Discrete Probability Distributions

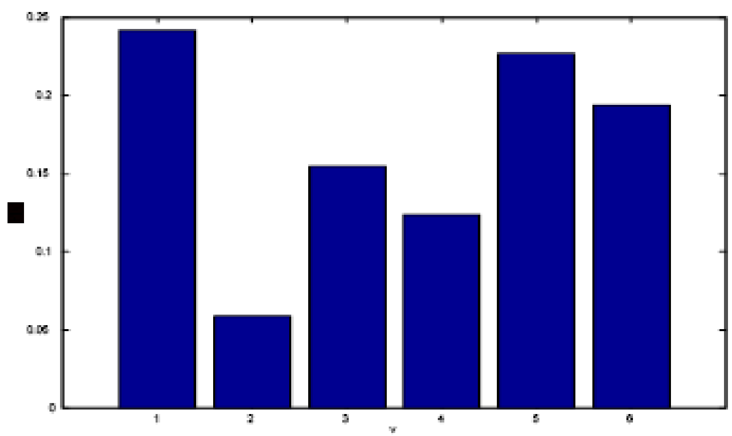
- The random variables only take on **discrete** values
  - e.g., throwing dice: possible values

$$v_i \in \{1, 2, 3, 4, 5, 6\}$$

- The probabilities sum to 1

$$\sum_i P(v_i) = 1$$

- Discrete distributions are particularly important in classification
- Probability Mass Function or Frequency Function (normalized histogram)



A “non fair” die

Multinomial distribution

Example: Multi-way coin toss, roll of dice

- **Data:** a set of N outcomes (multi-set)

  N_i - a number of times an outcome i has been seen

Model parameters: $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots \theta_k)$ **s.t.** $\sum_{i=1}^{k} \theta_i = 1$

θ_i - probability of an outcome i

Probability of data (likelihood)

$$P(N_1, N_2, \ldots N_k \mid \boldsymbol{\theta}, \xi) = \frac{N!}{N_1! N_2! \ldots N_k!} \theta_1^{N_1} \theta_2^{N_2} \ldots \theta_k^{N_k}$$

Multinomial distribution

ML estimate:

$$\theta_{i,ML} = \frac{N_i}{N}$$

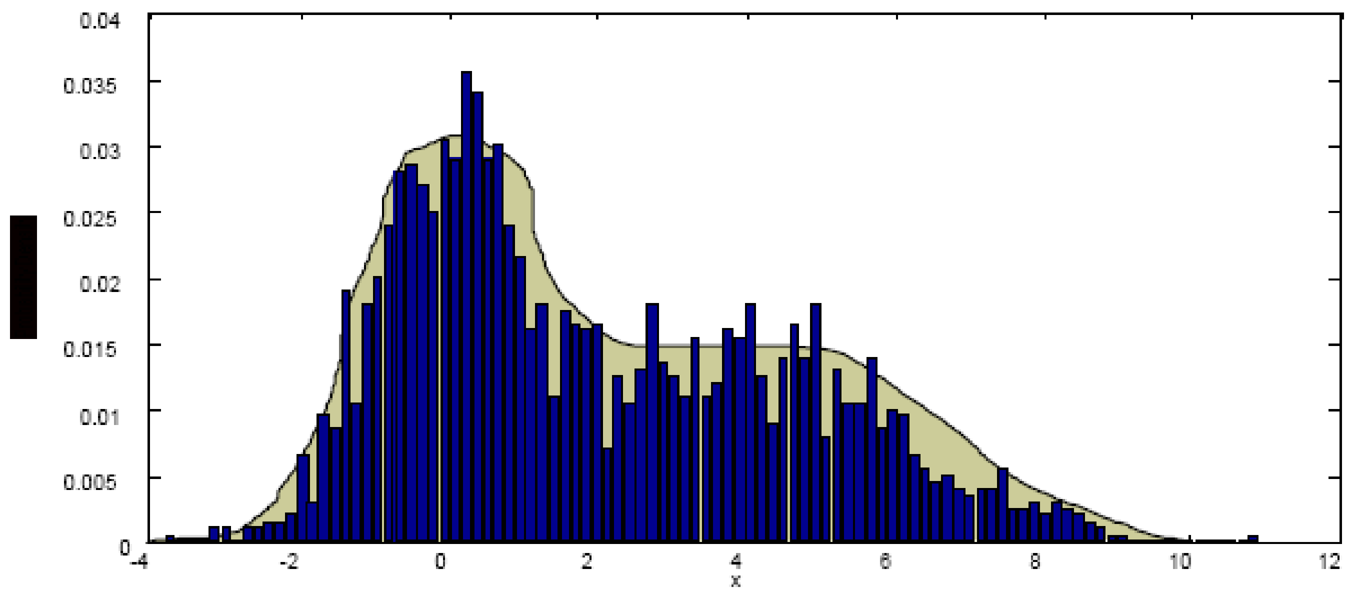
Continuous Probability Distributions

- Random variables take on **real values**.
- Continuous distributions are discrete distributions where the number of discrete values goes to infinity while the probability of each discrete value goes to zero.
- Probabilities become densities.
- Probability density integrates to 1.

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

Continuous Probability Distributions (cont'd)

- Probability Density Function $p(x)$



- Probability of an event:

$$P(a < x < b) = \int_a^b p(x)dx$$

A Classic Continuous Distribution

Normal Distribution

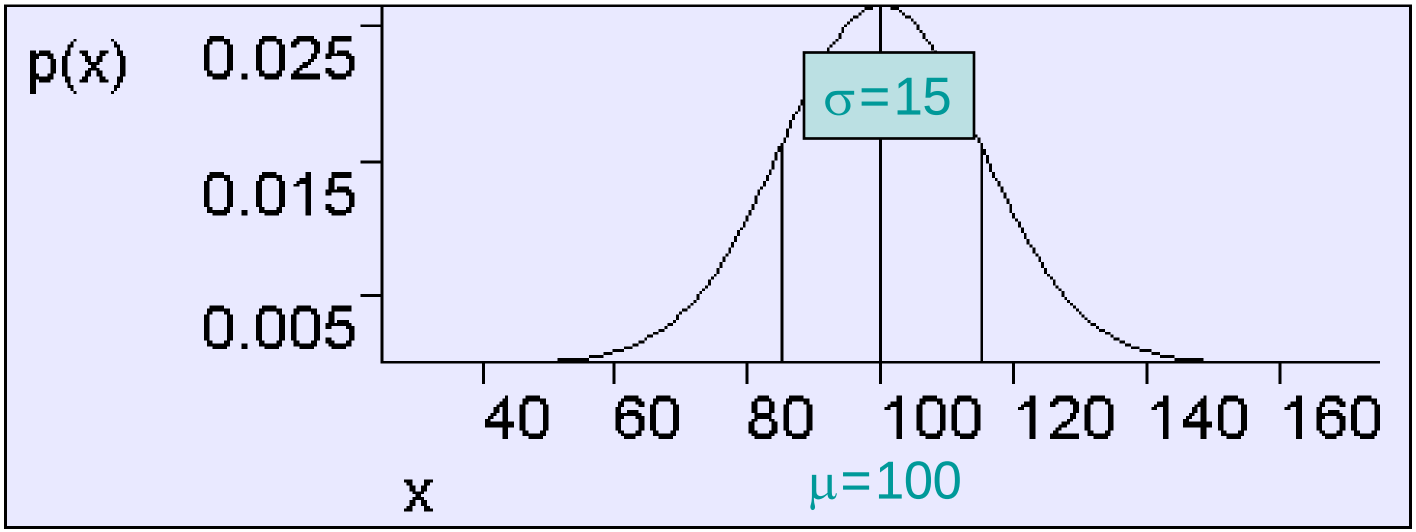
- The most important continuous distribution

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} \right)$$

- Also called Gaussian distribution after C.F.Gauss who proposed it
- Justified by the Central Limit Theorem:
  - roughly: "if a random variable is the sum of a large number of independent random variables, it is approximately normally distributed"
  - Many observed variables are the sum of several random variables

Shortform:

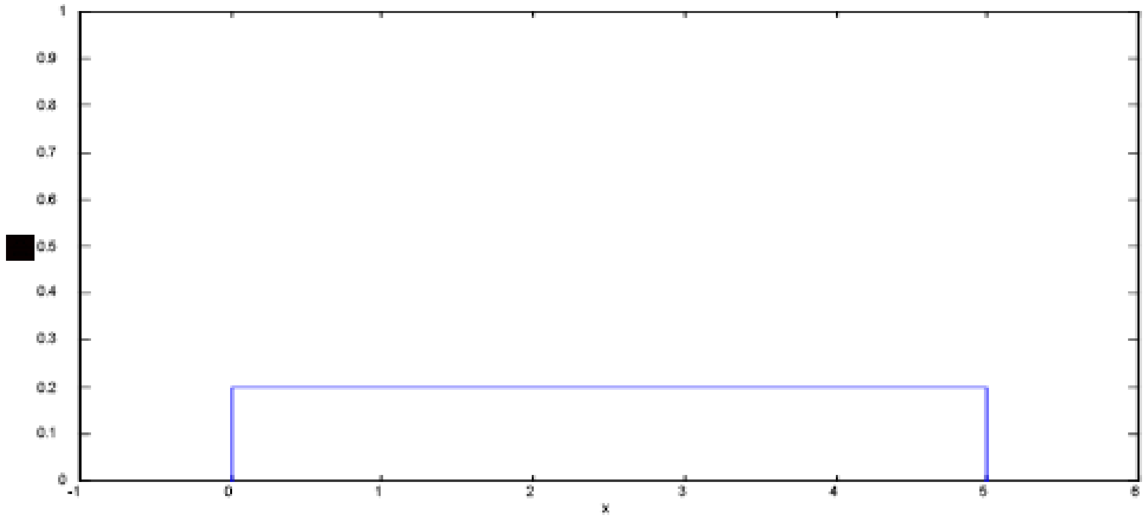
$$x \sim N(\mu, \Sigma)$$



Another Continuous Distribution

Uniform Distribution

- All data is equally probable within a bounded region R, $p(x)=1/R$.



Expected Values of Random Variables

- Definition for discrete random variables: $E\{\mathbf{x}\} = \sum_i \mathbf{x}_i P(\mathbf{x}_i) = \langle \mathbf{x} \rangle$
- Definition for continuous random variables: $E\{\mathbf{x}\} = \int_{-\infty}^{+\infty} \mathbf{x}_i p(\mathbf{x}_i) d\mathbf{x} = \langle \mathbf{x} \rangle$
- E{x} is often called the MEAN of x.
- E{x} is the "Center of Mass" of the distribution.

Expectation of Functions of Random Variables

$$E\{g(\mathbf{x})\} = ?$$

- as long as sum (or integral) remain bounded, just replace x*p(x) with g(x)*p(x) in E{}

- **Note: in general,** $E\{g(\mathbf{x})\} \neq g(E\{\mathbf{x}\})$
- **Other rules:**

$$E\{a\mathbf{x}\} = aE\{\mathbf{x}\}$$

$$E\{\mathbf{x}+\mathbf{y}\} = E\{\mathbf{x}\} + E\{\mathbf{y}\}$$

$$E\left\{\sum_i a_i \mathbf{x}_i\right\} = \sum_i a_i E\{\mathbf{x}_i\}$$

$$In\ general\ ,\ E\{\mathbf{x}\mathbf{y}\} \neq E\{\mathbf{x}\}E\{\mathbf{y}\}$$

Variance & Standard Deviation

- **Variance** $Var\{x\} = E\{(x - E\{x\})^2\} = E\{x^2\} - (E\{x\})^2$
- **Standard Deviation** $Std\{x\} = \sqrt{Var\{x\}}$

- the Var gives a measure of dispersion of the data

- Example I: What is the variance of a normal distribution?

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$$

- Example II: What is the variance of a uniform distribution x $\in$ [0, r] ?

$$Var\{x\} = \frac{r^2}{12}$$