Introduction to Probability Distributions

Manisha J

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.
  - For example, if you poll people about their voting preferences, the percentage of the sample that responds "Yes on Proposition 100" is a also a random variable (the percentage will be slightly differently every time you poll).
- Roughly, probability is how frequently we expect different outcomes to occur if we repeat the experiment over and over ("frequentist" view)

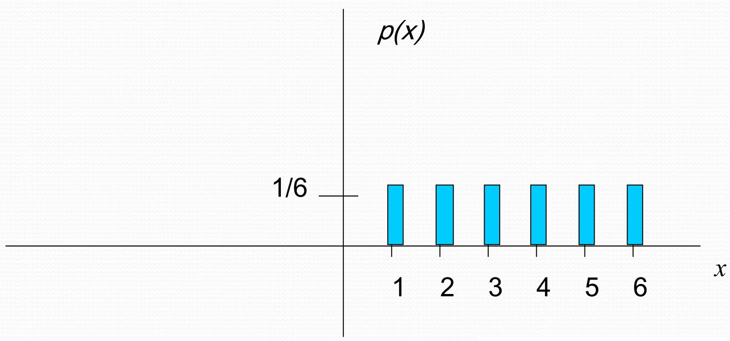
Random variables can be discrete or continuous

- **Discrete** random variables have a countable number of outcomes
  - Examples: Dead/alive, treatment/placebo, dice, counts, etc.
- **Continuous** random variables have an infinite continuum of possible values.
  - Examples: blood pressure, weight, the speed of a car, the real numbers from 1 to 6.

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

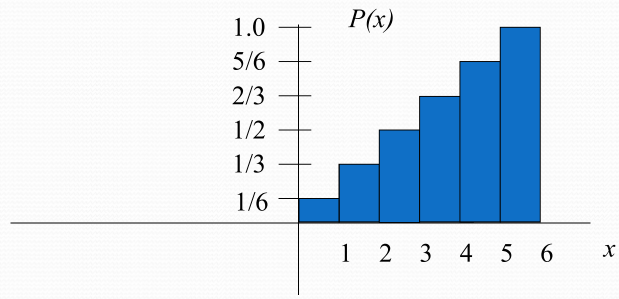


$$\sum_{\text{all } x} P(x) = 1$$

Probability mass function (pmf)

x	$p(x)$
1	$p(x=1)=1/6$
2	$p(x=2)=1/6$
3	$p(x=3)=1/6$
4	$p(x=4)=1/6$
5	$p(x=5)=1/6$
6	$p(x=6)=1/6$
	1.0

Cumulative mass function (CMF



Cumulative mass function

x	*P(x≤A)*
1	*P(x≤1)*=1/6
2	*P(x≤2)*=2/6
3	*P(x≤3)*=3/6
4	*P(x≤4)*=4/6
5	*P(x≤5)*=5/6
6	*P(x≤6)*=6/6

Practice Problem:

- The number of patients seen in the ER in any given hour 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 in a given hour:

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

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

c. At most 11 patients 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:

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

Discrete Distributions

Relative frequency distributions for "counting" experiments.

- Bernoulli Distribution
- Binomial Distribution
- Negative Binomial
- Poisson Distribution
- Geometric Distribution
- Multinomial Distribution

Yes-No responses.

Sums of Bernoulli responses

Number of trials to k^{th} event

Points in given space

Number of trials until first *success*

Multiple possible outcomes for each trial

Continuous Distributions

Foundations for much of statistical inference

- **Normal Distribution**
- Log Normal Distribution
- Gamma Distribution
- **Chi Square Distribution**
- **F Distribution**
- **t Distribution**
- Weibull Distribution
- Extreme Value Distribution (Type I and II)

Environmental variables

Time to failure, radioactivity

Basis for statistical tests.

Lifetime distributions

Continuous random variables are defined for continuous numbers on the real line. Probabilities have to be computed for all possible sets of numbers.

Illustration