Example : Flaws occur randomly along a thin copper wire. Let y denote the number of flaws in a length of L millimeters of wire and suppose that the average number of flaws in L millimeters is λ.

- Partition the length of wire into n subintervals of small length, say, 1 micrometer each.

- If the length of subintervals is small enough, the probability that more than one flaw occurs in the subinterval is negligible.

- Every subinterval has the same probability of containing a flaw, say $p = \lambda/n$.

- A subinterval contains a flaw is **independent** of other subintervals.

So we can model the distribution of y as approximately a Bin(n,p) random variable with n large and $np = \lambda$.

Approximate Bin(n,p)

with n large and $np = \lambda$

$$\begin{aligned} P(y) &= \binom{n}{y} p^y (1-p)^{n-y} \\ &= \frac{n(n-1)\cdots(n-y+1)}{y!}\Big(\frac{\lambda}{n}\Big)^y\Big(1-\frac{\lambda}{n}\Big)^{n-y} \\ &= \frac{n(n-1)\cdots(n-y+1)}{n^y}\Big(\frac{\lambda^y}{y!}\Big)\Big(1-\frac{\lambda}{n}\Big)^{n-y} \\ &\approx \frac{\lambda^y}{y!}\Big(1-\frac{\lambda}{n}\Big)^n \\ &\approx \frac{\lambda^y}{y!}e^{-\lambda} \end{aligned}$$

where the last approximation comes from

$$\begin{aligned} \log\Big(1-\frac{\lambda}{n}\Big)^n &= n\log\Big(1-\frac{\lambda}{n}\Big) \\ &= n\Big(-\frac{\lambda}{n}-\frac{1}{2}\frac{\lambda^2}{n^2}-\cdots\Big) \approx -\lambda \end{aligned}$$

Poisson Distribution

Consider a particular event occurs during a given unit (or interval) of time, assume that

- the events occur randomly through the interval.
- the probability that an event occurs in a given unit of time is the **same** for all the units.
- the number of events that occur in a given unit of time is **independent** of the number that occur in other units.

Such random experiment is called a **Poisson Process**

Let λ denote the mean or expected number of events in each unit.

The random variable y denotes the number of events occurring during that unit is a **Poisson random variable** with parameter λ

$$p(y) = \frac{\lambda^y e^{-\lambda}}{y!} \quad (y = 0, 1, 2, \ldots)$$

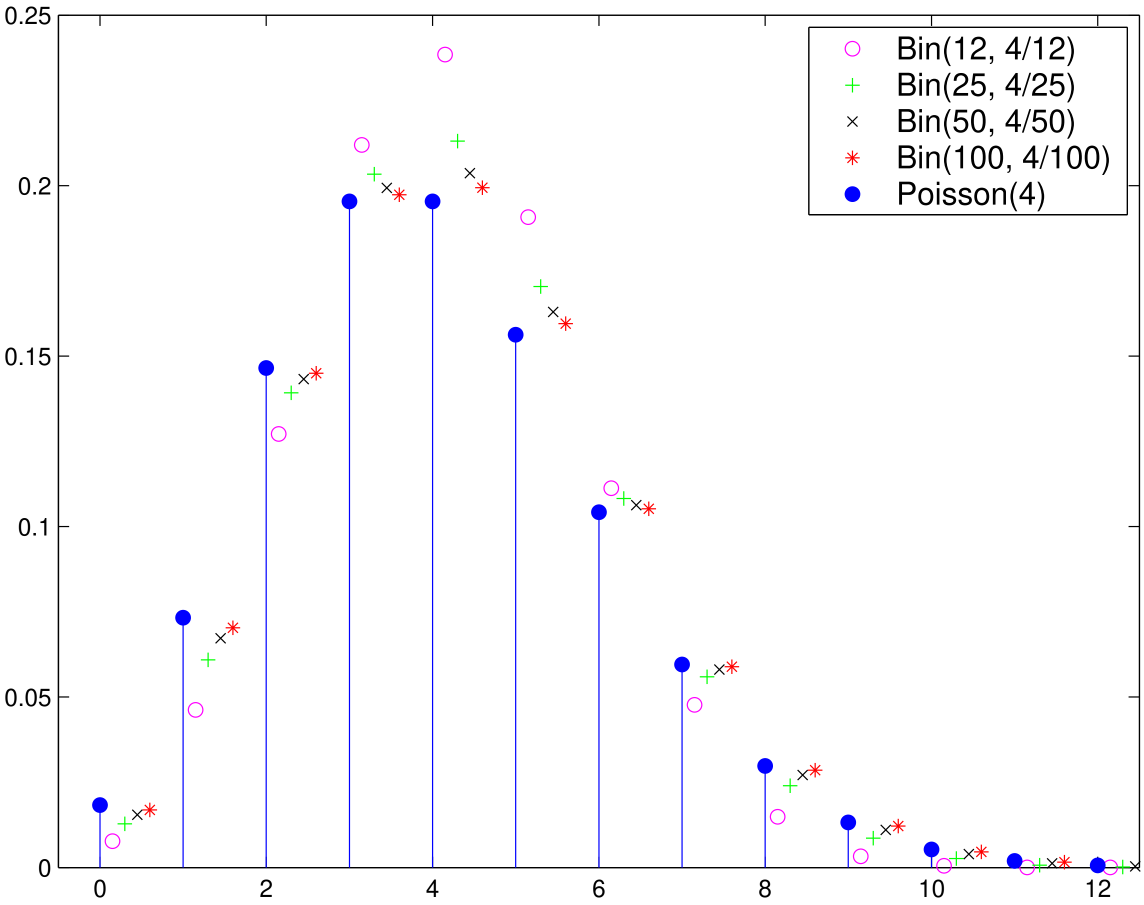
Mean and Variance

$$\mu = \lambda \quad \sigma^2 = \lambda$$

Show that $\mathbb{E}(y) = \lambda$ where y is a Poisson random variable with parameter λ.

$$\begin{aligned}
\mathbb{E}(y) &= \sum_{y=0}^{\infty} y \cdot p(y) \\
&= \sum_{y=1}^{\infty} y \times \frac{\lambda^y e^{-\lambda}}{y!} \\
&= \sum_{y=1}^{\infty} \frac{\lambda^y e^{-\lambda}}{(y-1)!} \\
&= \lambda \sum_{y=1}^{\infty} \frac{\lambda^{y-1} e^{-\lambda}}{(y-1)!} \\
&= \lambda \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x)!} \qquad x = y - 1 \\
&= \lambda
\end{aligned}$$

The following figure shows how Poisson approximates Binomial distribution.



Consistent Units

Example : For the case of the copper wire, suppose that the number of flaws follows a Poisson distribution with a mean of 2.3 flaws per millimeter.

Determine the probability of exactly 1 flaw in 2 millimeters of wire.

Let y denote the number of flaws in 2 millimeters of wire. Then, y has a Poisson distribution with

$$\mathbb{E}(y) = 2 \text{ mm} \times 2.3 \text{ flaws/mm} = 4.6 \text{ flaws}.$$

Therefore

$$P(y = 1) = e^{-4.6} \times 4.6.$$

It is important to use consistent units in the calculation of probabilities, means and variances involving Poisson random variables

Discrete Uniform Distribution

A random variable y has a **discrete uniform distribution** if

- it takes k possible values $(y_1, \dots, y_k)$
- each of the k values has equal probability, that is,

$$p(y = y_i) = 1/k, \quad i = 1, 2, \dots, k.$$

Examples : Toss a fair coin (Bernoulli(1/2)), Throw a die...

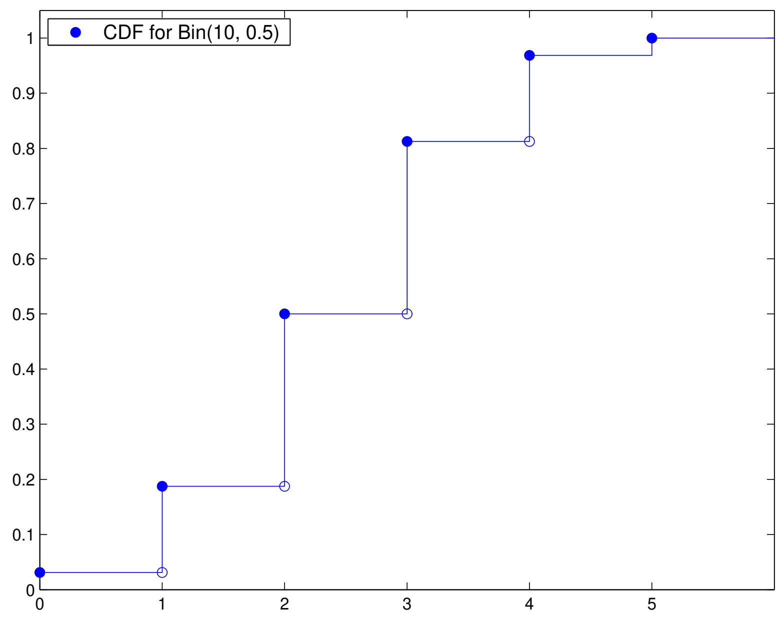
Cumulative Distribution Function

The **cumulative distribution function** $F(y_0)$ for a random variable y is equal to

$$F(y_0) = P(y \leq y_0).$$

Property of CDF

- $0 \leq F(y) \leq 1$.
- $F(y)$ is a monotonically increasing function.



Continuous Random Variable

A **continuous random variable** y is one that has the following three properties

1. y takes on an uncountably infinite number of values in the interval $(-\infty, \infty)$.

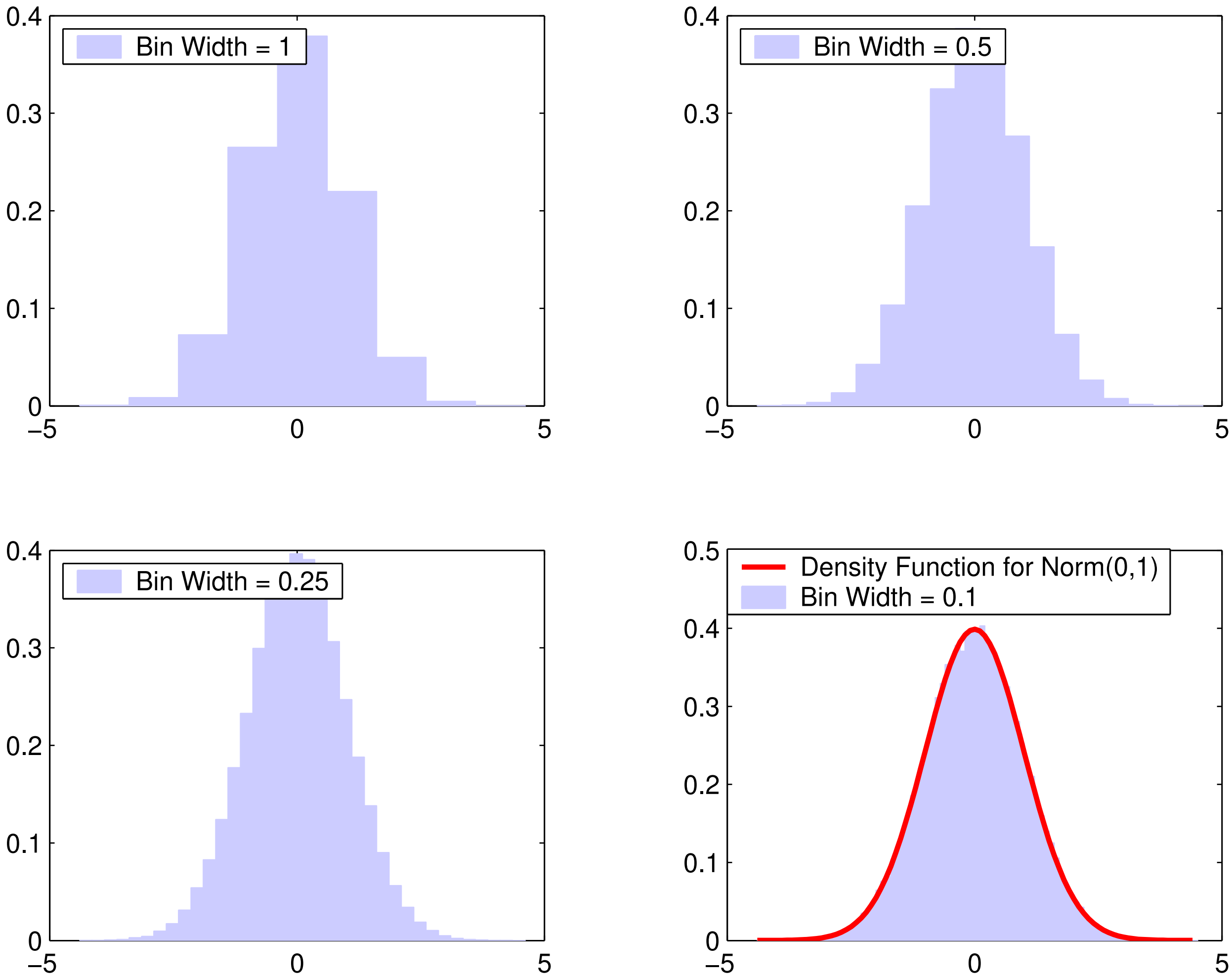
2. The CDF $F(y)$ is continuous.

3. The probability that y equals any one particular value is 0.

Histogram Revisit

Histogram is an approximation to a probability density function.

$$\text{Height} = \text{Relative Freq} \ / \ \text{Bin Width}$$



Density Function

- **Motivation**

$$\frac{F(y+\delta)-F(y)}{\delta} \to \frac{dF(y)}{dy},$$

  when $\delta \to 0$.

- **Definition** If $F(y)$ is the cumulative distribution function for a continuous random variable y, then the **density function** $f(y)$ for y is

$$f(y) = \frac{dF(y)}{dy}$$

- It follows from the definition that

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

Properties of a Density Function

1. $f(y) \geq 0$, and $f(y) = 0$ for y values that can not occur.

2. $\int_{-\infty}^{\infty} f(y)dy = F(\infty) = 1$

3. $P(a < y < b) = \int_a^b f(t)\ dt$.

Example 5.1 Let c be a constant and consider the density function

$$f(y) = \begin{cases} cy & \text{if } 0 \leq y \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

a. Find the value of c.

Since $\int_{\infty}^{\infty} f(y)dy = 1$, we have

$$\int_{-\infty}^{\infty} f(y)dy = c\int_0^1 y\ dy = c/2$$

$$\Longrightarrow c = 2$$

Expectation

The **expected values** of y and $g(y)$ are

$$\begin{aligned} \mathbb{E}(y) &= \int_{-\infty}^{\infty} y f(y) dy \\ \mathbb{E}[g(y)] &= \int_{-\infty}^{\infty} g(y) f(y) dy \end{aligned}$$

Properties of Expectation

- $\mathbb{E}(c) = c$, where c is a constant
- $\mathbb{E}(cy) = c\mathbb{E}(y)$
- $\mathbb{E}[g_1(y) + \cdots g_k(y)] = \mathbb{E}[g_1(y)] + \cdots \mathbb{E}[g_k(y)]$

Variance

- Let y be a continuous random variable with $\mathbb{E}(y) = \mu$, then

$$\sigma^2 = \mathbb{E}[(y-\mu)^2] = \int_{-\infty}^{\infty} (y-\mu)^2 f(y) dy$$

- Another expression for σ^2

$$\sigma^2 = \mathbb{E}(y^2) - \mu^2 = \int_{-\infty}^{\infty} y^2 f(y) dy - \mu^2$$

$$\begin{aligned}
\sigma^2 &= \int_{-\infty}^{\infty} (y^2 - 2y\mu + \mu^2) f(y)\ dy \\
&= \int_{-\infty}^{\infty} y^2 f(y)\ dy - 2\mu \int_{-\infty}^{\infty} y f(y)\ dy \\
&\qquad + \mu^2 \int_{-\infty}^{\infty} f(y)\ dy \\
&= \int_{-\infty}^{\infty} y^2 f(y)\ dy - 2\mu \cdot \mu + \mu^2 \\
&= \int_{-\infty}^{\infty} y^2 f(y)\ dy - \mu^2
\end{aligned}$$

Let y be a (discrete or continuous) random variable with mean μ and variance σ^2, then

$$\mu_{(c+y)} = \mu + c, \quad \mu_{cy} = c\mu.$$

$$\sigma^2_{(c+y)} = \sigma^2, \quad \sigma^2_{cy} = c^2\sigma^2.$$

$$\begin{aligned} \sigma^2_{(c+y)} &= \mathbb{E}[(c+y) - \mu_{(c+y)}]^2 \\ &= \mathbb{E}[c+y-(c+\mu)]^2 \\ &= \mathbb{E}(y-\mu)^2 = \sigma^2 \end{aligned}$$

$$\begin{aligned} \sigma^2_{cy} &= \mathbb{E}(cy - \mu_{cy})^2 \\ &= \mathbb{E}(cy - c\mu)^2 \\ &= c^2\mathbb{E}(y-\mu)^2 = c^2\sigma^2 \end{aligned}$$

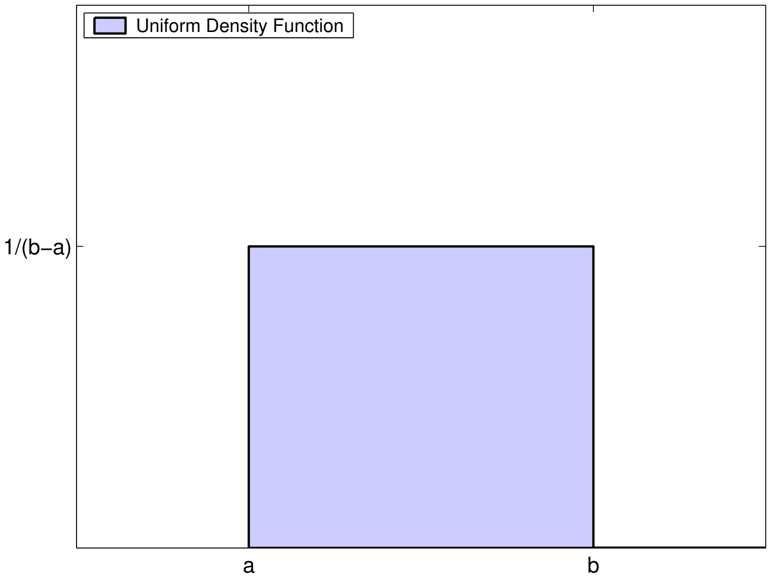
Uniform Distribution

A continuous random variable y with probability density function

$$f(y) = 1/(b-a), \qquad a \le y \le b$$

is a **continuous uniform random variable**

$$\mu = \frac{(a+b)}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12}$$



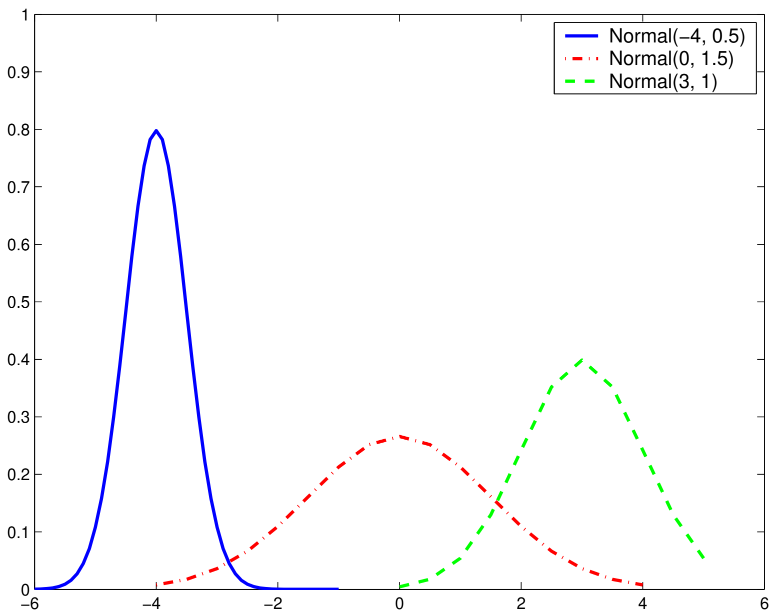
Normal Distribution

A continuous random variable y with probability density function

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

is a **normal random variable** with mean μ and variance σ^2.

Specially, the normal random variable with $\mu = 0$ and $\sigma^2 = 1$ is called **standard normal variable**.



- (Theorem 5.4) If y is a normal random variable with mean μ and variance σ^2, then

  $$z = \frac{y - \mu}{\sigma}$$

  is a standard normal variable.

- The entries in Table 4 of Appendix II are the area between 0 and $z > 0$ under a standard normal density curve.

- How to get CDF from the table?

  Let $T(z)$ denote the entry corresponding to z, then

  $$F(z) = \begin{cases} 0.5 + T(z) & \text{if } z \geq 0 \\ 0.5 - T(|z|) & \text{if } z < 0 \end{cases}$$

Example : The line width of for semiconductor manufacturing is assumed to be normally distributed with a mean of 0.5 micrometer and a standard deviation of 0.05 micrometer.

1. What is the probability that a line width is greater than 0.62 micrometer?

   The value $y = 0.62$ corresponds to a z value of
   $$z = \frac{y - \mu}{\sigma} = \frac{0.62 - 0.5}{0.05} = 2.4$$
   The entry corresponding to 2.4 in the table is 0.4918, thus
   $$\begin{aligned} P(y \geq 0.62) &= P(z \geq 2.4) = 1 - F(2.4) \\ &= 1 - (0.5 + 0.4918) \\ &= 0.5 - 0.4918 = 0.0082 \end{aligned}$$

2. What is the probability that a line width is between 0.47 and 0.63 micrometer?
   $$\begin{aligned} P(0.47 \leq y \leq 0.63) &= P(-0.6 \leq z \leq 2.6) \\ &= F(2.6) - F(-0.6) \\ &= 0.5 + T(2.6) - (0.5 - T(0.6)) \\ &= 0.2257 + 0.4953 = 0.721 \end{aligned}$$

3. The line width of 90% of samples is below what value?

   From the table, find that the value of z which has CDF equal to 90%, that is, $z = F^{-1}(0.9)$. From the table, the closest z value is 1.28. Thus

$$y = \sigma z + \mu = 1.28 \times 0.05 + 0.5 = 0.564.$$