Oregon State University

Design Under Uncertainty

ME 615 Spring 2020

Dr. Chris Hoyle

MIME

Normalization and Standardization

- **Normalization** of x_i to a 0-1 scale (z_i):
  - $z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$
  - You can "undo" the normalization by solving for x_i
- **Standardization** of x_i to a $\mu=0$, $\sigma=1$ scale (z_i):
  - $z_i = \frac{x_i - \mu}{\sigma}$
  - You can "undo" the standardization by solving for x_i

Frequently Used Distribution Expressions

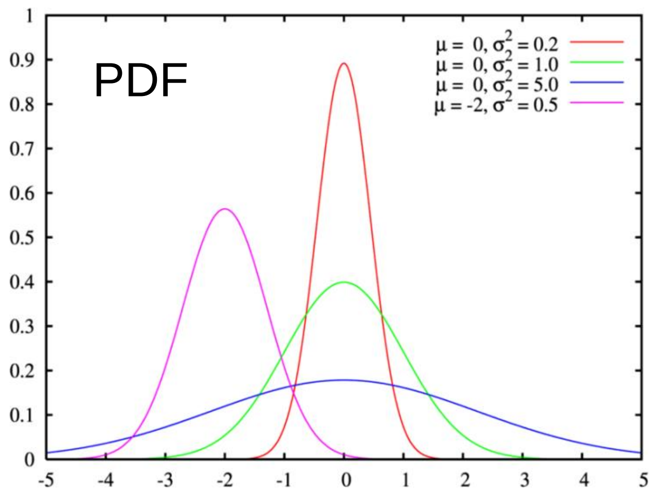
Oregon State University College of Engineering

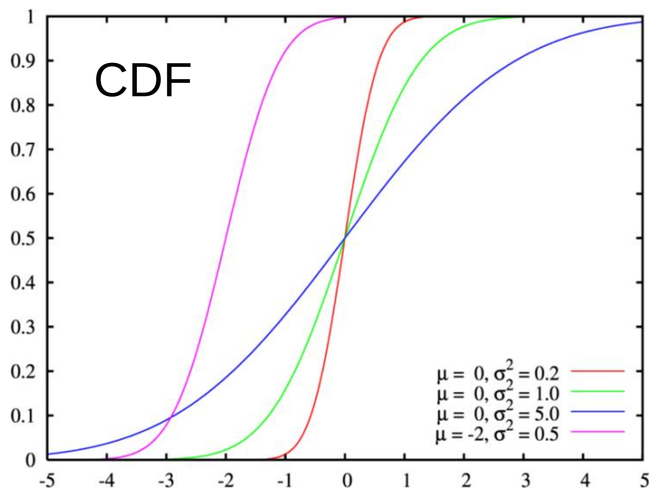
- Normal (Gaussian) distribution

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right],\quad -\infty<x<\infty,\ -\infty<\mu<\infty,\ \sigma>0$$

- Standard normal distribution

$$\phi(z)=\frac{1}{\sqrt{2\pi}}\exp\left[-\frac{z^2}{2}\right],\quad -\infty<z<\infty,\ \mu=0,\ \sigma=1 \qquad z=\frac{X-\mu}{\sigma}$$



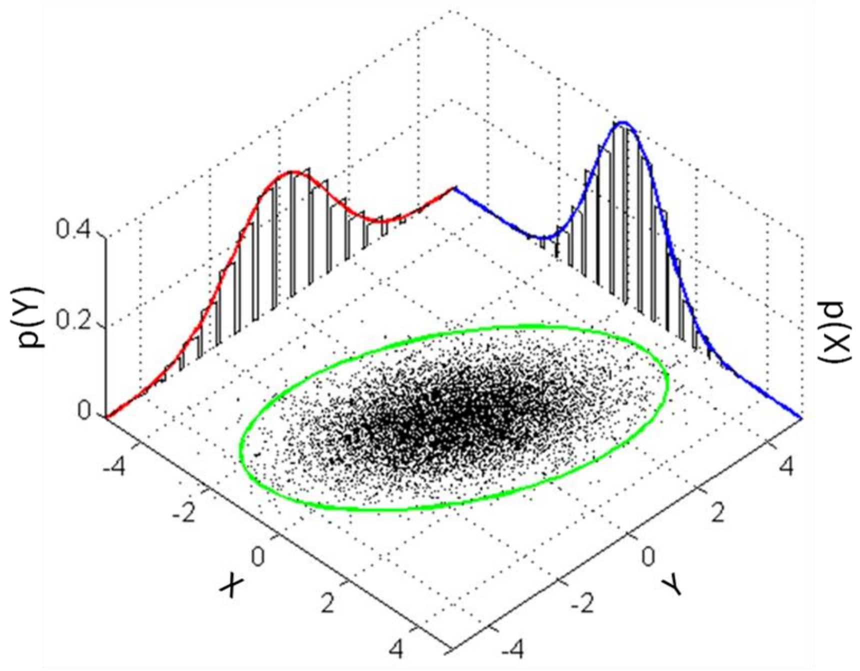


Frequently Used Distribution Expressions

- Mulitvariate Normal distribution

$$f_{\mathbf{X}}(x_1, \dots, x_k) = \frac{\exp(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}})}{\sqrt{(2\pi)|\boldsymbol{\Sigma}|}}$$

PDF



Frequently Used Distributions

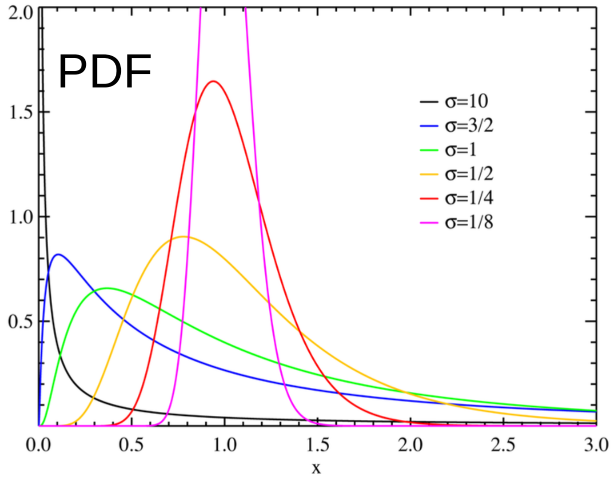
- Lognormal distribution
  - Can be derived from a normal distribution $N(\mu,\sigma)$

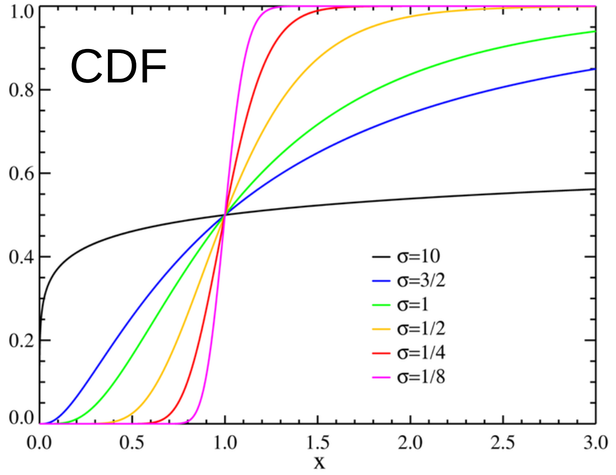
$$f(x;\mu,\sigma)=\frac{1}{\sigma x\sqrt{2\pi}}\exp\left[-\frac{1}{2\sigma^2}(\log x-\mu)^2\right],\quad x\geq 0,\ -\infty<\mu<\infty,\ \sigma>0$$

  - Mean and variance

Mean and variance of normal distribution

$$\mu_X=\exp\left(\mu+\frac{1}{2}\sigma^2\right),\quad \sigma_X^2=\exp\left(\sigma^2-1\right)\exp\left(2\mu^2+1\right)$$



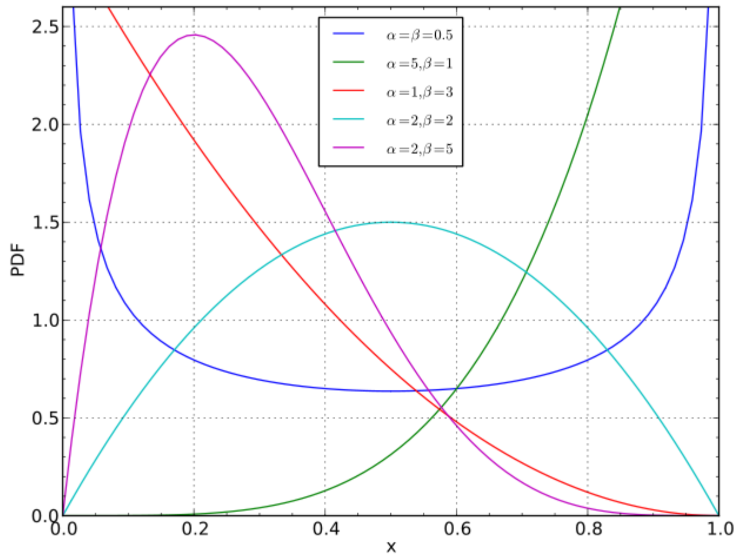


Frequently Used Distributions

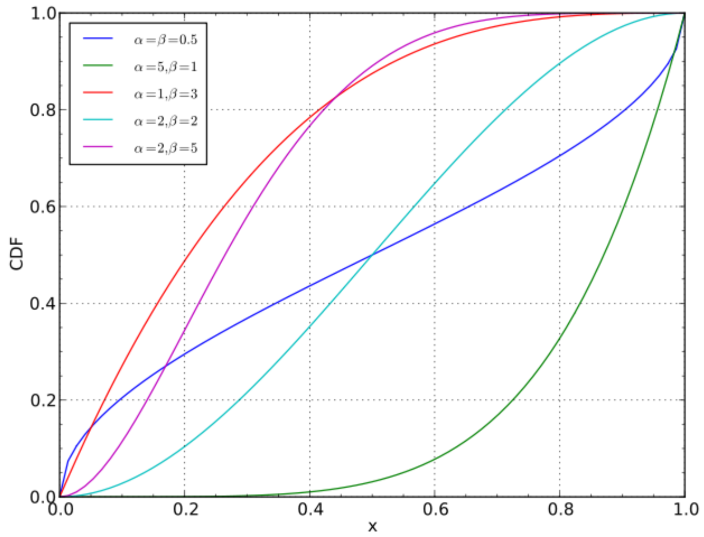
Oregon State University
College of Engineering

- Beta Distribution
  - $f(x;\ \alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}$ ***You must normalize x***
  - Mean and Variance
    - $\mu_X = \frac{\alpha}{\alpha+\beta}, \quad \sigma^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

PDF



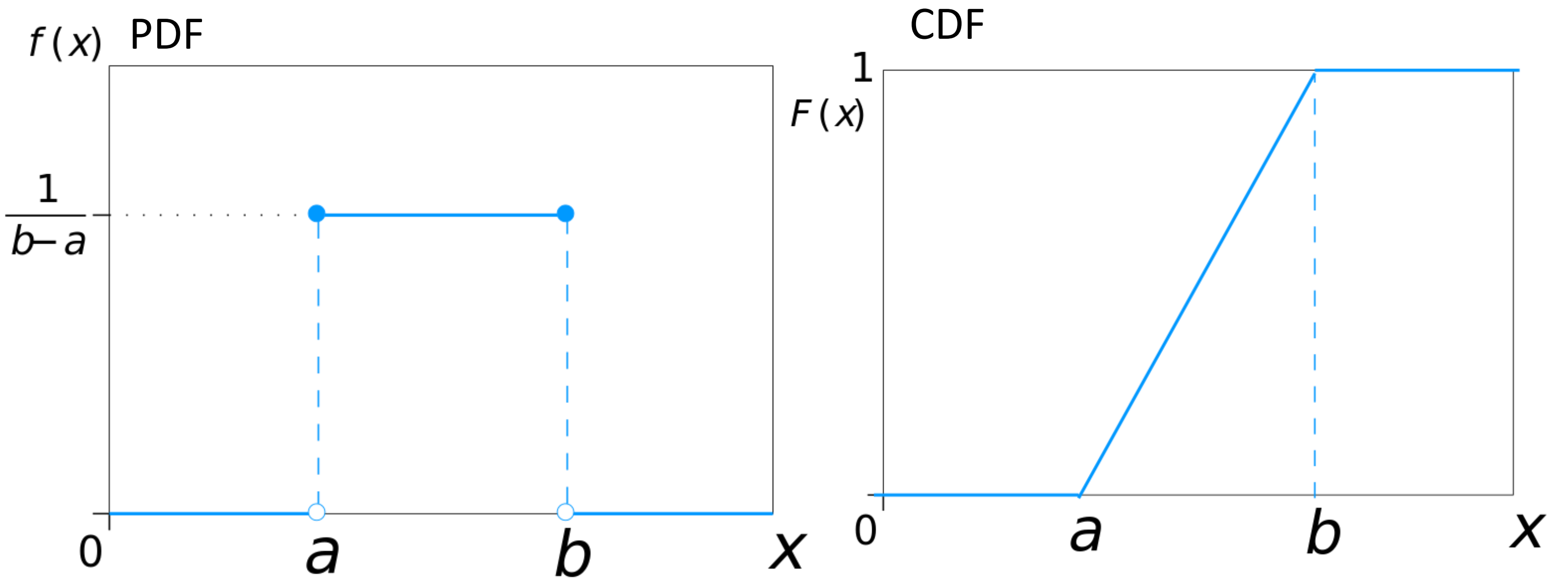
CDF



Frequently Used Distributions

Oregon State University
College of Engineering

- Uniform Distribution
  - $f(x; a, b) = \frac{1}{b-a} \quad a \leq x \leq b$
  - Mean and Variance
    - $\mu_X = \frac{1}{2}(a+b), \;\; \sigma^2 = \frac{1}{12}(b-a)^2$



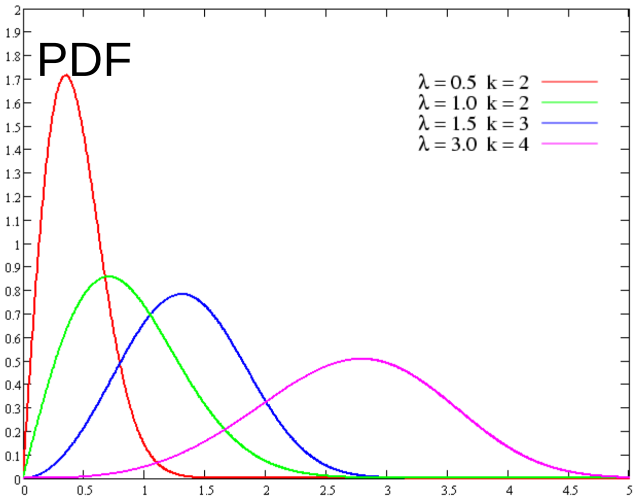
Frequently Used Distributions

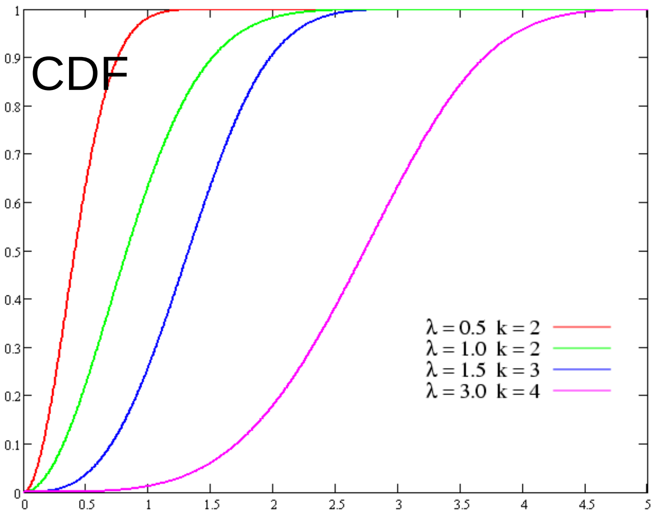
- Weibull distribution

$$f\left(x;k,\lambda\right)=\frac{k}{\lambda}\left(\frac{x}{\lambda}\right)^{k-1}\exp\left[-\left(\frac{x}{\lambda}\right)^{k}\right]\quad x>0,\ \lambda>0,\ k>0$$

  - Mean and variance

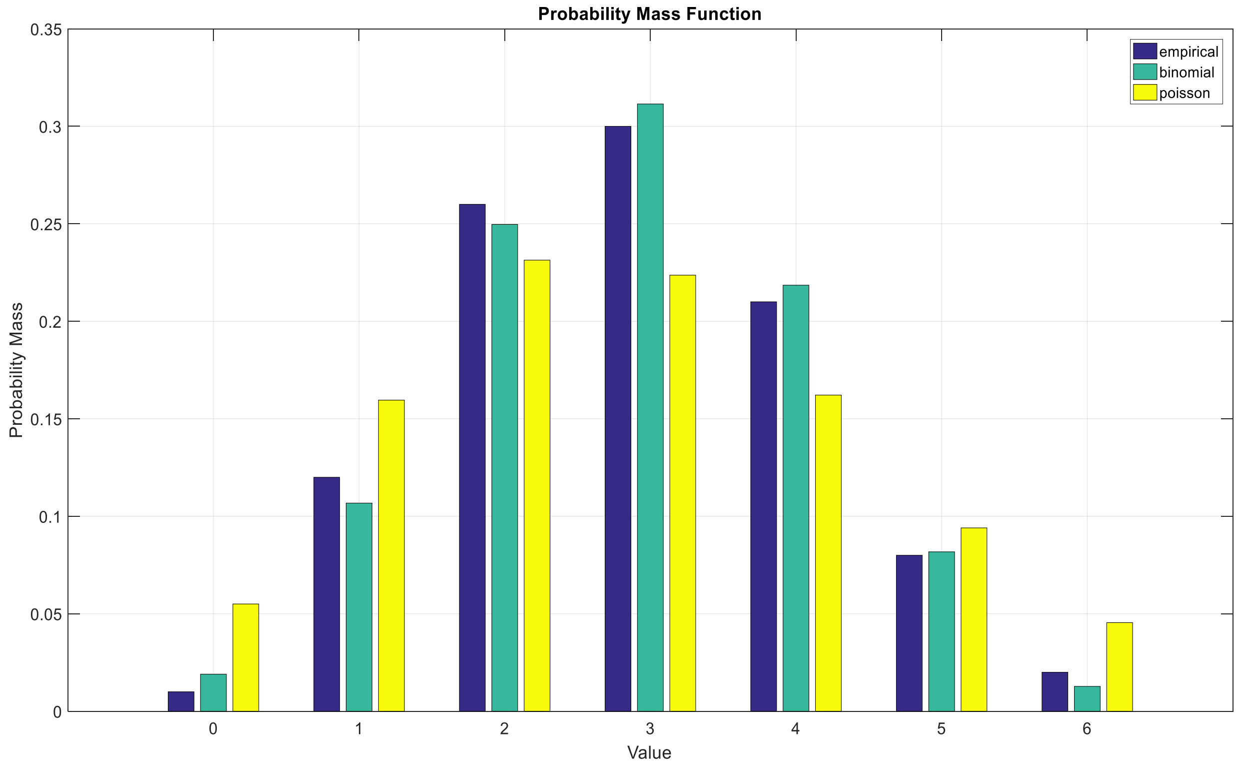
$$\mu_X=\lambda\Gamma\left(1+\frac{1}{k}\right),\ \sigma_X^2=\lambda\left\{\Gamma\left(\frac{2}{k}+1\right)+\left[\Gamma\left(\frac{1}{k}+1\right)\right]^2\right\}$$





Discrete Distributions

- Common Discrete Distributions (i.e. for count data) are:
  - Binomial: the number of successes in a sequence of *n* independent experiments
  - Negative Binomial: number of successes in a sequence of independent experiments before a specified number of failures, *r*, occurs.
  - Poisson: number of occurrences in a fixed interval of time.



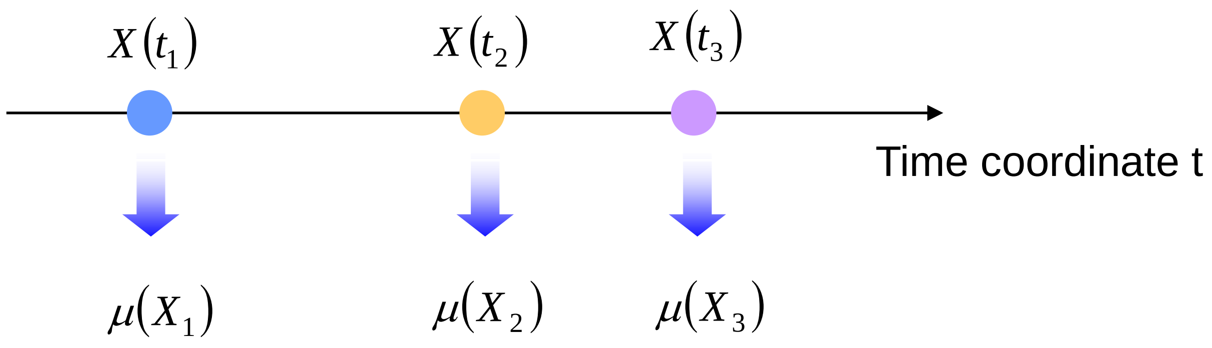
Random Process

- A stochastic process is defined as a collection of random variables defined on a common probability space, where Ω is a sample space, F is a σ -algebra, and *P* is a probability measure, and the random variables, indexed by some set *T* (which can be discrete or continuous)

$$\{X(t) : t \in T\}$$

- More generally we can thing of it as a function of two variables, $t \in T$ and $\omega \in \Omega$

$$\{X(t, \omega) : t \in T\}$$

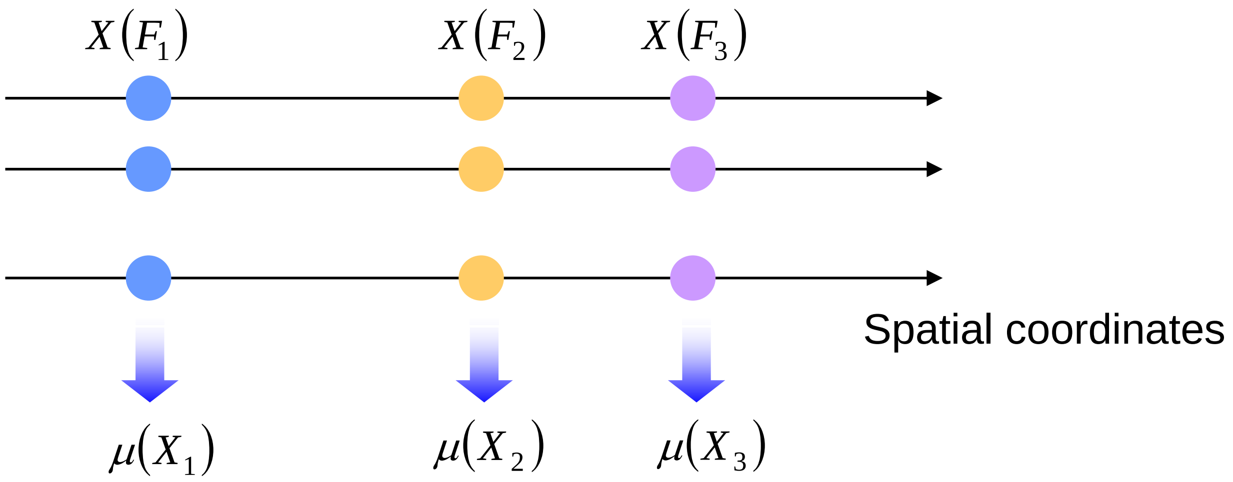


Expected value

$$\mu(x) = E[H(x)]$$

Random field

- A **random [field](#) *F*** is a generalization of a [stochastic process](#) such that the underlying parameter need no longer be a simple [real](#) or integer valued "time", but can instead take values that are multidimensional [vectors](#).
  - $\{F_t : t \in T\}$
- For example if $F = [y, t] \in space - time$ $\{X(F, \omega) = X(y, t, \omega) : t \in T\}$



Expected value

$$\mu(x) = E[H(x)]$$

autocovariance
$$\mathrm{cov}(x_1, x_2) = E[(H(x_1) - \mu(x_1))(H(x_2) - \mu(x_2))]$$

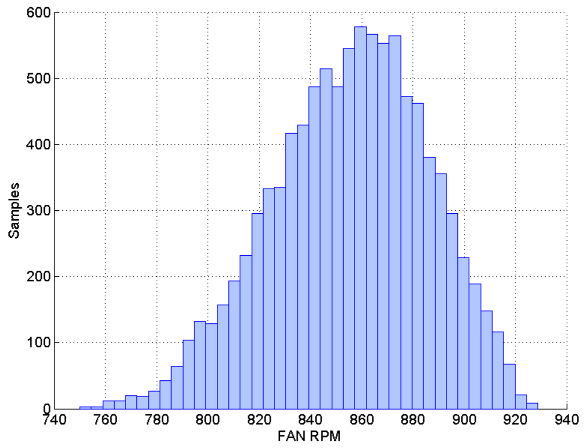
autocorrelation
$$\alpha(x_1, x_2) = E[H(x_1) \cdot H(x_2)]$$

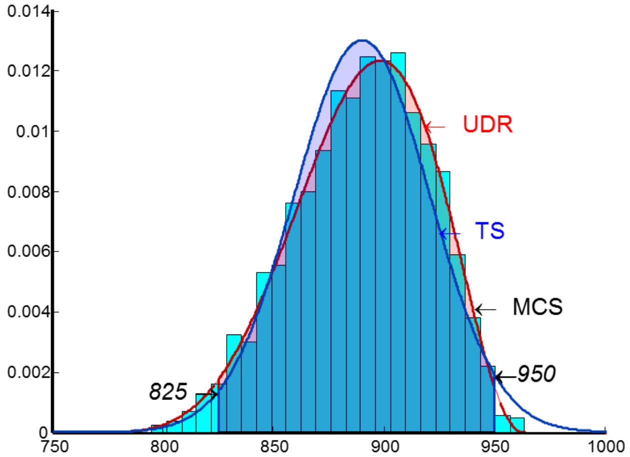
How do you know which distribution to use?

Oregon State University
College of Engineering

- If you have experimentally collected data, you could “fit” a distribution to the data

1. Plot your data using a histogram
2. Look at the wikipedia pages for the parametric distributions we studied.
3. Theorize a few distribution types.
4. Evaluate the fit using a method shown in the next 2 slides.

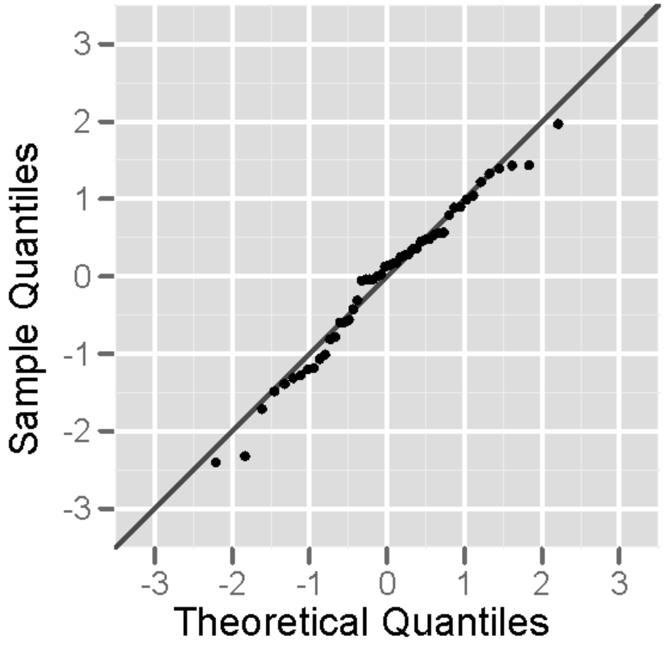




Determination of Distribution and Test

Oregon State University
College of Engineering

- Probability plotting
  - Check linear relationship between 'ordered observation' and their 'sample quantile' (estimated cumulated probability) in specially scaled probability paper
  - Evaluates fit based on overall quality of the fit of the distribution



Determination of Distribution and Test

- *Example – yield stress and tensile strength of carbon steel*

Yield stress

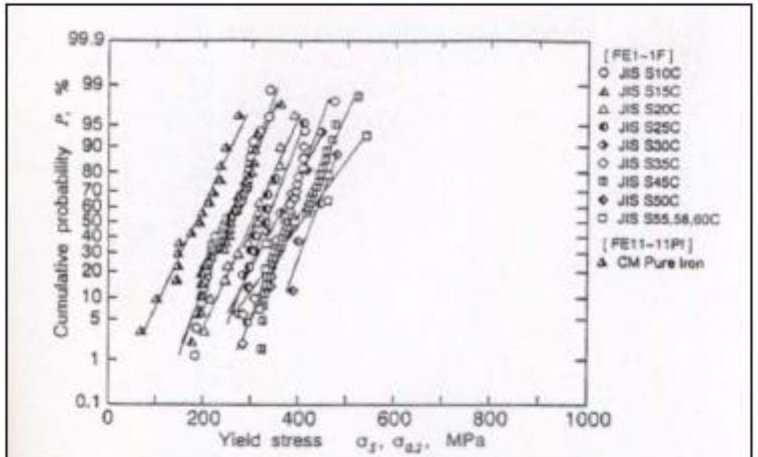


図 2.3.1.a 純鉄および機械構造用炭素鋼の降伏応力または0.2%耐力の分布

Tensile strength

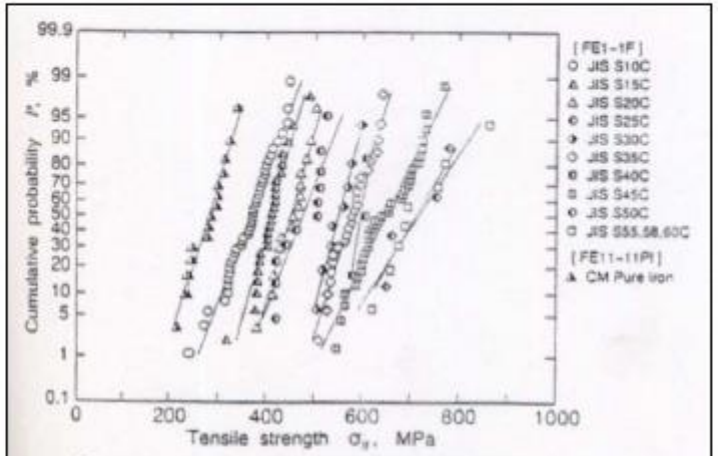


図 2.3.1.b 純鉄および機械構造用炭素鋼の引張強度の分布

Normal probability paper

→ Normal distribution

From Japanese society of material science

Probability Plotting

Probability Plotting

- Use Matlab probplot for visual representation

To do yourself, follow these steps:

1. Sort data from smallest to largest value (Matlab *sort*). This is your x value.
2. Use the following transformation to create the y (z) value for the normal distribution:

$$z_i = \Phi^{-1}\left(\frac{i-a}{n+1-2a}\right)$$

for $i = 1, 2, \ldots, n$, where

$a = 3/8$ if $n \leq 10$ and

0.5 for $n > 10$,

3. Plot x vs. y and use `fitlm` in Matlab to get the linear regression fit.
4. Repeat for other distribution assumptions

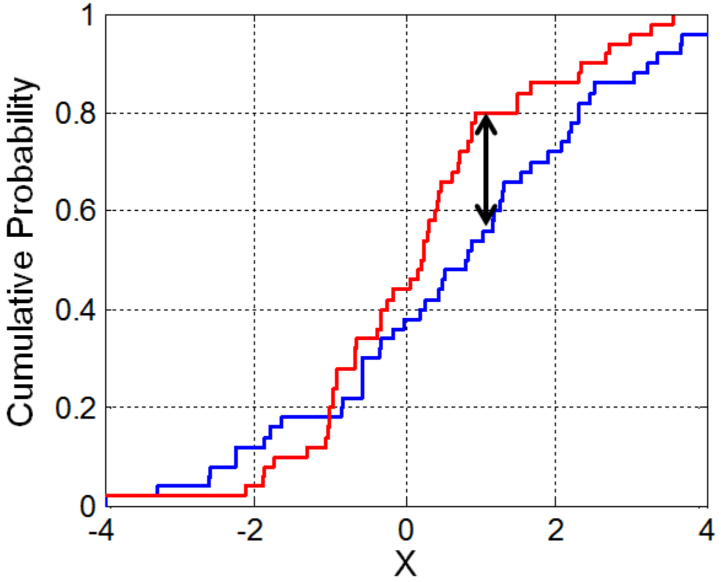
Probability Plotting

Oregon State University
College of Engineering

- To use probability plotting, one would need:
  - Sample data
  - Transformations for all distribution types considered (need to look these up in a book or other source)
- Pick distribution with highest R^2

Determination of Distribution and Test

- Kolmogorov-Smirnov test (K-S test)
  - Compare observed cumulative frequency with CDF of assumed distribution
    - $D_n = \sup_x |F_N(x) - F(x)|$
  - Evaluates based on worst case parts of fit.



How do you know which distribution to use?

Oregon State University College of Engineering

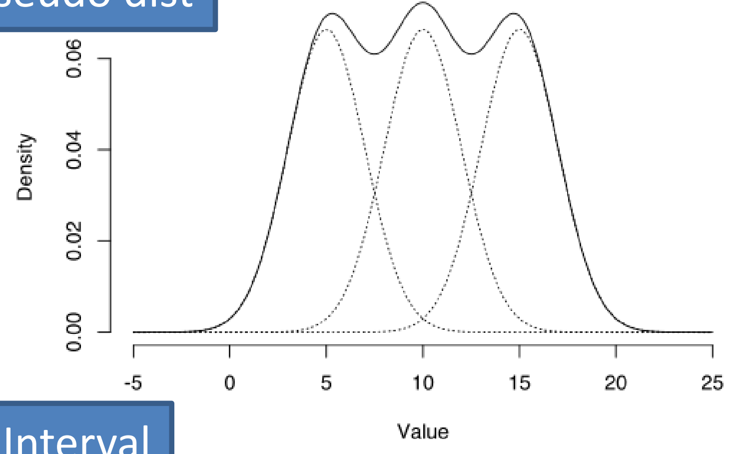
- If you don't have data, you can still estimate a distribution based upon experience or known bounds (*subjective probability*).
- Popular choices
  - Normal Distribution
    - Pros: Symmetric, mathematics are easy
    - Cons: Symmetric, bounded [-∞, ∞]
  - Lognormal Distribution
    - Pros: Never less than 0 (or greater than 0), bounded [0, ∞]
    - Cons: Non symmetric, math gets more difficult
  - Beta Distribution
    - Pros: Bounded on interval [a, b], can be symmetric or non-symmetric
    - Cons: Math gets more difficult
  - Uniform Distribution
    - Pros: Bounded on interval [a, b], Everything is equi-probable, easy math
    - Cons: Everything is equi-probable

Design Under Uncertainty

Oregon State University
College of Engineering

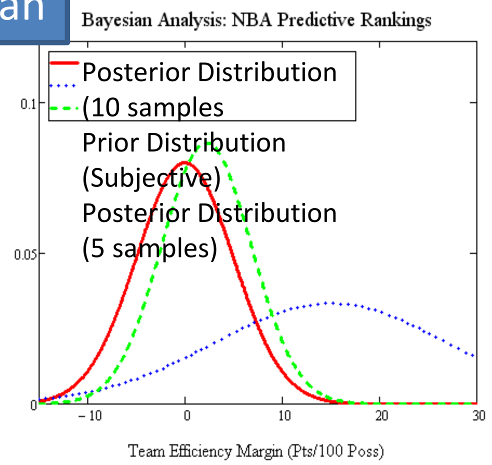
- Works best for aleatory uncertainties
- We can potentially represent epistemic uncertainties:
  - Range of mean and std deviation: “pseudo” distributions
  - Intervals: Interval math
  - Bayesian approach: use prior and posterior distributions

Pseudo dist



Interval

Bayesian



Design Under Uncertainty

Oregon State University
College of Engineering

- Dealing with Ontological uncertainties is an open issue
- Classification of Noise and Control factor uncertainty is useful:
  - Type I Design: Noise factor uncertainty
  - Type II Design: Control factor uncertainty
  - Combination of Type I and Type II