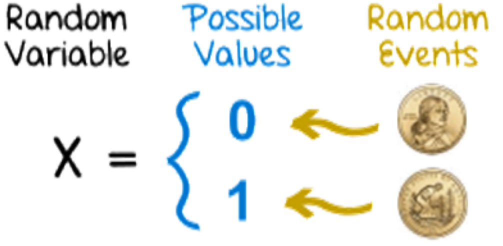
Probability Distributions

Random Variable

- A random variable *x* takes on a defined set of values with different probabilities.
  - For example, if you roll a die, the outcome is random (not fixed) and there are 6 possible outcomes, each of which occur with probability one-sixth.
  - Tossing of a coin



- Roughly, <u>probability</u> is how frequently we expect different outcomes to occur if we repeat the experiment over and over ("frequentist" view)

Two Types of Random Variables

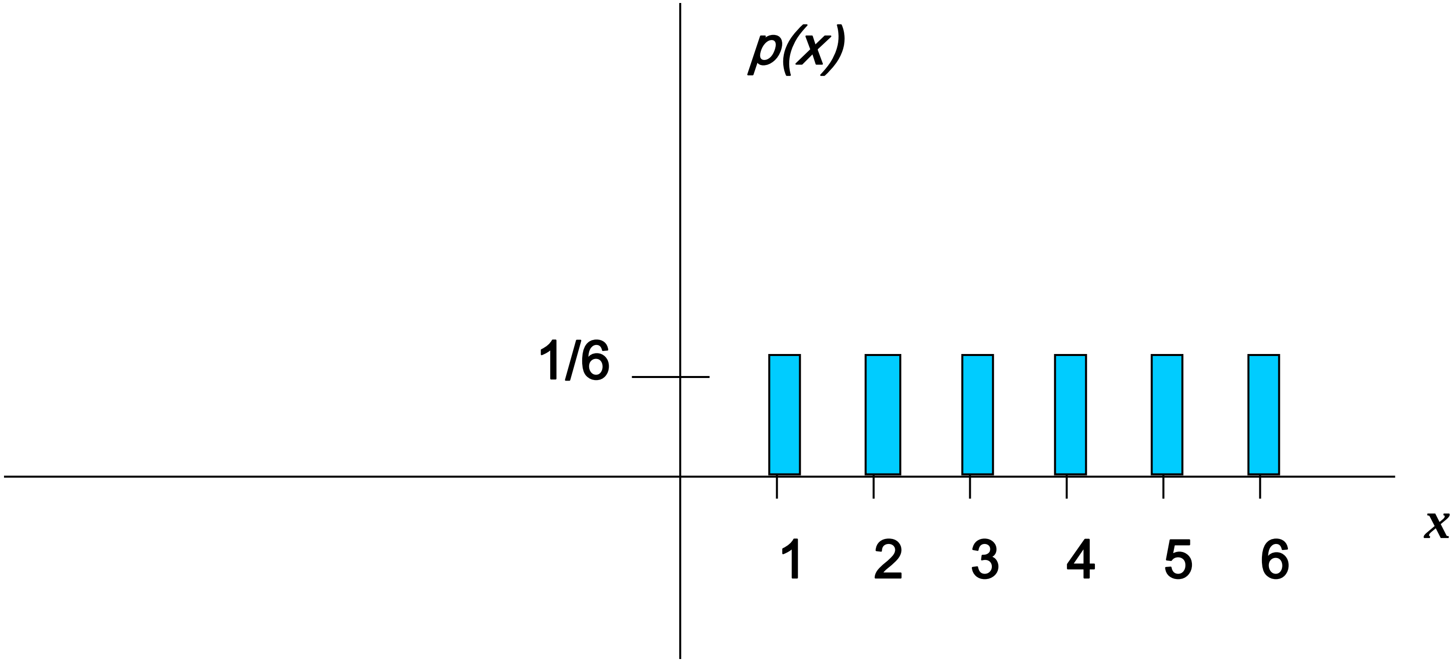
- **Discrete random variables**
  - Number of sales
  - Number of calls
  - Shares of stock
  - People in line
  - Mistakes per page

- **Continuous random variables**
  - Length
  - Depth
  - Volume
  - Time
  - Weight

Probability functions

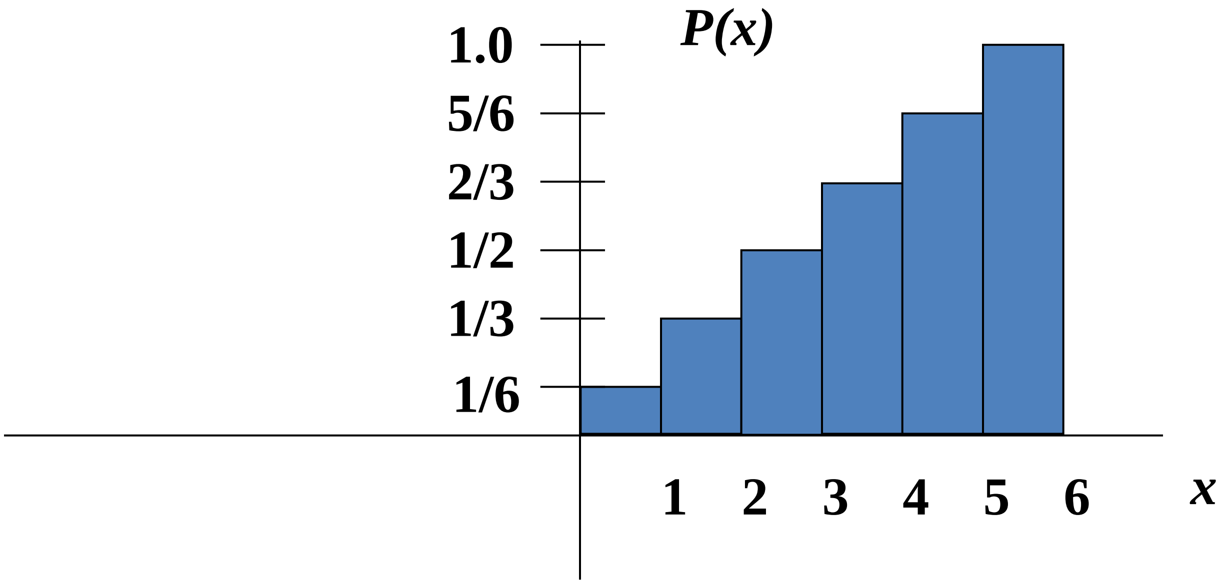
- A probability function maps the possible values of *x* against their respective probabilities of occurrence, *p(x)*
- *p(x)* is a number from 0 to 1.0.
- The area under a probability function is always 1.

Discrete example: roll of a die



$$\sum_{\text{all } x} P(x) = 1$$

Cumulative distribution function (CDF)



Practice Problem:

- The number of ships to arrive at a harbor on any given day is a random variable represented by *x*. The probability distribution for *x* is:

x	10	11	12	13	14
P(x)	.4	.2	.2	.1	.1

Find the probability that on a given day:

a. exactly 14 ships arrive $p(x=14) = .1$

b. At least 12 ships arrive $p(x \geq 12) = (.2 + .1 + .1) = .4$

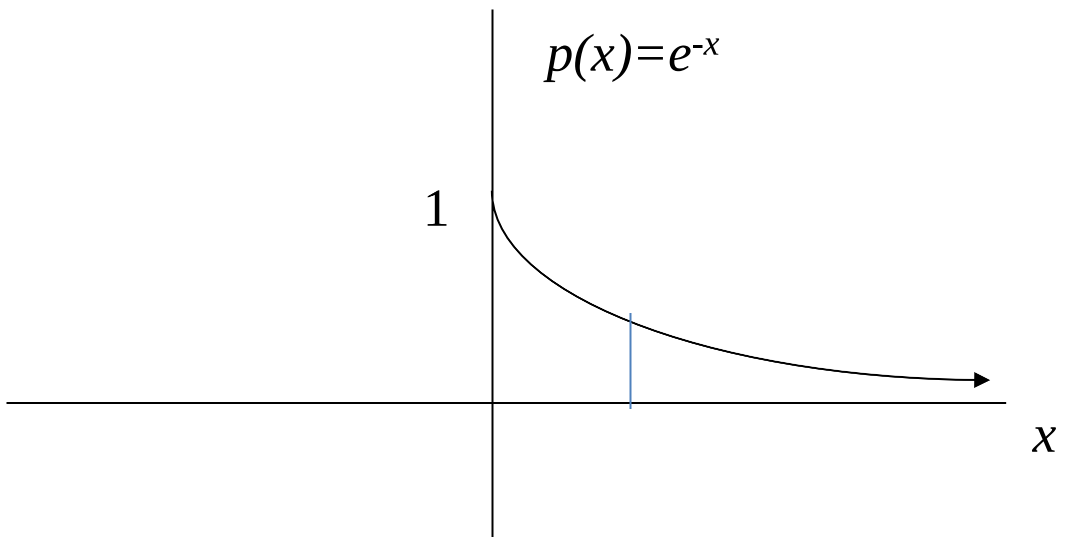
c. At most 11 ships arrive $p(x \leq 11) = (.4 + .2) = .6$

Continuous case

- The probability function that accompanies a continuous random variable is a continuous mathematical function that integrates to 1.
  - For example, recall the negative exponential function (in probability, this is called an "exponential distribution"): $f(x) = e^{-x}$
  - This function integrates to 1:

$$\int_0^{+\infty} e^{-x} = -e^{-x} \Big|_0^{+\infty} = 0 + 1 = 1$$

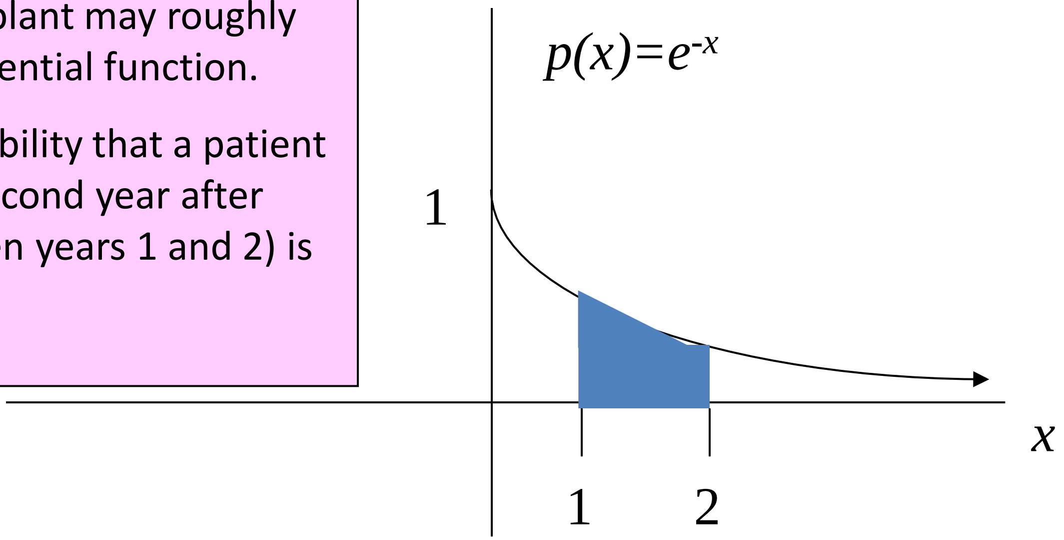
Continuous case: “probability density function” (pdf)



For example, the probability of *x* falling within 1 to 2:

Clinical example: Survival times after lung transplant may roughly follow an exponential function.

Then, the probability that a patient will die in the second year after surgery (between years 1 and 2) is 23%.



$$P(1 \le x \le 2) = \int_1^2 e^{-x} = -e^{-x} \Big|_1^2 = -e^{-2} - -e^{-1} = -.135 + .368 = .23$$

Expected Value and Variance

- All probability distributions are characterized by an expected value (mean) and a variance (standard deviation squared).

Expected value of a random variable

- Expected value is just the average or mean (μ) of random variable *x*.
- It's sometimes called a "weighted average" because more frequent values of X are weighted more highly in the average.
- It's also how we expect X to behave on-average over the long run ("frequentist" view again).

Expected value, formally

Discrete case:

$$E(X) = \sum_{\text{all } x} x_i \, p(x_i)$$

Continuous case:

$$E(X) = \int_{\text{all } x} x_i \, p(x_i) dx$$

Gambling (or how casinos can afford to give so many free drinks…)

A roulette wheel has the numbers 1 through 36, as well as 0 and 00. If you bet $1 that an odd number comes up, you win or lose $1 according to whether or not that event occurs. If random variable X denotes your net gain, X=1 with probability 18/38 and X= -1 with probability 20/38.

$E(X) = 1(18/38) – 1 (20/38) = -\$.053$

On average, the casino wins (and the player loses) 5 cents per game.

The casino rakes in even more if the stakes are higher:

$E(X) = 10(18/38) – 10 (20/38) = -\$.53$

If the cost is $10 per game, the casino wins an average of 53 cents per game. If 10,000 games are played in a night, that's a cool $5300.

Variance/standard deviation

σ^2=Var(x) =E(x-μ)^2

"The expected (or average) squared distance (or deviation) from the mean"

$$\sigma^2 = Var(x) = E[(x-\mu)^2] = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

Variance, continuous

Discrete case:

$$Var(X) = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

Continuous case?:

$$Var(X) = \int_{\text{all } x} (x_i - \mu)^2 p(x_i) dx$$

Practice Problem

On the roulette wheel, X=1 with probability 18/38 and X= -1 with probability 20/38.

- We already calculated the mean to be = -$.053. What's the variance of X?

Answer

$$\sigma^2 = \sum_{\text{all } x} (x_i - \mu)^2 p(x_i)$$

$$= (+1 - -.053)^2 (18/38) + (-1 - -.053)^2 (20/38)$$

$$= (1.053)^2 (18/38) + (-1 + .053)^2 (20/38)$$

$$= (1.053)^2 (18/38) + (-.947)^2 (20/38)$$

$$= .997$$

$$\sigma = \sqrt{.997} = .99$$

Standard deviation is \$.99. Interpretation: On average, you're either 1 dollar above or 1 dollar below the mean.

Important discrete distributions

Binomial Probability Distribution

A binomial random variable ***X*** is defined to the number of "successes" in *n* independent trials where the *P("success")* = *p* is constant.

Notation: $X \sim BIN(n,p)$

In the definition above notice the following conditions need to be satisfied for a binomial experiment:

1. There is a fixed number of *n* trials carried out.
2. The outcome of a given trial is either a "success" or "failure".
3. The probability of success (*p*) remains constant from trial to trial.
4. The trials are independent, the outcome of a trial is not affected by the outcome of any other trial.

B.D Pmf

- If $X \sim$ BIN(n, p), then

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad x = 0,1,\ldots,n.$$

Examples:::

1. Tossing a coin and considering heads as success and tails as failure.
2. Checking items from a production line: success = not defective, failure = defective.
3. Phoning a call centre: success = operator free; failure = no operator free.

Example: Treatment of Kidney Cancer

- Suppose we have $n = 40$ patients who will be receiving an experimental therapy which is believed to be better than current treatments which historically have had a 5-year survival rate of 20%, i.e. the probability of 5-year survival is
$p = .20$.
- Thus the number of patients out of 40 in our study surviving at least 5 years has a binomial distribution, i.e. $X \sim BIN(40,.20)$.

The Poisson Distribution

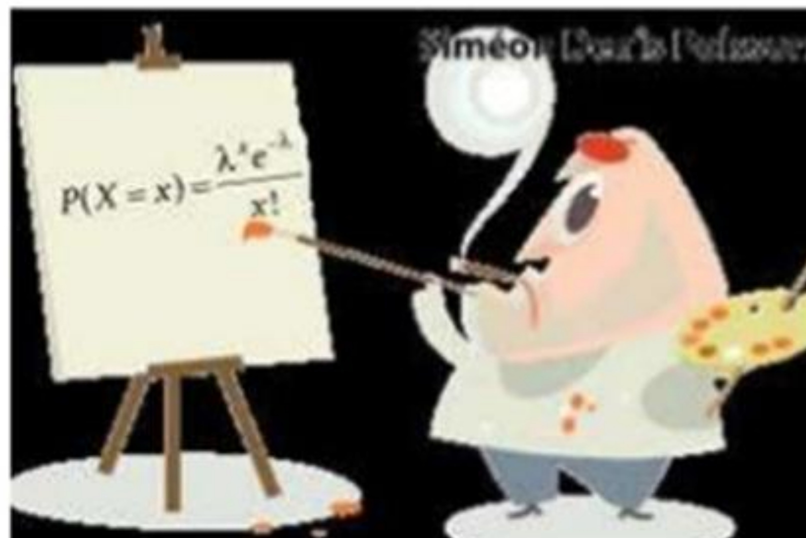
We can use the Poisson distribution to estimate the probability of arrivals at a car wash in one hour or the number of leaks in 100 miles of pipeline. Bell Labs uses it to model the arrival of phone calls.

- The probability that there are *r* occurrences in a given interval is given by $\frac{\lambda^r}{r!} e^{-\lambda}$.

Where,

λ = Mean no. of occurrences in a time interval

r =No. of trials.



EXAMPLES:

1. Number of telephone calls in a week.
2. Number of people arriving at a checkout in a day.
3. Number of industrial accidents per month in a manufacturing plant.

GEOMETRIC DISTRIBUTION

- The probability distribution of the number *X* of Bernoulli trials needed to get one success, supported on the set { 1, 2, 3, ...}

- The probability of success on each trial is p, then the probability that the kth trial (out of k trials) is the first success is

$$\Pr(X = k) = (1-p)^{k-1}p$$

for k = 1, 2, 3, ...

Examples of Geometric PDF

- First car arriving at a service station that needs brake work
- Flipping a coin until the first tail is observed
- First plane arriving at an airport that needs repair
- Number of house showings before a sale is concluded

The Hypergeometric Distribution

- Randomly draw *n* elements from a set of *N* elements, without replacement. Assume there are *r* successes and *N-r* failures in the *N* elements.
- The hypergeometric random variable is the number of successes, *x*, drawn from the *r* available in the *n* selections.

The Hypergeometric Distribution

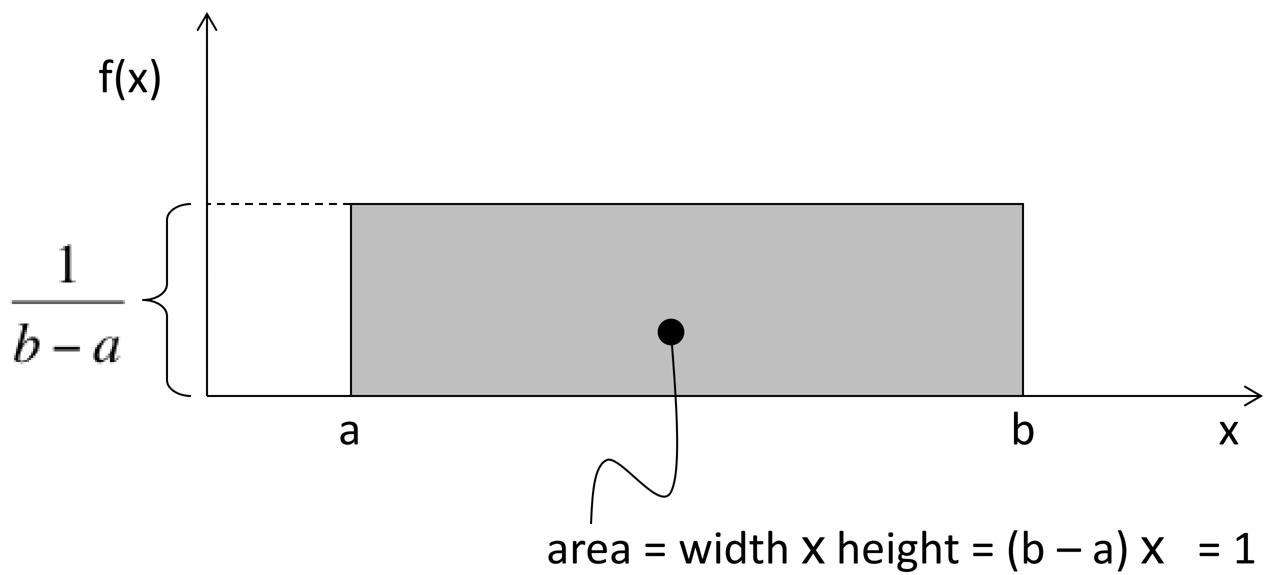
$$P(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

where
N = the total number of elements
r = number of successes in the *N* elements
n = number of elements drawn
X = the number of successes in the *n* elements

Uniform Distribution

- Consider the uniform probability distribution
- It is described by the function:

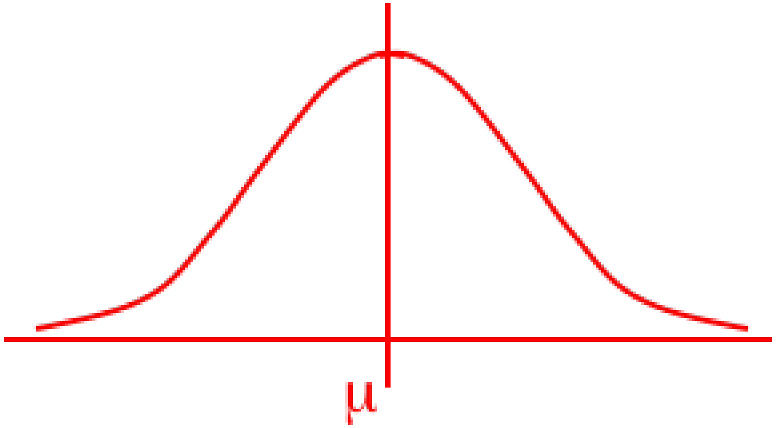
$$f(x) = \frac{1}{b-a}, \quad where \quad a \le x \le b$$



The Normal Distribution

- The normal distribution is the most important of all probability distributions. The probability density function of a normal random variable is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < \infty$$



- It looks like this:
- Bell shaped,
- Symmetrical around the mean μ

Exponential Distribution

- The exponential distribution which has this probability density function:

$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Gamma Distribution

A continuous random variable X is said to have a gamma distribution if the probability density function of $\boldsymbol{X}$ is

$$f(x) = \frac{1}{\beta^{\alpha}\,\Gamma(\alpha)}\, x^{\alpha-1} e^{-\frac{x}{\beta}} \quad for \quad x \geq 0$$

where the parameters α and β satisfy $\alpha > 0, \beta > 0.$

Beta Distribution

A random variable X is said to have a beta distribution with parameters α, β if the pdf of X is

Beta – 1

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1$$

Beta – 2

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1+x)^{-(\alpha+\beta)}, x > 0$$

THANK YOU