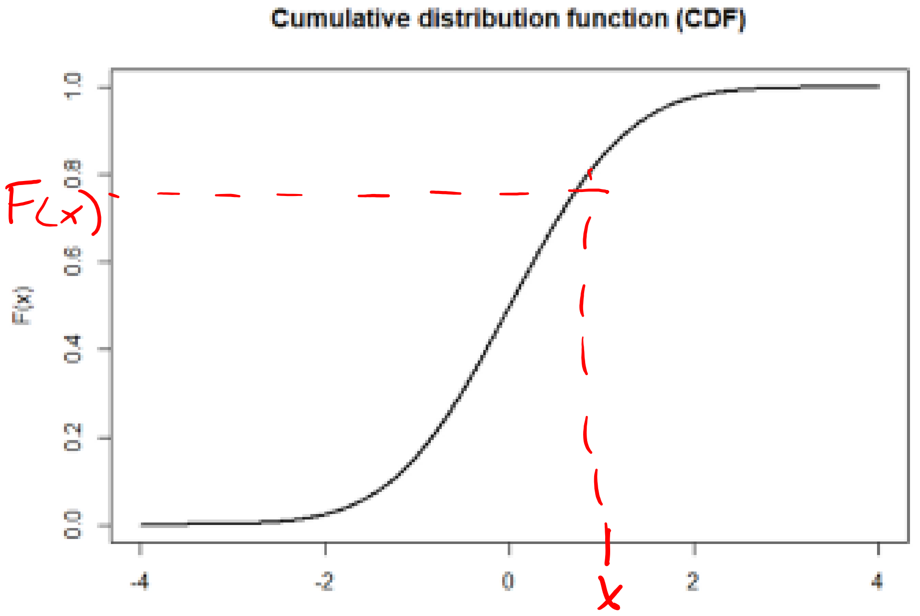


Cumulative Density Functions (CDF)

- Integral of the PDF
- The probability that the random variable X takes some value **less than or equal to** the specific value x

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

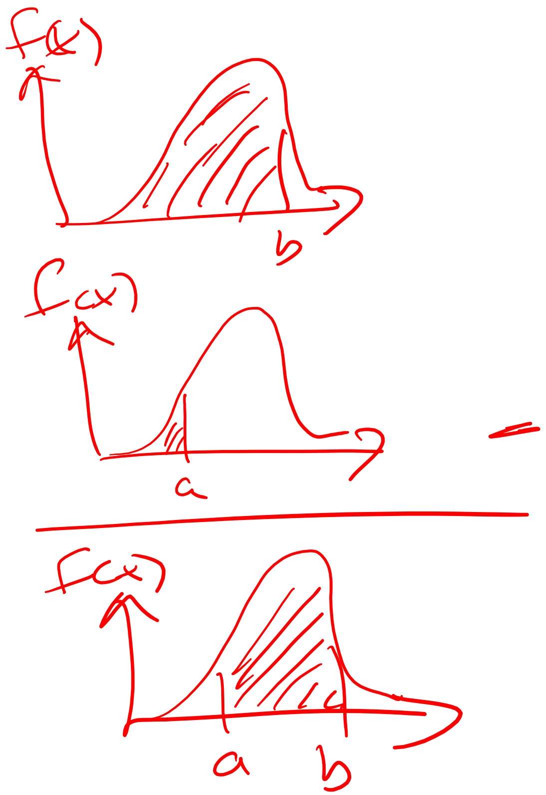




Cumulative Density Functions (CDF)

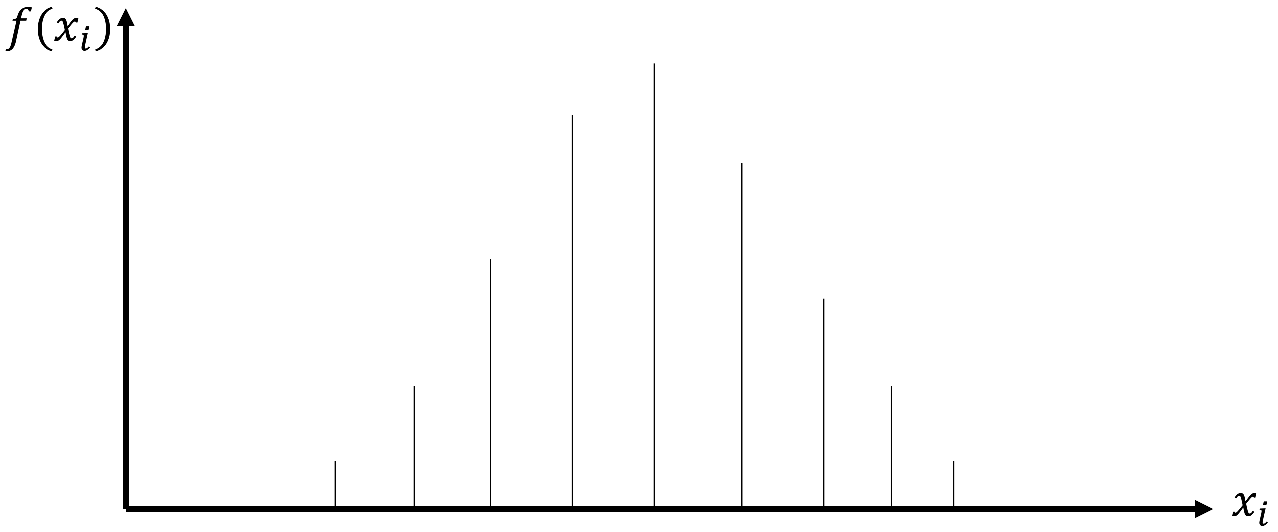
- How would you write the probability that some random variable X is between $x = a$ and $x = b$ (i.e., $P(a \leq X \leq b)$) in terms of the CDF?

$$P(a \leq X \leq b) = F(b) - F(a)$$



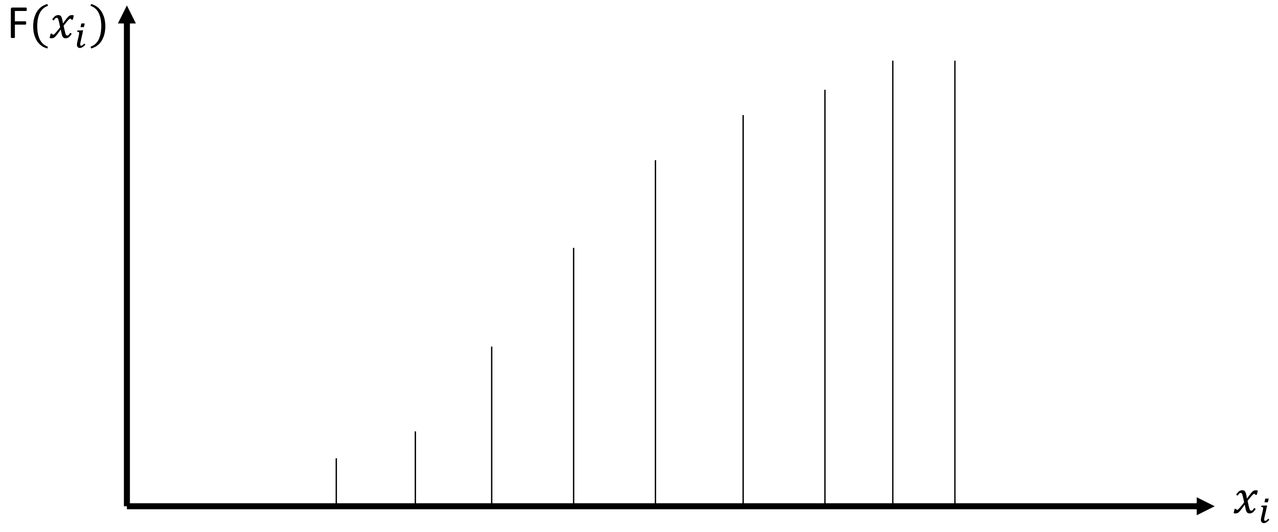
Discrete Probability Density Function (PDF)

1. $f(x_i) \geq 0 \quad for\ all\ x_i$ (function is non-negative)
2. $\sum f(x_i) = 1$
3. $P(x_m \leq X \leq x_n) = \sum_{i=m}^{n} f(x_i)$
4. $\mathrm{P}(X = x_i) = f(x_i)$

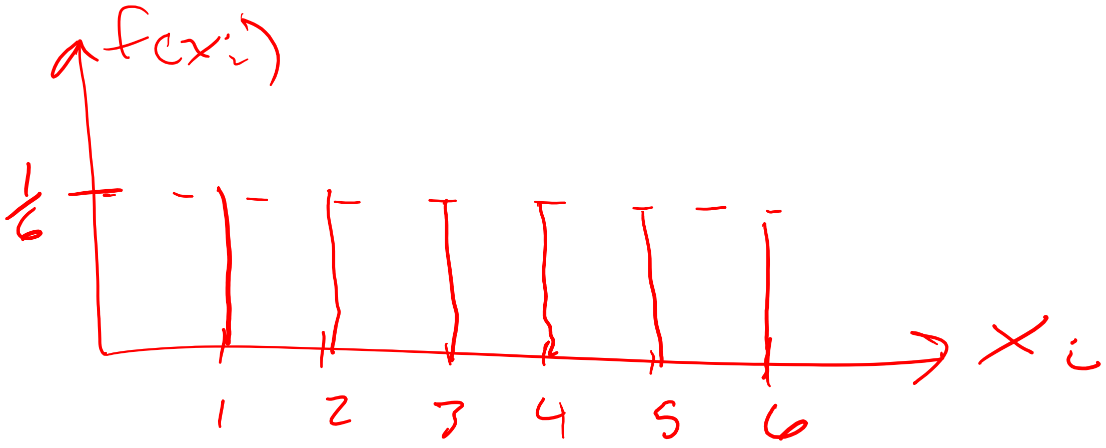


Discrete Cumulative Density Function (CDF)

$$P(X < x_m) = F(x_m) = \sum_{i=1}^{m} f(x_i)$$



Draw the PDF for the roll of a 6 sided die

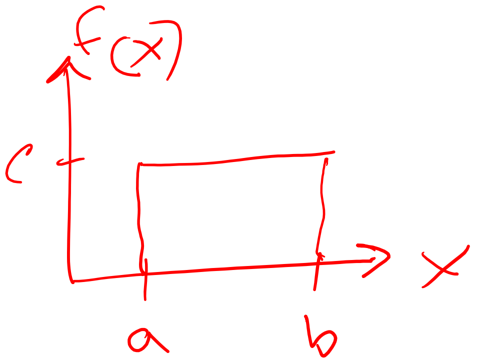


What is the probability that the any die roll will be between 2 and 5?

$$P(2 \leq X \leq 5) = \sum_{i=2}^{5} \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

Example 2.1

- The PDF of a uniform distribution is given by $f(x) = c$ for $a \leq x \leq b$. What is the CDF for this distribution?



$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{a}^{x} c\,dx$$

$$F(x) = c\left[x\right]_{a}^{x} = \boxed{c(x-a)}$$

Metrics to Describe Random Variables

Mean, Variance, Covariance, Correlation Coefficient

Linear combination of random variables

Department of MECHANICAL ENGINEERING
THE UNIVERSITY OF UTAH

Arithmetic Mean

- Continuous random variable

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

- Discrete random variable

$$\mu_x = \sum_{i=1}^{n} x_i f(x_i)$$

where there are n possible values for the discrete random variable X

Department of MECHANICAL ENGINEERING THE UNIVERSITY OF UTAH

Arithmetic Mean

- Mean of a sample of data

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

However, we call this $\bar{x}$ to indicate that it is a **sample** mean.

Variance

- A description of how wide the distribution is (i.e., how far data points deviate from the mean)
- Continuous random variable

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx$$

- Discrete random variable

$$\sigma_x^2 = \sum_{i=1}^{n} (x_i - \mu_x)^2 f(x_i)$$

Sample Variance and Standard Deviation

- **Sample variance** is the same equation as the variance for a discrete random variable, but we use s_x^2 instead of σ_x^2

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_x)^2$$

Any distribution

- **Standard deviation**

$$\sigma_x = \sqrt{\sigma_x^2}, \qquad s_x = \sqrt{s_x^2}$$

Covariance

- A measure of how “similar” two random variables, X and Y are
- General equation

$$Cov(X,Y) = E\left[X - E[X]\right] - E\left[Y - E[Y]\right]$$

where $E[]$ is the expected value (i.e. mean)

- Continuous random variable

$$\sigma_{xy} = \mu_{XY} - \mu_X \mu_Y = \int\int xyf(x,y)dxdy \; - \int xf(x)dx \; \int yf(y)dy$$

where $f(x,y)$ is the joint probability density function

- Discrete random variable – use general equation

Covariance of Samples

- For samples of two random variables (X and Y) of size n, the sample covariance is:

$$s_{xy} = \frac{1}{n-1}\sum_{i}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

Notes on Covariance

- the size of x and y must be the same.

- $s_{xy} = s_{yx}$

Correlation Coefficient

- Correlation Coefficient is the normalized covariance

$$C_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$-1 < C_{xy} < 1$$

- (In ME 2550 we gave the correlation coefficient the symbol R)

Correlation Coefficient

$$R = \frac{\sum_{i=1}^{n}(y_i - \bar{y})\,(x_i - \bar{x})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$= \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\Sigma (x_i - \bar{x})^2 \, \Sigma (y_i - \bar{y})^2}}$$

$$s_{xy} = \frac{1}{n-1} \Sigma (x_i - \bar{x})(y_i - \bar{y})$$

$$s_x = \sqrt{\frac{1}{n-1} \Sigma (x_i - \bar{x})^2}$$

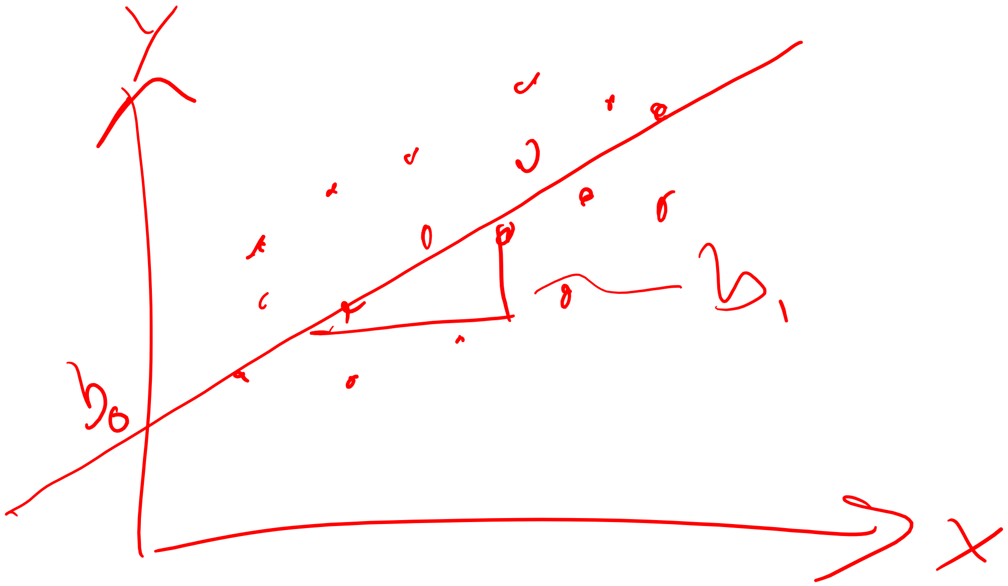
$$s_y = \quad ''$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

$$= \frac{\cancel{\frac{1}{n-1}} \Sigma (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\cancel{\frac{1}{n-1}} \Sigma (x_i - \bar{x})^2} \sqrt{\cancel{\frac{1}{n-1}} (y_i - \bar{y})^2}}$$

Correlation Coefficient

Linear Regression equation



$$y = b_0 + b_1 x$$

$$b_1 = \frac{S_y}{S_x} C_{xy}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Useful Excel and Matlab Functions

- Excel
  - average() -- mean
  - var.s() – variance of a sample
  - stdev.s() -- standard deviation of a sample
  - covariance.s() - covariance of a sample
  - correl() – correlation coefficient
- Matlab
  - mean() – mean
  - var() – variance
  - std() – standard deviation
  - cov(x,y) – covariance matrix in which diagonals are s_x^2 and s_y^2, off-diagonals are s_{xy}.
  - corrcoef(x,y) = matrix in which the off-diagonals are the correlation coefficient.

Department of MECHANICAL ENGINEERING
THE UNIVERSITY OF UTAH

Example 2.2

- You get a set of cylindrical shafts from a supplier. They give you the diameters and a measure of the surface roughness.
- Calculate the mean and standard deviation for each parameter?
- Are the parameters correlated?

D [mm]	1.022	0.998	0.988	0.989	0.996	1.009	1.007	1.000	1.002
S [um]	0.0972	0.0999	0.0996	0.0936	0.0985	0.0995	0.0984	0.1019	0.0992

Linear Combinations of Random Variables

- If we linearly combine two or more Random Variables, what is the mean and variance of the result?
- Consider the random variables X, Y, and Z

$$Z = aX + bY$$

- What is the μ_Z

$$\mu_z = a\mu_x + b\mu_y$$

Department of MECHANICAL ENGINEERING THE UNIVERSITY OF UTAH

Linear Combinations of Random Variables

- What is the standard deviation of Z (σ_z)?

$$\sigma_z^2 = a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\sigma_{xy}$$

$$\sigma_z = \sqrt{a^2\sigma_x^2 + b^2\sigma_y^2 + 2ab\sigma_{xy}}$$

If X & Y are independent $(\sigma_{xy} = 0)$

$$\sigma_z = \sqrt{a^2\sigma_x^2 + b^2\sigma_y^2}$$

Linear Combinations of Random Variables

- What if we have more than two variables?

$$Z = aX_1 + bX_2 + cX_3 \;\ldots$$

$$\mu_z = a\mu_{x_1} + b\mu_{x_2} + c\mu_{x_3}$$

$$\sigma_z^2 = a^2\sigma_{x_1}^2 + b^2\sigma_{x_2}^2 + c^2\sigma_{x_3}^2 + 2ab\,\sigma_{x_1x_2} + 2ac\,\sigma_{x_1x_3} + 2bc\,\sigma_{x_2x_3} \ldots$$

Types of Distributions

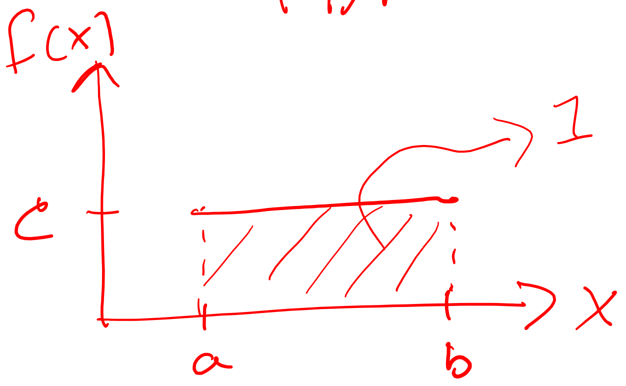
Uniform, Normal, Weibull

Department of
MECHANICAL ENGINEERING
THE UNIVERSITY OF UTAH

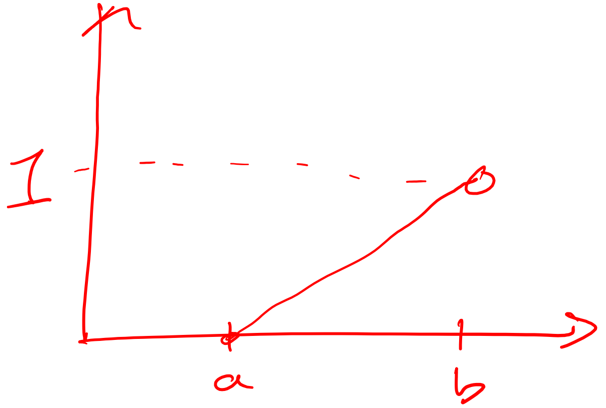
Uniform Distribution

- Every possible outcome of a random experiment is equally likely to occur

PDF



$F(x)$ CDF $\left(F(x) = \int f(x)\,dx\right)$



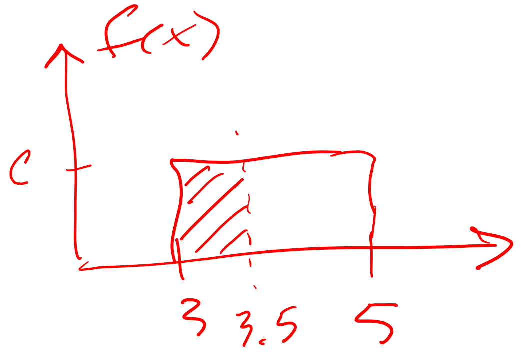
$$F(x) = \int_a^x c\,dx = c\left[x\right]_a^x = c(x-a)$$

$$F(b) = c(b-a) = 1$$

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

Example 2.3.1

The probability that machine that needs service is uniformly distributed between 3-5 years. What is the probability that the machine needs repair before 3.5 years?



$$\int_3^{3.5} f(x)\,dx = F(3.5)$$

$$c = \frac{1}{b-a} = \frac{1}{5-3} = \frac{1}{2}$$

$$F(x) = c(x-a) \rightarrow x = 3.5$$

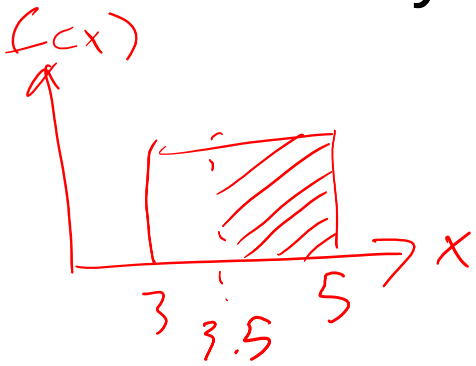
$$F(3.5) = \frac{1}{2}(3.5-3) = \frac{1}{2}\,\frac{1}{2} = \frac{1}{4}$$

25% probability machine needs repair

Department of MECHANICAL ENGINEERING THE UNIVERSITY OF UTAH

Example 2.3.2

What is the probability that the machine does NOT needs repair before 3.5 years?



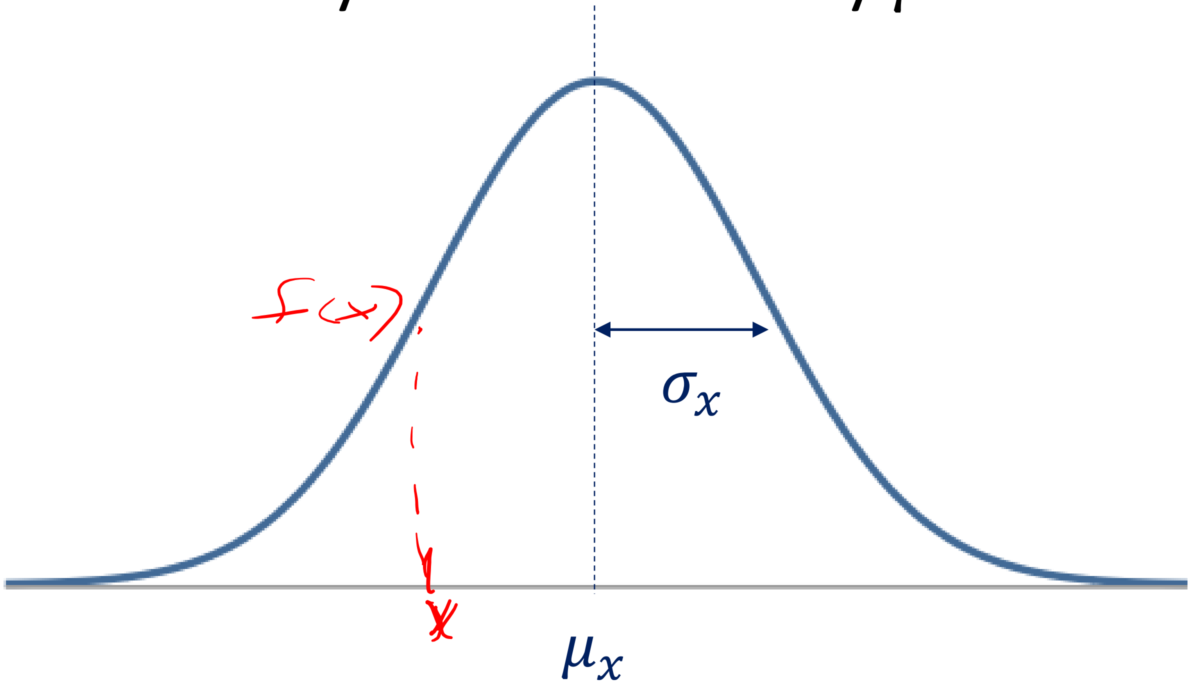
$$1 - F(3.5) = 1 - 0.25$$

$$\boxed{P(X > 3.5) = 0.75}$$

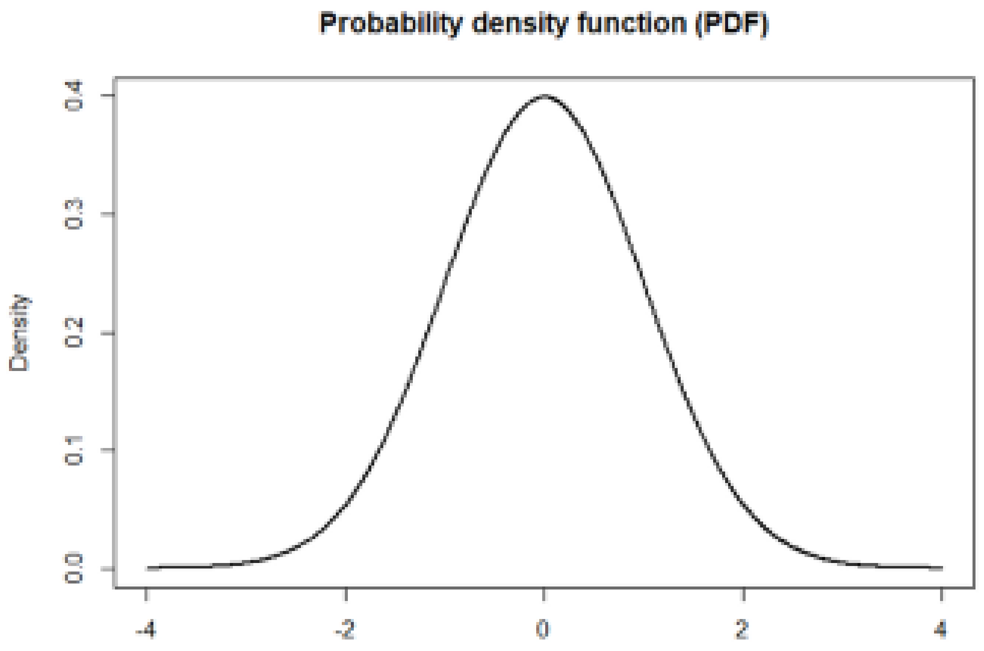
Department of
MECHANICAL ENGINEERING
THE UNIVERSITY OF UTAH

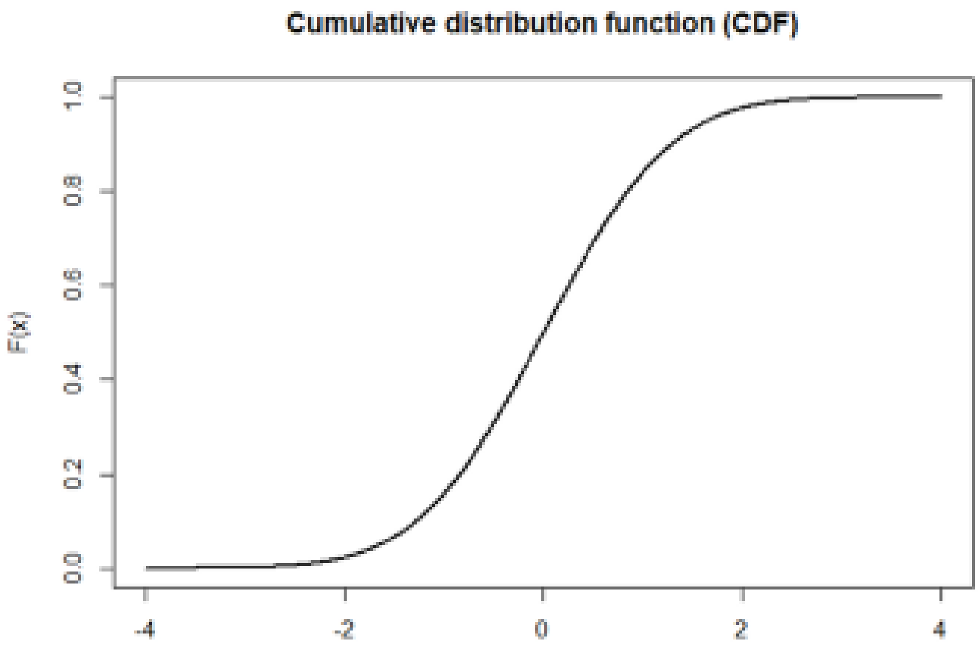
Normal (or Gaussian) Distribution

- Most common
- Many natural effects follow Normal distribution
- Denoted as $N(\mu, \sigma)$ → Normal with mean μ and standard deviation σ
- Normal distribution is fully characterized by μ and σ



Normal (or Gaussian) Distribution





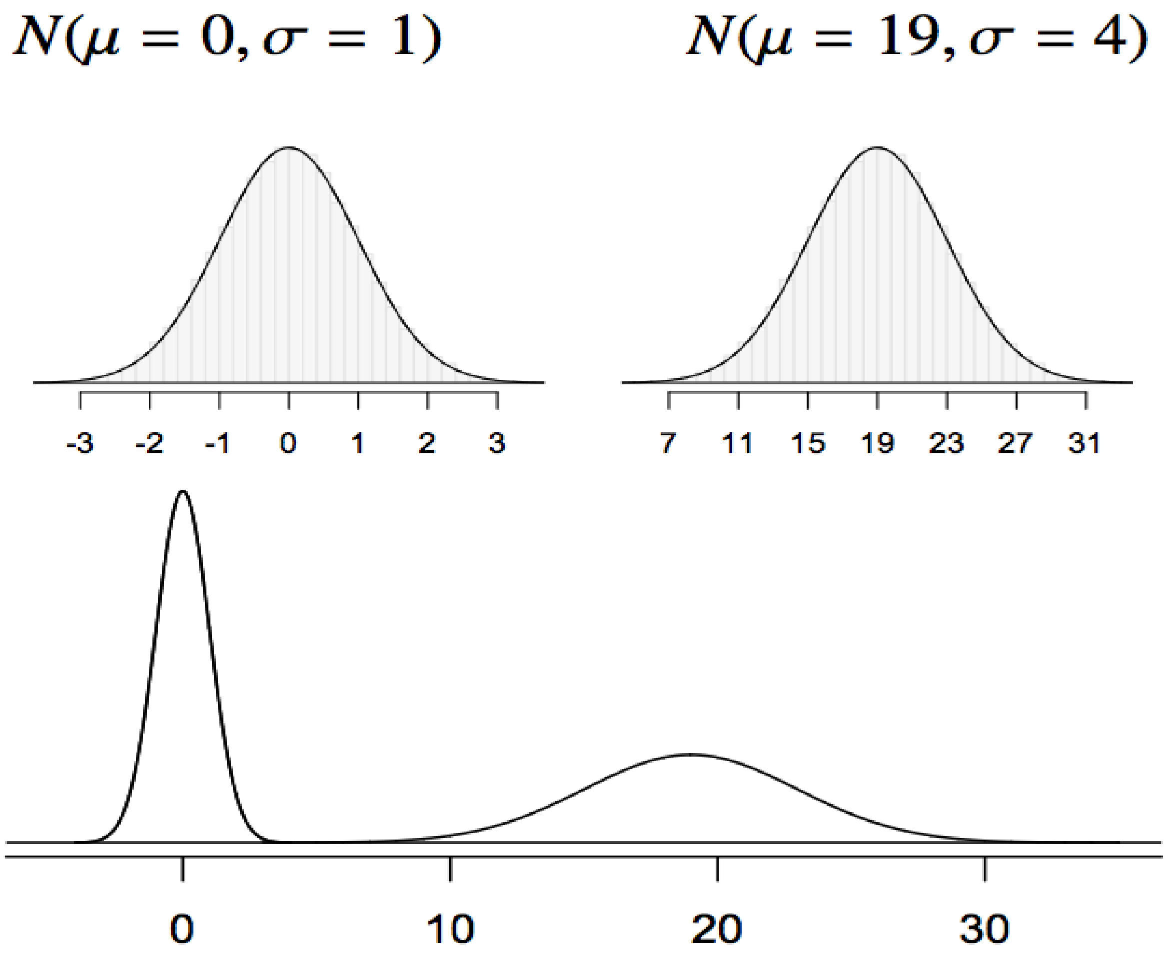
$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Does not have a simple closed form.

Normal Distribution with Different Parameters

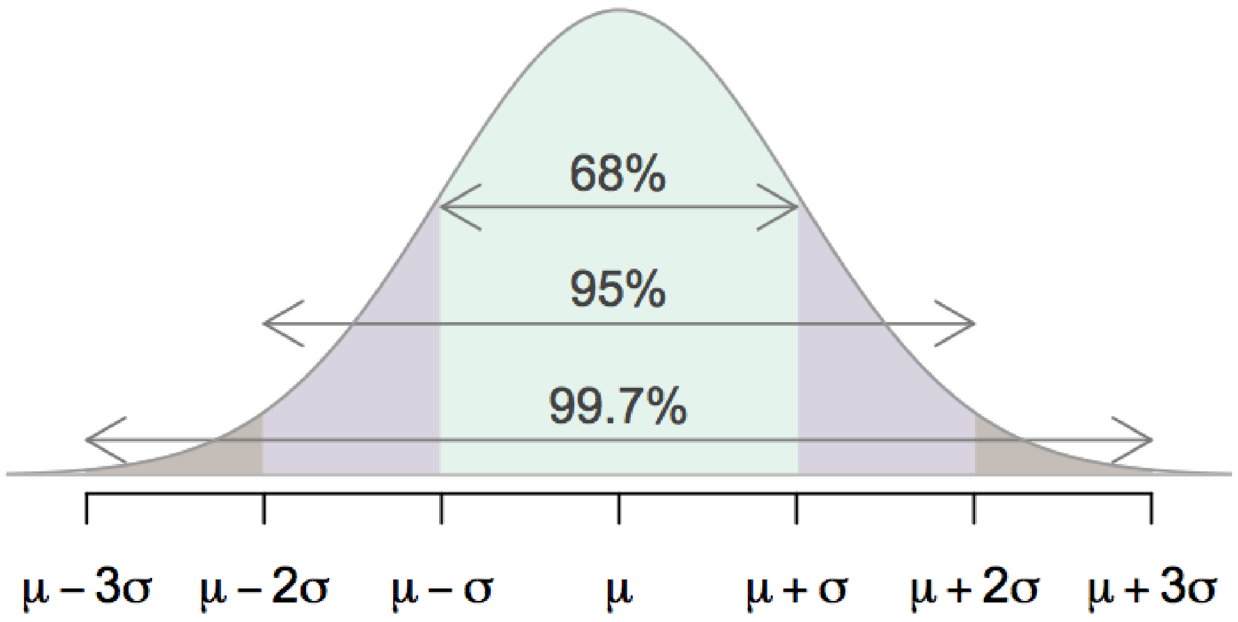
μ: mean, σ: standard deviation



68-95-99.7 Rule

For nearly normally distributed data,

- about 68% falls within 1 SD of the mean,
- about 95% falls within 2 SD of the mean,
- about 99.7% falls within 3 SD of the mean.



Standard Normal Distribution

$$N(\mu, \sigma) \ \rightarrow \ N(0,1)$$

- Can transform any normal distribution to the standard normal distribution by transforming $\mathrm{X} \rightarrow Z$ where Z is the Standard Normal Random Variable

$$Z = \frac{X - \mu_x}{\sigma_x}$$

- So, Z is the number of standard deviations from the mean

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(z)^2}{2}}$$ PDF

- There are pre-calculated tables for the CDF $F(z)$

Ex 2.4.1

1.4

z

z = 2.17

Ex 2.4.2

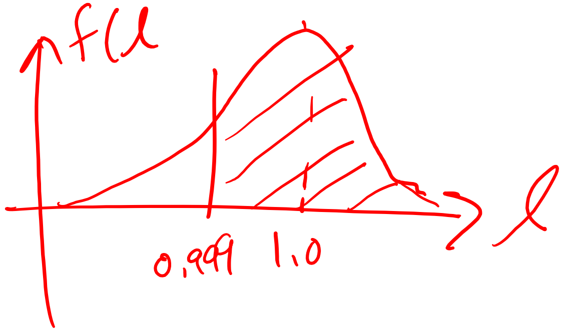
z-value

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

F(z)

Example 2.4.1

- Consider a steel tube cut to a length of 1m. The length L is a random variable (normally distributed) with $\mu_L = 1m$, $\sigma_L = 2mm$.
- What is the probability that the length (l) of a randomly selected sample will exceed 0.999 m?



$P(L > 0.999)$

Transform $L \to Z$

$$z = \frac{l - \mu_L}{\sigma_L} = \frac{0.999 - 1}{0.002} = -0.5$$

$$P(Z > -0.5) = P(Z < 0.5) = F(z)$$

$$\boxed{F(z) = 0.6915} \to P(L > 0.999)$$

Example 2.4.1

Example 2.4.2

- What standard deviation must be achieved to ensure that 98.5% of all processed tubes are shorter than 1.022m?

$$z = \frac{\ell - \mu_L}{\sigma_L}$$

know ℓ, know μ_L,

Find z for $P(Z_1 < z) = 0.985$

$$F(z) = 0.985$$

$$z = 2.17$$

$$\sigma_L = \frac{\ell - \mu_L}{z} = \frac{1.022 - 1}{2.17} = 0.00092\,m$$

$$\boxed{\sigma_L = 0.92\,mm}$$

Department of
MECHANICAL ENGINEERING
THE UNIVERSITY OF UTAH

Useful Excel and Matlab Functions for Normal Distributions

- Excel
  - norm.dist(x,mean,stdev,cum) -- Returns the cumulative distribution function (CDF) if cum = 1. That is the area under the normal distribution curve up to x.
  - norm.inv(probability, mean, stdev) – Returns the value (i.e x) associated with "probability" for the inverse normal CDF with specified mean and standard deviation.
  - standardize(x,mean,stdev) – Returns the z value associated with x.
- Matlab
  - normpdf(x,mu,sigma) – Returns the value of the PDF of the normal distribution.
  - normcdf(x,mu,sigma) – Returns the CDF of the normal distribution. That is the area under the normal distribution curve up to x.
  - norminv(p,mu,sigma) – Returns the value (i.e., x) associated with the probability p for the inverse normal CDF with specified shape and scale parameters.
  - standardize(x,mu,sigma) – Returns the z value associated with x.

Department of MECHANICAL ENGINEERING
THE UNIVERSITY OF UTAH

Weibull Distribution

- Often used to describe life and durability of mechanical elements (i.e., time to failure)
- Very flexible distribution
- PDF of Weibull distribution is dependent on 3 parameters
  - x_0: minimum expected value for the random variable X
  - λ: scale parameter
  - k: Weibull slope (shape parameter) → failure rate
    - ~~$k < 1$: failure rate ↓ over time~~
    - ~~$k = 1$: failure rate is constant~~
    - ~~$k > 1$: failure rate ↑ over time~~

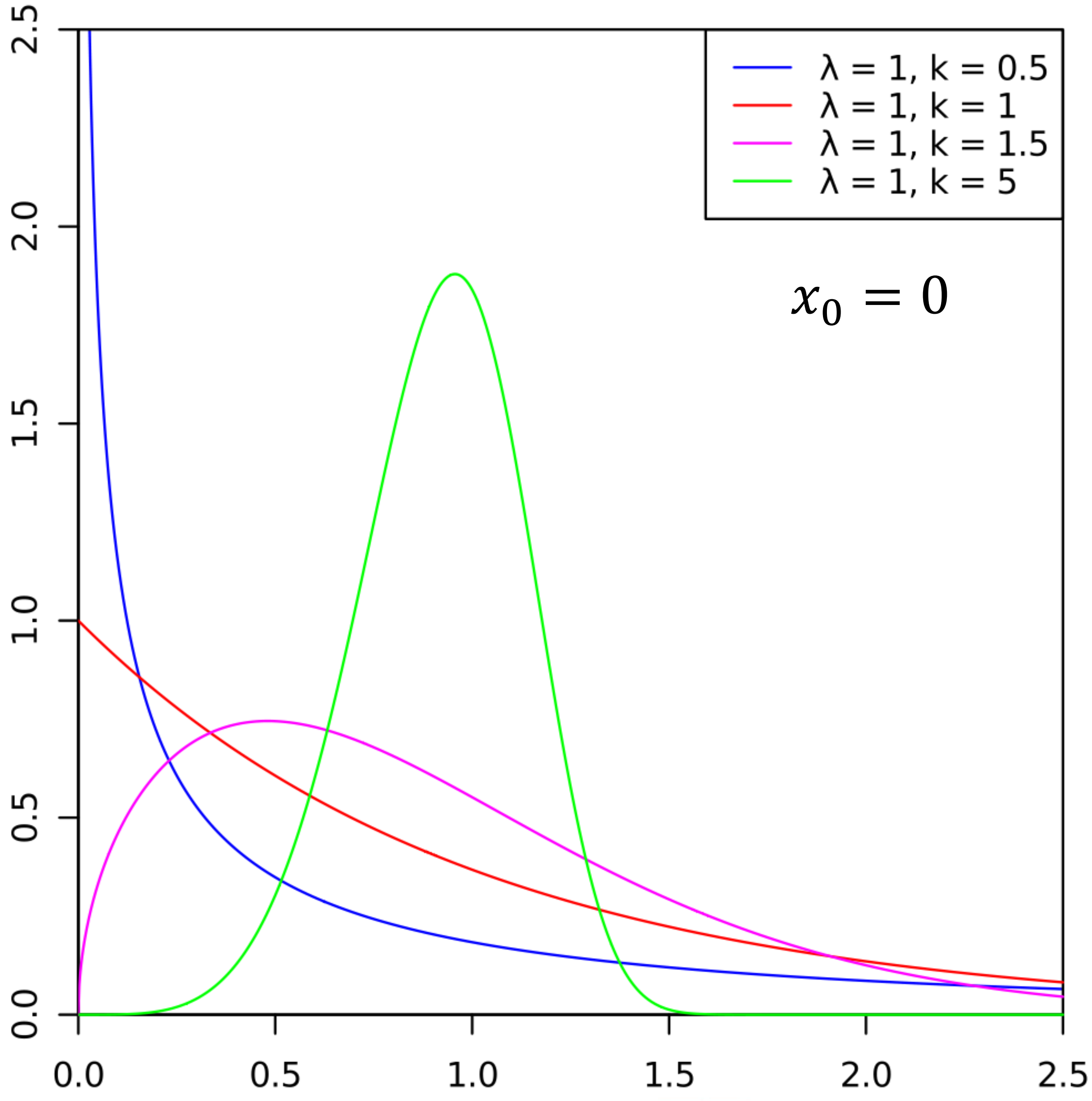
→ Ignore this

Weibull Distribution

$$f(x) = \frac{k}{\lambda}\left(\frac{x - x_0}{\lambda}\right)^{k-1} e^{-\left(\frac{x-x_0}{\lambda}\right)^k}$$

$$F(x) = \int_{x_0}^{x} f(x)dx = 1 - e^{-\left(\frac{x-x_0}{\lambda}\right)^k}$$

Important note, these equations are written slightly differently in both the Raeymaekers book and Shigley's. I am almost certain that their form is incorrect. Use this form.



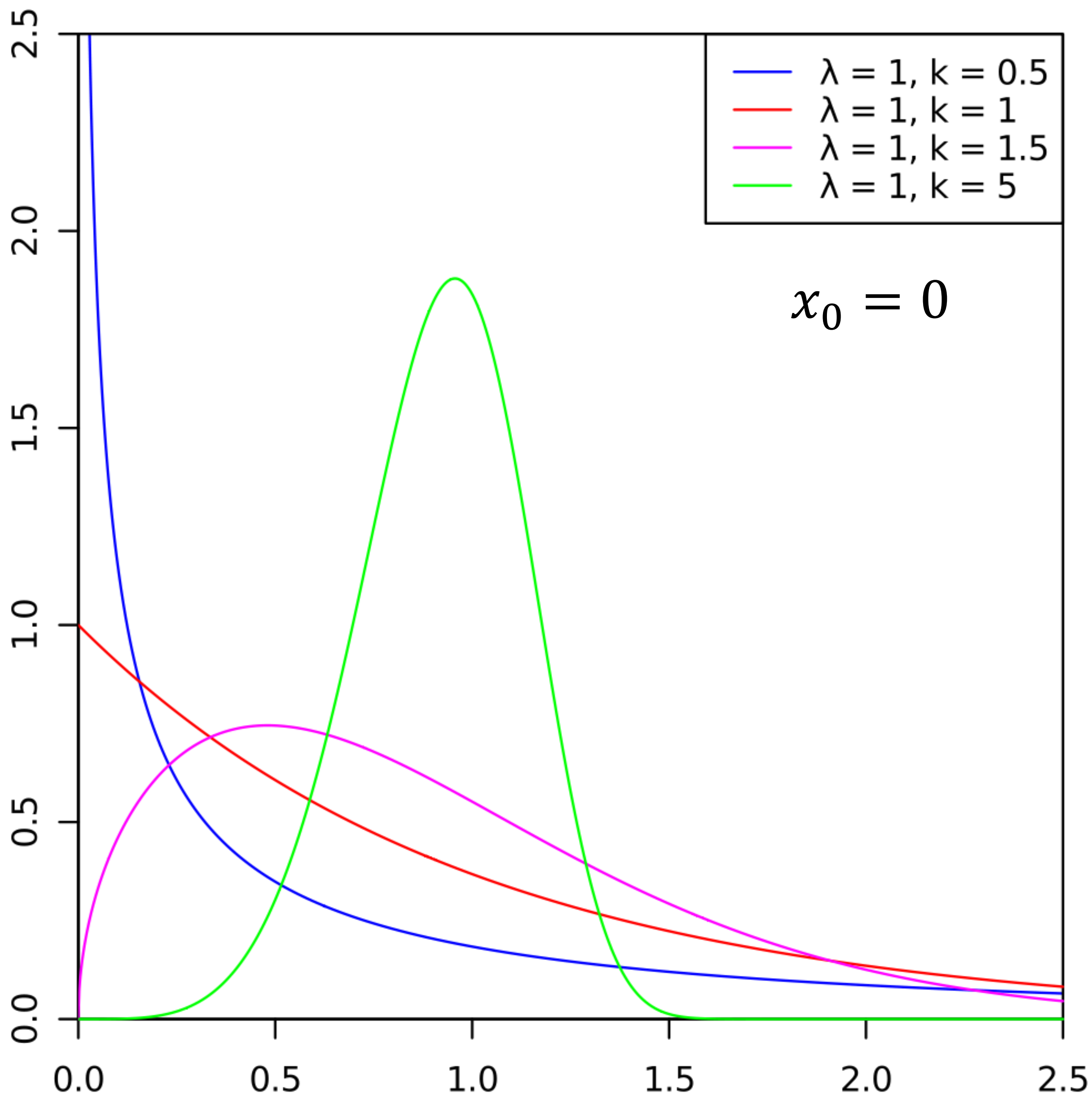
Department of MECHANICAL ENGINEERING THE UNIVERSITY OF UTAH

Weibull Distribution

If $x_0 = 0$, the equations reduce to

$$f(x) = \frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$$

$$F(x) = \int_{x_0}^{x} f(x)dx = 1 - e^{-\left(\frac{x}{\lambda}\right)^k}$$

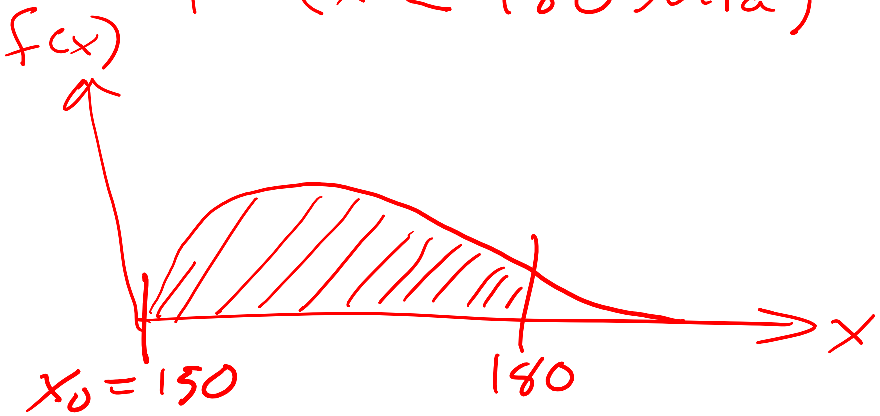


Example 2.5.1

- You are ordering steel cylinders from vendor. The vendor says that the ultimate strength of the material follows a Weibull distribution with parameters $x_0 = 150\ MPa, \lambda = 250\ MPa, \mathrm{k} = 3$.
- You estimate that in use the cylinders will experience a maximum stress of 45 MPa. Your design factor is 4. What percentage of cylinders will not meet your design factor?

$$\sigma = n\sigma_{max} = 4 \cdot 45\ MPa = 180\ MPa$$

$$P(x < 180\ MPa) \rightarrow F(180)$$



Example 2.5.1

$$F(180) = 1 - e^{-\left(\frac{x - x_0}{\theta}\right)^k} = 1 - e^{-\left(\frac{180-150}{250}\right)^3}$$

$$F(180) = P(x < 180) = \boxed{0.0017}$$

We expect 0.17% of parts to have a design factor below 4.

Department of MECHANICAL ENGINEERING
THE UNIVERSITY OF UTAH

f(x)
k ↓
λ ↑
x
x₀

Useful Excel and Matlab Functions for Weibull Distributions

- Excel
  - weibull.dist(x,alpha,beta,cumulative) – Returns evaluation of Weibull distribution function at x, with alpha scale parameter and beta shape parameter
  - Excel does not have an inverse Weibull function.
- Matlab
  - wblpdf(X,A,B) – Returns the value of the PDF of the Weibull distribution.
  - wblcdf(X,A,B) – Returns the CDF of the Weibull distribution. That is the area under the Weibull distribution curve up to X.
  - wblinv(P,A,B) – Returns the value (i.e., X) associated with the probability P for the inverse Weibull CDF with specified mean and standard deviation.
  - wlbrnd(A,B) – Returns a random value selected from a Weibull distribution with scale parameter A and shape parameter B.
  - wblstat(A,B) - Returns the mean and variance of the Weibull distribution with scale parameter A and shape parameter B.

Summary

- Random variable → variable whose value depends on the outcome of a random event (or experiment)
  - A random variable will be described by some distribution
- Metrics to describe distribution → mean, variance, standard deviation, covariance and correlation
  - Linear combinations of random variables → useful for tolerance stackups
- Common distributions for this class → uniform, Normal, Weibull

Department of MECHANICAL ENGINEERING
THE UNIVERSITY OF UTAH