SMART CHOICES

PROBABILITY

- Many real-life problems are probabilistic
- Probability used in quality control, educational testing, business predictions, etc.

Probability Basics

- In the situations of interest
  - More than one possible outcome (e.g. When a dice is rolled, there are 6 possible outcomes)
  - Uncertainty about the actual outcome
- Situation is called an experiment
- Outcomes are called events
- Events which are not lists of other events are called elementary events (e.g. Face of dice on a roll is 4)
- Events which can be decomposed into elementary events are called compound event (e.g. An even numbered face comes up in the roll of a dice)

LECTURE NOTES

Chapter 8

Computational Statistics

Assigning Probability

If an event E consists of m of the n equally likely elementary events of the experiment:

- Then, the probability of the event is given by

  P(E) = m/n

- E.g. probability of an even face in a roll of a dice

  n = 6

  m = 3

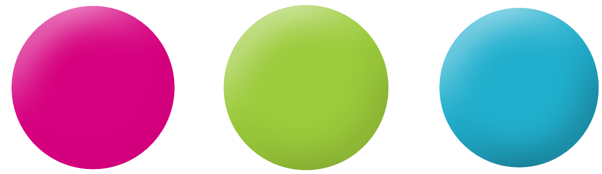
  P(E) = m/n = 3/6 = 0.5

Combinatorial Analysis

- **Theorem:** If k experiments are performed, and there are n_i different outcomes for the i^{th} experiment, $I \leq i \leq k$, then there are $n_1 \times n_2 \times \ldots n_k$ different outcomes when all k experiments are performed and order matters. This is written as $\prod_{i=1}^{k} n_i$
- **Theorem:** For a population of n elements, there are $n(n-1)\ldots(n-k+1)$ distinct ordered samples of size k if samples are taken without replacement. This is written as $(n)^k$.

Example: 5 dice are rolled. What is the probability that no two dice show the same face?

- N = 6^5
- M = $(6)^5$
- Probability = $\frac{(6)_5}{6^5} = \frac{5}{54}$



Example: 2 balls drawn from an urn containing 3 balls. What is the probability of drawing a specific ball?

- Calculate n
  - First ball can be picked in 3 ways
  - Second ball can be picked in 2 ways
  - So, n = 3 x 2 = 6
- Calculate m
  - Selected ball may be first or second to be picked
  - If first, 2 ways to pick second ball
  - If second, 2 ways to pick first ball
  - Total of 2 + 2 = 4 ways. So, m = 4
- Probability = 4/ 6 = 2/3

SAMPLING

We often have to make decisions based on limited data.

- A sip of wine to judge the quality of a bottle
- A drop of blood to determine infection
- Opinions of about 2,000 people to determine the mood of the nation
- 30 stocks in the DJI to determine the state of a market comprising over 3,000 stocks
- A spoonful for a kid to determine the likeability of a new dish

Most of these decisions would be impossible without sampling.

- If you need to drink the whole bottle to determine its quality, the assessment is too late
- If the doctor needs all of your blood to determine the infection, it would be too late
- If the administration needed to run the census every month to determine what the country wanted, it would be too expensive

If order does not matter, for example, a hand of bridge, and *k* samples are taken without replacement from *n* objects, the number of elementary outcomes is

$$\frac{(n)_k}{k!} = \frac{n!}{k!(n-k)!} = C(n,k)$$

This is because each of the $(n)_k$ ordered outcomes can be rearranged *k!* times.

Example: In bridge, what is the probability that a given hand contains no spades?

Unordered Sampling

- n = number of ways a hand can be dealt = C(52, 13)
$$= \frac{52.51.50.49.48.47.46.45.44.43.42.41.40}{13.12.11.10.9.8.7.6.5.4.3.2.1}$$
- m = number of ways a hand can be dealt without a spade = C(39, 13)
$$= \frac{39.38.37.36.35.34.33.32.31.29.28.27}{13.12.11.10.9.8.7.6.5.4.3.2.1}$$
- Probability

$$= \frac{39.38.37.36.35.34.33.32.31.30.29.28.27}{52.51.50.49.48.47.46.45.44.43.42.41.40} = 0.0128 \approx 1.3\%$$

Interesting Early Use From 1279

Trial of the Pyx

King of England, Edward I, wanted a procedure to ensure that the coins minted met the defined standards for gold or silver content.

- Pyx derived from Greek word for box
- Referred to the container holding the coins to be sampled
- Coins in the pyx were probably selected at random from the mint's output

The sampled coins were compared to a plate stored in a protected vault.

3

UNEQUALLY LIKELY OUTCOMES

What if the outcomes are not equally likely?

- e.g. toss a coin twice
- Can get 0, 1 or 2 heads

What is the probability of getting 2 heads?

The **sample space** of an experiment is the set of all elementary outcomes of the experiment.

- E.g. S = {{H,H}, {H,T}, {T, H}, {T, T}} for the above experiment
- In our example, P(0) = P(2) = ¼; P(1) = ½

FUNCTIONS

- Often we like to know how the probability changes as a variable changes (e.g. probability of buying by income, age, etc.)
- Most of these outcomes of interest are not equally likely
- Convenient to represent probability as a function of the variable
- Can find optimum, rate of change etc.
- Let *X* and *Y* be sets
- Function *f* from *X* to *Y* is a rule which assigns to each of some (not necessarily all) elements *x* of *X*, an element *f(x)* of *Y*
- Domain of *f* is the subset $D \subset X$ to which f assigns elements of *Y*
- Range of f is the subset $R \subset Y$ of elements so assigned from *Y*
- We write *f*: $D \subset X \rightarrow R \subset Y$

PROBABILITY MEASURE

In the general case of N elementary outcomes E_1, … E_n, which may not be equally likely,

- Let $P(E_i) = p_i \geq 0$ with $\sum_{i=1}^{n} p_i = 1$
- Define $P(E) = \sum P(E_i)$; sum taken over $E_i \subset E$, as the probability of event $E \subset S$
- Then P is a probability measure on the finite sample space S

Rules

- Normality

  $P(S) = 1$

- Non-negativity

  $P(E) \geq 0 \ \forall \ E \subset S$

- Additivity

  $P(E + F) = P(E) + P(F)$ for disjoint sets E and F

RANDOM VARIABLE

4

In experiments in which outcomes are not equally likely, it is useful to associate random variables with experimental outcomes.

Random variable (RV)

- A number x(ζ) assigned to every outcome ζ of an experiment

  e.g. gains in a lottery, number of heads when a coin is tossed 100 times

- A random variable is a function whose domain is the set S of all possible outcomes of the experiment

  e.g. when a die is thrown, gain for face i = 10i

How are the concepts of probability related to random variables?

In experiments, a common question of interest is What is the probability that $x \leq x_1$; or $x_1 \leq x \leq x_2$? (e.g. probability that the person weighs less than 180 lbs?)

Probabilities are associated with events.

How do we calculate the probability that a RV meets some specified constraints?

- We start with the set of experimental outcomes for which the RV meets the specified constraints

  e.g. $\{x \leq x_1\}$

- $\{x \leq x1\}$ is the set of experimental outcomes for which the RV meets the specified constraints

  In our example, the set has 5 elements

  probability $x(\zeta) \leq 180 = 5/7$

- RV satisfies the following properties

  $P\{x = \infty\} = 0$
  $P\{x = -\infty\} = 0$

- In general, $P\{x \leq x\}$ of event $\{x \leq x\}$ depends on x because elements in the event $\{x \leq x\}$ change with x

PROBABILITY DISTRIBUTIONS

Random Variables and Distribution Functions

- $P\{x \leq x\}$ is denoted by $F(x)$
- $F(x)$ is called the cumulative distribution function of the random variable x

  $F(x) = P\{\mathbf{x} \leq x\}\ \forall\ x \text{ in } -\infty \leq x \leq +\infty$

- For any x, $F(x)$ gives us the probability that the RV associated with the outcome is less than or equal to x

Outcome (ζ)	Weight (lbs) = Random Variable X (ζ)
John	250
Joe	175
Jill	135
Jane	145
Jack	170
Jay	195
Jean	140

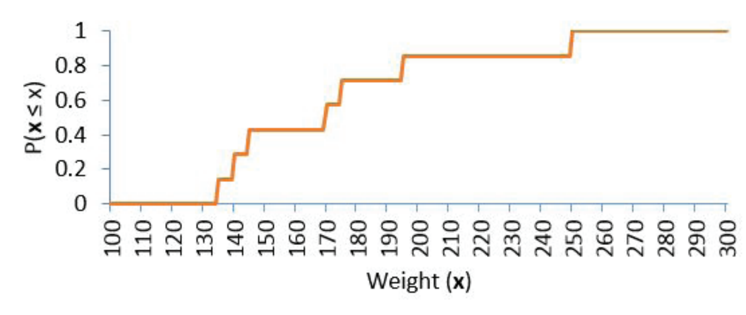
Cumulative Distribution Function

- Let us consider our weight example

  No person weighs below 135 lbs, hence probability $x(\zeta) < 135 = 0$

- Everybody weighs less than or equal to 250 lbs

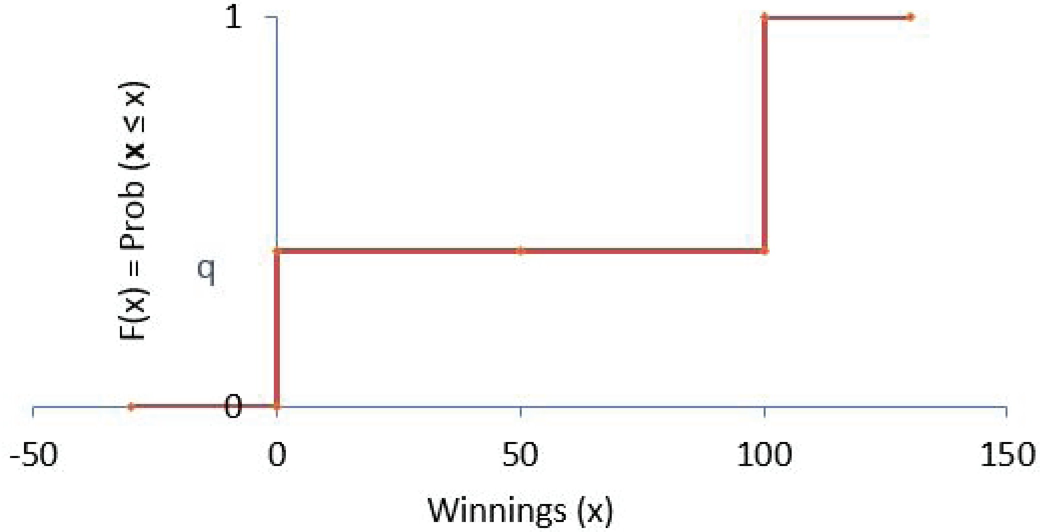
  probability $x(\zeta) \leq 250 = 1$



Cumulative Probability Distribution Example:

The only way to earn between \$0 and \$100 is to draw tails, which gives you earnings of \$0.

- This happens with probability q



Another Distribution Example:

A fair coin is tossed twice. Random variable x represents the number of heads. Find F(x).

- There are 4 equally likely, mutually exclusive cases

  {TT}, {HT}, {TH}, {HH}

  x(TT) = 0; x(HT) = x(TH) = 1; x(HH) = 2

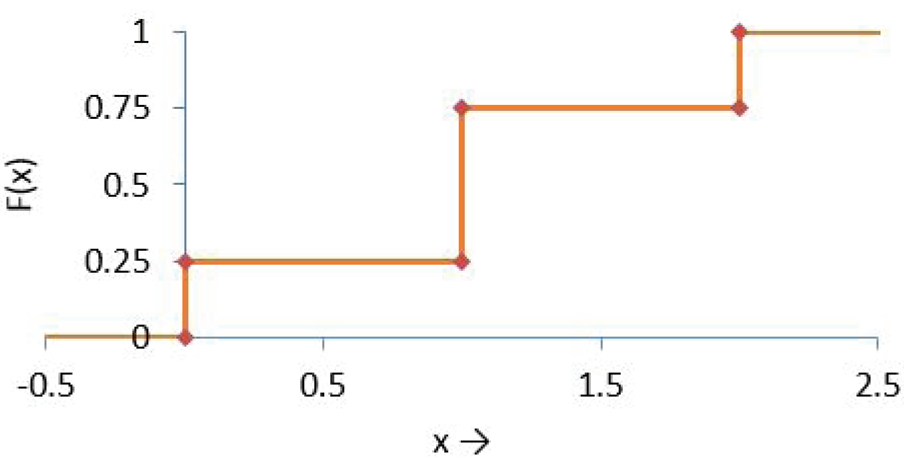
- We have the four cases

  if x < 0, F(x) = 0

  If x ≥ 2, F(x) = 1

  If 0 ≤ x < 1, F(x) = P{TT} = ¼

  If 1 ≤ x < 2, F(x) = P{TT} + P{HT} + P{TH} = ¾



Properties of Distribution Functions

- $F(-\infty) = 0$
- $F(+\infty) = 1$
- if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$

  Non-decreasing property
- $P\{x > x\} = 1 - F(x)$
- $P\{x_1 < x \leq x_2\} = F(x_2) - F(x_1)$

Continuity of a Random Variable

Random variable x is called continuous if the distribution function is continuous.

$$F(x-) = F(x)$$

- For a continuous random variable

  $P\{x = x\} = 0$

PROBABILITY DENSITY FUNCTION

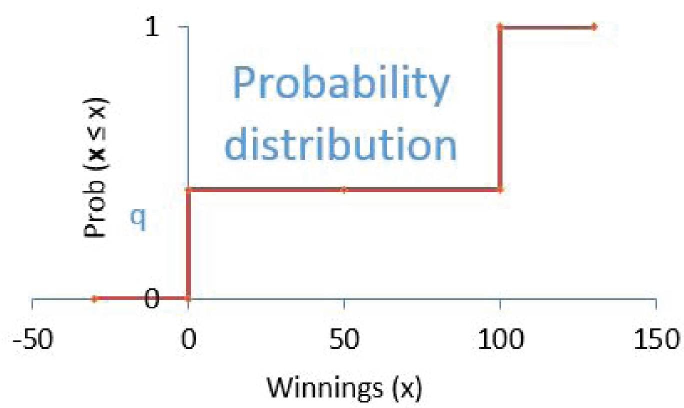
The derivative of the probability distribution function (p.d.f.) is called the probability density function.

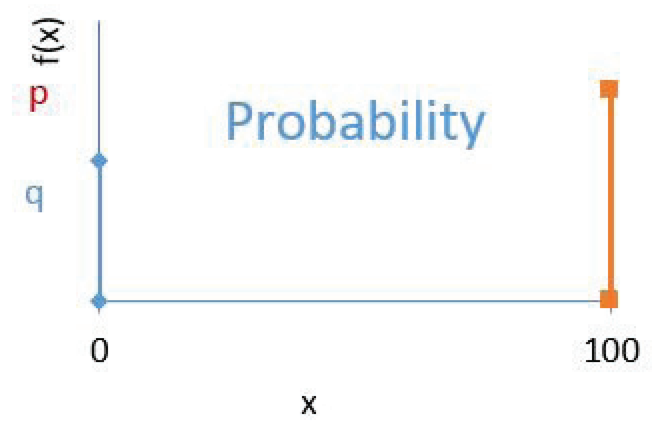
- Written as f(x) of the random variable **x**

$$f(x) = \frac{dF(x)}{dx}$$

- Given the p.d.f., the probability distribution F(x) is obtained as

$$F(x) = \int_{x=-\infty}^{x} f(u).du$$





Properties of the p.d.f.

- Since F(x) is monotonically non-decreasing

  $f(x) \geq 0 \;\forall\; x$

- If x is a continuous random variable, f(x) is a continuous function
- Since $F(\infty) = 1$,
- $P(x_1 < x \le x_2) = F(x_2) - F(x_1) =$

  Probability that RV **x** lies in interval (x_1, x_2) is the area under the pdf in (x_1, x_2)

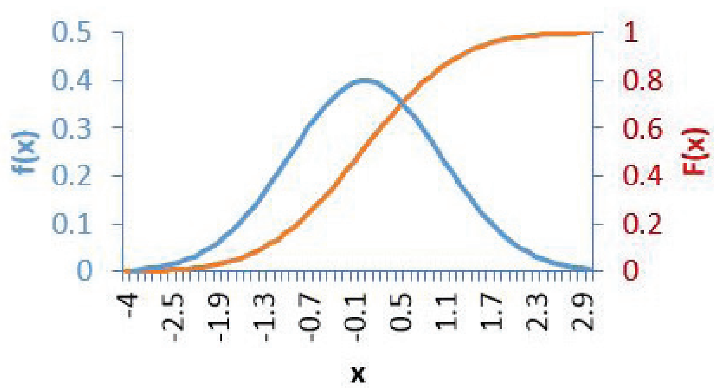
CONTINUOUS RANDOM VARIABLES

Gaussian distribution

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

Parameters are μ and σ, often written as $x \sim N(\mu, \sigma^2)$.

- One of the most important distributions
- Natural phenomena follow N
- Average of independent, identically distributed random variables
- Hence, this is also called the normal distribution
- Values available in tables



Francis Galton (1822-1911) was a first cousin of Charles Darwin who never worked to earn a living but measured everything compulsively. This is his view on Normal Distribution:

> **[T]he "Law of Frequency of Error" … reigns with serenity and in complete self-effacement amidst the wildest confusion. The huger the mob … the more perfect is its sway. It is the supreme law of Unreason. Whenever large samples of chaotic elements are taken in hand … an unsuspected and most beautiful form of regularity proves to have been latent all along.**

8

EXPONENTIAL DISTRIBUTION

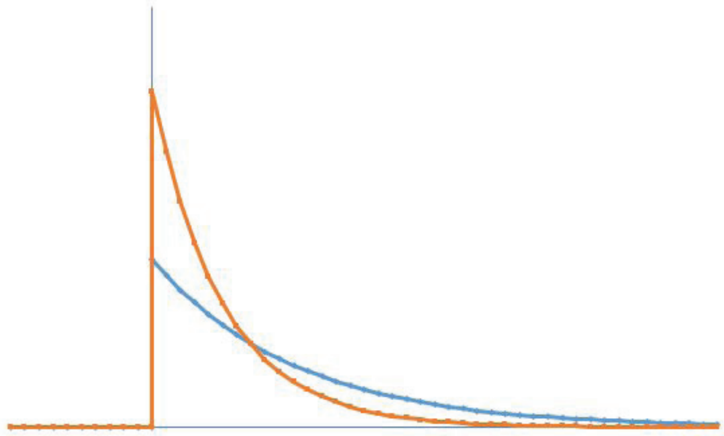
If events over non-overlapping intervals are independent, then the distribution of waiting times for these events is exponential.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & otherwise \end{cases}$$

- $F(x) = 1 - e^{-\lambda x}$

Example: The waiting time at a restaurant is exponentially distributed with a mean of 5 (λ = 1/3) minutes. Probability wait time > 10 minutes?

$P(t > 10) = 1 - F(10) = 1 - (1 - e^{-10/5}) = 0.135$



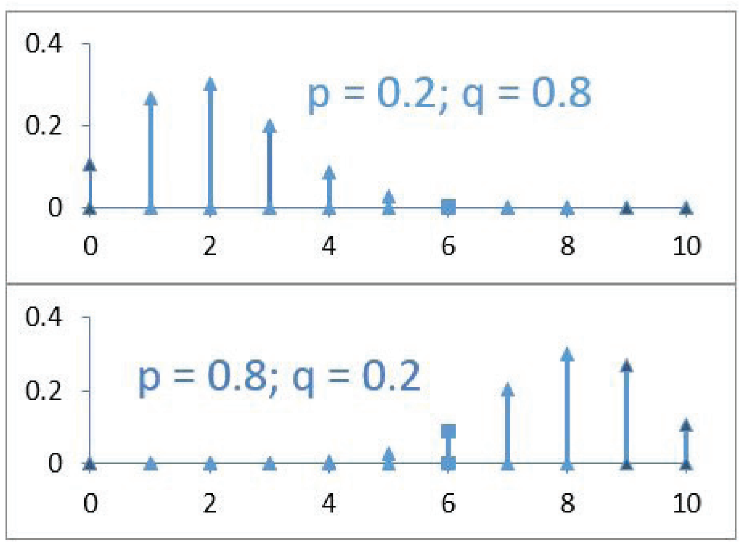
BINOMIAL DISTRIBUTION

If n Bernoulli experiments are performed, with probability of success = p, the total number of favorable outcomes, **y**, is called a Binomial random variable.

P{y = k} = (nk)pk .qn-k

$p + q = 1; 0 \le k \le n$

For example, if a coin is tossed multiple times, what is the probability of drawing 5 heads?



BERNOULLI DISTRIBUTION

x is Bernoulli distributed if **x** takes the values 0 and 1 with probabilities

$P\{x = 1\} = p; P\{x = 0\} = q = 1 - p$

- Coin toss is an example

Priors

For some distributions, it is possible to specify prior belief. How confident are you about the parameters of the distribution? E.g. for the binomial distribution, the beta distribution is the prior distribution. You can express confidence in your estimate of p and q by raising α and β.

POISSON DISTRIBUTION

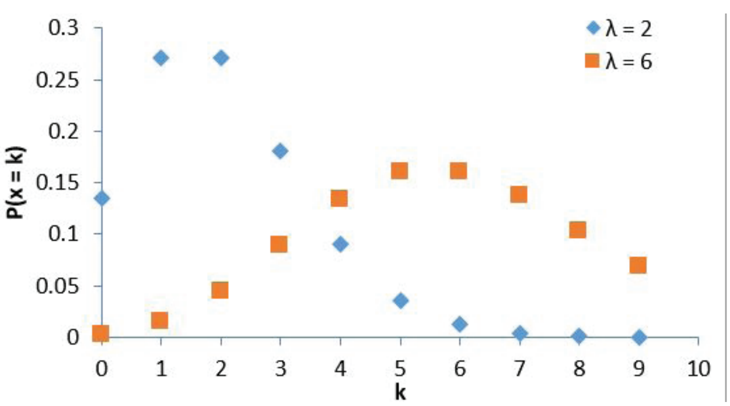
The number of occurrences of a rare event in a large number of trials follows a Poisson distribution.

Examples:

- number of telephone calls at an exchange over a fixed duration
- number of printing errors in a book
- number of winning tickets among those purchased in a large lottery

The p.d.f. of a random variable x that takes on values 0, 1, 2, … , ∞ with a Poisson distribution is given by

- P{x = k} = $e^{-\lambda}\dfrac{\lambda^{-k}}{k!}$
- It follows that. for a Poisson distribution
$$\frac{P_{k-1}}{P_k} = \frac{k}{\lambda}$$
- Increasing for k < λ and decreasing for k > λ



Poisson Example

A spacecraft has 20,000 components. The probability that any one component may fail is 10-4. The mission will fail if 5 or more components fail. How likely is that?

- When n is large and p is small, we can use the Poisson distribution, with λ = np = 20,000*10-4 = 2
- P(x ≥ 5) = 1 – P(x ≤ 4) = 1 – {P(0) + P(1) + … + P4)}
- 1 – e^{-2}(1 + 2 + 2 + 4/3 + 2/3) = 0.052

P.D.F.'S IN BUSINESS RESEARCH

A common example is in modeling human behavior (e.g. consumer behavior).

If I am selling a product, different people will be willing to spend different amounts for the product.

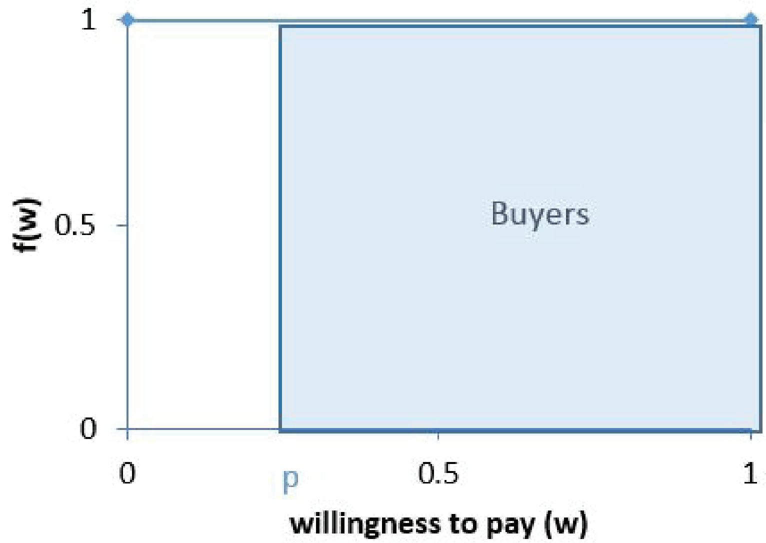
- Often called "willingness to pay"
- Willingness to pay follows a p.d.f.
- Let us consider a simple example, uniform distribution

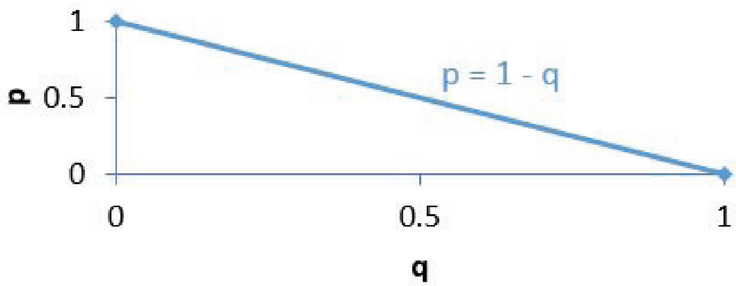
If I sell the product for a price above a consumer's willingness to pay, they won't buy, else they will buy 1 unit

- Then, demand

$$q = \int_{w=p}^{1} dw$$

- 1 – p
  - Downward sloping demand curve





Statistics on Random Variables

If the outcomes of an experiment are probabilistic, what can we "expect" the results to be?

- This is called the expected value or the expectation of the experiment
- The expected outcome after repeated trials of the experiment
- Calculated as the expectation of the p.d.f. associated with the RV

11

EXPECTATION

Given random variable **x**, the expectation of **x**, denoted as E(**x**) is defined as the weighted mean of the p.d.f. of **x.**

- Depending upon how the probabilities are specified, it may be calculated as:

$$E(x) = \begin{cases} \int_{-\infty}^{\infty} x dF(x) \\ \int_{-\infty}^{\infty} x f(x) dx \\ \sum_{\substack{overall\, x| \\ p(x)>0}} x p(x) \end{cases}$$

Example: Suppose you are given even odds on Roulette (double your money if you guessed right). What are the expected winnings by consistently betting on an odd outcome?

Random variable **y** = amount won by betting $1 on odd outcome.

- **y** = 1 if outcome x is odd; **y** = -1 if x is even
- P{**y** = 1} = 18/37; P{**y** = -1} = 19/37
- E[**y**] = 1.P{1} + (-1).P(-1) = -1/37 = -0.027

Expectation Implications

In the Roulette example, you will not lose $0.027/ game.

- You can only win or lose $1.
- Rather, this is the loss/ $ you should expect if you played the game repeatedly.

Same thing happens in the stock market.

- Expectation of a normal distribution = μ (mean)
- Mean annual stock returns (including dividends) during 1960 – 1995 = 11.2%

So, stock returns in most years should be around 11%, right?

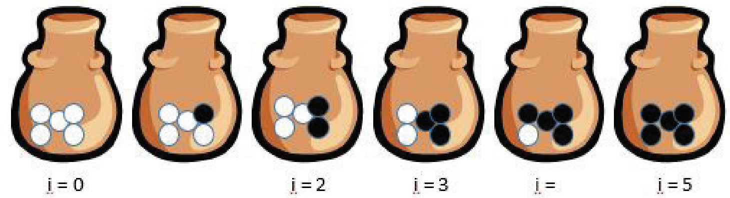
Quoting Warren Buffett's annual report. 2004:

The average return of a distribution isn't a very useful point forecast.

"In one respect, 2004 was a remarkable year for the stock market, ... an investor's return, including dividends, from owning the S&P has averaged 11.2% annually (well above what we expect future returns to be). But if you look for years with returns anywhere close to that 11.2% – say, between 8% and 14% – you will find only one before 2004. In other words, last year's "normal" return is anything but."

CONDITIONAL PROBABILITY

Conditional probability is a way to incorporate additional information into probability estimate.



Say, we have n + 1 urns numbered 0 to n. The i^{th} urn contains i white balls and (n – i) black balls

Question 1:

- Each urn has the same number of balls – n
- Hence, each urn is equally likely to be picked
- Hence, probability the ball came from urn n = $\frac{1}{n}$

Question 2:

- Total number of white balls = 0 + 1 + ... + n = $\frac{n(n+1)}{2}$
- White balls in urn i = i
- Hence, probability = $\frac{2i}{n(n+1)}$

What is the impact of the additional information that the ball picked is white?

- Probability estimate is more precise
- Conditional probability = $\frac{2i}{n(n+1)}$
- Unconditional probability = $\frac{1}{n}$
- Conditional probability =

$$\frac{2 * i * unconditional\ probability}{(n+1)}$$

This estimate is more precise:

The ball is less likely to have come from an urn with few white balls.

Another Example:

Is September through October a good time to own stocks?

- Anecdotally, a terrible time: Sep-Oct ’29 (-40%), Oct 19 ‘87 (-22%), Sep-Oct 02 (-16%), Oct 08 (-25%)
- Unconditional mean only slightly different from 0
- However
  - If S&P 500 on 08/31 is more than 10% below 6-month moving average, Sep-Oct mean return = +2.5%
  - If S&P 500 on 08/31 is more than 10% above 6-month moving average (2009), Sep-Oct mean return = -5.6%

BAYE’S RULE

- A formal rule to calculate conditional probabilities.
- Conditional probability of interest is calculated in terms of other conditional probabilities.
- Useful if various other conditional probabilities are known.
- Given that the number of times in which an unknown event has happened and failed: Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probabilities that can be named.

  E.g. if 10 out of a sample of 100 pins are defective, what is the probability that the total output of pins will contain between 9% and 11% defectives?

We often like to know the probability that A_i occurs, given the additional information that B has occurred.

- Written as $P(A_i \mid B)$
- e.g. which urn (A_i) did the ball come from, given that the ball drawn was white (B)
- or, the probability that a person who tests positive for a disease actually has the disease
- Note that this limits the sample space to those events which also include B

Say the elementary outcomes are equally likely

- If A can occur in n_A ways, B can occur in n_B ways, & both can occur in n_{AB} ways
- Given than B has occurred, A can only occur in n_{AB} ways (m)
- B occurs in n_B ways (n)
- Then, P(A | B) = $$\frac{n_{AB}}{n_B} = \frac{n_{AB}/n}{n_B/n} = \frac{P(AB)}{P(B)}$$
- So, $P(A_i \mid B) = P(A_iB)/ P(B)$
- Consider the numerator, $P(A_iB)$
  - Using the same logic as before
  - $P(B \mid A_i) = P(A_iB)/ P(A_i)$

    therefore
  - $P(A_iB) = P(B \mid A_i).P(A_i)$
- Now consider the denominator, P(B)
  - If all the ways A_i, in which outcome B can be obtained, satisfy a special property
  - they are mutually exclusive and collectively exhaustive (MECE)
  - Then we can write $P(B) = P(B|A_1).P(A_1) + \ldots + P(B|A_n).P(A_n)$ = $$\sum_{i=1}^{n} P(B \mid A_i).P(A_i)$$
- Putting the two together
- This is called Bayes’ rule

- A_i are often called hypotheses, or causes
- $P(A_i)$ is the apriori probability of hypothesis A_i
- $P(A_i \mid B)$ is the aposteriori probability of A_i given B

Bayes' Theorem

Bayes' theorem is often used to estimate the probability of a hypothesis or cause.

- The problem is that the unconditional (apriori) probabilities A_j are rarely known
- E.g. P(Sunny) = 0.9; P(Cloudy) = 0.1

Acutal Weather	Forecast Sunny (F_s)	Forecast Cloudy (F_c)	Forecast Iffy (F_i)
Sunny	0.8	0.1	0.1
Cloudy	0.4	0.4	0.2

- If the forecast is sunny, the probability it will actually be sunny =

$$P(S \mid F_s) = \frac{P(S).P(F_s \mid S)}{P(S).P(F_s \mid S) + P(C).P(F_s \mid C)} = \frac{0.72}{0.76} \cong 0.95$$

Bayes' Rule Example:

- Say Moffitt develops a test to detect cancer, with P(T | C) = 0.95, P(T' | C') = 0.95
  - i.e. if a tested patient has cancer, the test will detect it with 95% probability
  - if the patient does not have cancer, the test will report no cancer with 95% probability
- What is P(C | T)?
  - i.e. if the test result indicates cancer, what is the probability that the patient actually has cancer?

$$P(C \mid A) = \frac{P(A \mid C).P(C)}{P(A \mid C).P(C) + P(A \mid C').P(C')}$$

- Say, P(C) = 0.005

$$P(C \mid A) = \frac{(0.95).(0.005)}{(0.95).(0.005) + (0.05).(0.995)} = 0.087$$

Then

- i.e. though the test is highly accurate, only in 8.7% of the cases, will a patient who tests positive for cancer actually have cancer
  - But this is a great improvement over unconditional probability of 0.005

REFERENCES

Edward O. Thorpe, Elementary probability, 1966, LOC: 66-16130

Emanuel Parzen, Modern probability theory and its applications, 1960, LOC: 60-6456

Athanasios Papoulis and S. Unnikrishna Pillai, Probability, random variables and stochastic processes, 4th Ed, 2002, ISBN: 0-07-366011-6

Jean Tirole, The Theory of Industrial Organization, 1988, ISBN: 0-26-220071-6

Peter L. Bernstein, "Against the gods: The remarkable story of risk," 1996, ISBN: 0-471-12104-5

Trevor Hastie, Robert Tibshirani, and Jerome Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2001, ISBN: 0387952845

Geoffrey J. McLachlan and Thriyambakam Krishnan, The EM algorithm and extensions, 1997, ISBN: 0471123587