Permutations & Combinations

Permutation: an arrangement of *n different objects*. For example, if we have 3 objects, A, B and C, then the possible arrangements are ABC, ACB, BAC, CAB, CBA, and BCA. Basic Rule: for **n** elements, the total number of arrangements is $a_1\times a_2\times\ldots\times a_n$. **Example**: Post Code. Assume Letter, Digit, Digit. Then the total number of *locations* is 26×10×10 = 2600. For choosing **n** elements to fit into n spaces, for the first *position* we have n choices; for the *2nd* position we have (n-1) choices; ...., for the *nth* position we have 1 choice. So the total number of **permutations** is n(n-1)(n-2)...2×1 = n!. **Convention:** 0! = 1.

Permutations of n elements taken r at a time. The total number is n(n-1)...(n-r+1) [where (n-r+1) is r]. So we **have** ${}^{n!}/_{(n-r)!}$ elements. Example: If **repetitions** are not permitted, (i) How many *3 digit numbers* can be formed from the six digits 2, 3, 5, 6, 7, 9? A: n = 6, r = 3, so ${}^{n!}/_{(n-r)!} = {}^{6\times5\times4\times3\times2\times1}/_{3\times2\times1} = 120$. (ii) How many of these are **smaller** than 400? A: Take case “1st position = 2”. Now n = 5, r = 2, so ${}^{n!}/_{(n-r)!} = {}^{5!}/_{3!} = 20$. Now take **case** “1st position = 3”. 20 possibilities again. So the total number under 400 is 40. (iii) How many of *these are even?* A: Must end in a “2” or a “6”, so again **40** possibilities.

Assume that the n elements are *not all different*. Example: In how many ways can we arrange the letters of ZOONOOZ? n = 7, Permutations = 7!. But we have **2 Z’s and 4 O’s**. So the number of permutations is ${}^{7!}/_{2!\times4!} = 105$.

Combinations

A combination is a **selection** of elements in which order does not matter. We want to select *r objects from n different objects*. The number of permutations is ${}^{n!}/_{(n-r)!}$. If r = 3 and n = 6, then $x_1x_5x_3$ and $x_5x_3x_1$ are different **permutations** — but the **same combination**. The number of *combinations* is ${}^{n!}/_{(n-r)!r!}$. This is the binomial *coefficient*, $\binom{n}{r} = {}^{n!}/_{(n-r)!r!}$.

Example: In how many ways can a committee of *3 women and 2 men* be chosen from 7 women and 5 men? Women: $\binom{7}{3} = {}^{7\times6\times5}/_{1\times2\times3} = 35$ ways. Men: $\binom{5}{2} = {}^{5\times4}/_{1\times2} = 10$ ways. So total **number** = 35×10 = 350 ways.

The Binomial Theorem

$(a+b)^n = \sum_{r=0}^{n}\binom{n}{r}a^{n-r}b^r$. Example: $(a+b)^3 = \binom{3}{0}a^3b^0 + \binom{3}{1}a^2b + \binom{3}{2}ab^2 + \binom{3}{3}a^0b^3 = {}^{3!}/_{0!(3-0)!}a^3 + {}^{3}/_{1}a^2b + {}^{3\times2}/_{1\times2}ab^2 + {}^{3!}/_{0!3!}b^3 = a^3 + 3a^2b + 3ab^2 + b^3$. **Important**: $\binom{n}{r} = {}^{n!}/_{(n-r)!r!} = {}^{n!}/_{(n-r)!(n-(n-r))!} = \binom{n}{n-r}$. Example: A **student** is to answer 8 out of 10 exam questions. How many choices? A: n = 10, r = 8, so $\binom{n}{r} = \binom{10}{8} = \binom{10}{2} = {}^{10\times9}/_{1\times2} = 45$. If he answers **the** first 3 questions, how many choices? A: n = 7, r = 5 so $\binom{n}{r} = \binom{7}{5} = \binom{7}{2} = 21$. Q: If he must answer **at least 4 out of the first 5 questions**, how many choices? A: Case (i): He answers exactly 4 out of the first 5: n = 5, r = 4 so $\binom{5}{4} = \binom{5}{1} = 5$. For the 2nd part (remaining 5) he has **4** choices: n = 5, r = 4, ..., = 5. So the total number of choices is 25. Case (ii): Answers 5 out of the first 5. No choice in 1st part. For the 2nd part, he chooses 3, so n = 5, r = 3 gives $\binom{5}{3} = \binom{5}{2} = 10$ choices. So the **total** number of choices here is 10; Grand total of the number of possible choices = 10+25 = 35.

Subsets

Q: Find the number of **possible subsets of a set** with n elements. $({}^{n}_{0})$ = the number of subsets with *0 elements*; $({}^{n}_{1})$ = the number of subsets with *1 element;* $({}^{n}_{r})$ = the number of subsets with *r elements*. We have a **summation**: $({}^{n}_{0})+({}^{n}_{1})+...+({}^{n}_{n}) = \sum_{r=0}^{n}\binom{n}{r} = \sum_{r=0}^{n}\binom{n}{r}1^{n-r}1^{r} = (1+1)^{n} = 2^{n}$. **Another** method: $X = \{x_1,x_2,...,x_n\}$. There are 2 choices to indicate whether an **element** belongs to a subset or not for each element of X. Using the *basic* rule, we have 2×2×2×2... (n times) = 2^n.

Probability Theory

The set S of all possible outcomes of some **random** experiment is called the *sample space*. Example: "Toss a coin twice". Here, the sample space S is {HH, HT, TH, TT}. A particular outcome is called a sample **point**. An event A is any set of outcomes (any subset of S). The event {a}, $a \in S$, is called an *elementary* event.

Example: {HH} is an elementary event; {HH, HT, TH} = *"at least 1 head"*. The empty set ϕ is called the **impossible** event. The set S is called the **certain** event. The collection of all the event subsets of S is called the **Borel field**, ε. ε is the *power* set of S. $|\varepsilon|$ is the cardinality of a set (ε) and is the number of **elements** of the set. (Here of the set ε). Note: $|\varepsilon|$ is 2^n if $|S| = n$.

Tutorial

Q: How many ways can we select a **chairperson**, **vice-chairperson, secretary** and **treasurer** from a group of 18 people? A: Using the basic rule, we have 18×17×16×15 = 73440 choices. Using *Permutations*, $^{18!}/_{(18-4)!} = {}^{18!}/_{14!}$. Q: If *3 flags are available* and all are to be used, how many **signals** can be sent? A: Permutations = $^{n!}/_{(n-r)!} = {}^{3!}/_{(3-3)!} = {}^{3!}/_{0!} = {}^{3!}/_{1} = 6$ signals. Q: If at **least** 2 flags must be raised, how many *signals* can be sent? A: For 3 flags raised, we have **6** signals. For 2 flags, we have $^{3!}/_{(3-2)!} = {}^{3!}/_{1!} = 6$ signals. So we have 12 different signals in all.

Q: How many **13 card bridge hands** can be dealt from a pack of 52 cards? A: *Order* doesn't matter, so the number of hands is $({}^{52}_{13}) = {}^{52!}/_{39!13!}$. Q: How many *13 card bridge hands* can be dealt containing all 4 aces? A: Take the 4 aces out, so that there are 48 cards left, and select **another** 9 — in $({}^{48}_{9}) = {}^{48!}/_{39!9!}$ different ways. Q: How many deals of 4 hands (13 cards in each) are there? A: **1st** method: take all 52! permutations = $^{52!}/_{13!\times13!\times13!\times13!} = 52!/(13!)^4$. 2nd method: **1st** hand: $({}^{52}_{13})$ ways. **2nd** hand: $({}^{39}_{13})$ ways. **3rd**: $({}^{26}_{13})$ ways. **4th:** $({}^{13}_{13})$ ways. **Total** number of deals = $({}^{52}_{13})({}^{39}_{13})({}^{26}_{13})({}^{13}_{13}) = {}^{52!}/_{39!13!}\times{}^{39!}/_{26!13!}\times{}^{26!}/_{13!13!} = 52!/(13!)^4$.

Q: **Show** that $({}^{2n}_{2}) = 2({}^{n}_{2})+n^2$. A: Take the LHS. $^{2n!}/_{(2n-2)!2!} = {}^{(2n-1)(2n)}/_{2}$. [**Note**: We *know that* 2n! = 2n(2n-1)(2n-2)... The bit in blue onwards **is** (2n-2)!. So, 2n! = 2n(2n-1)(2n-2)!]. So we **have** n(2n-1) after *cancellation*. RHS = $2({}^{n!}/_{(n-2)!2!})+n^2 = 2\ {}^{n(n-1)}/_{1\times2} + n^2$ [**same** trick used] = $n(n-1)+n^2 = 2n^2-n = n(2n-1)$. QED; the *identity* is true.

Q: What is the **coefficient** of x^7y^3 in $(x+y)^{10}$? A: We get it from $({}^{10}_{3}) = {}^{10\times9\times8}/_{1\times2\times3} = 120$. Q: In how many **ways** can a set of n elements be split into *2 nonempty subsets?* Q: Find the number of **possible** permutations of elements AAABC *taken* 3 at a time. These two questions will be discussed **later**.

9th February 1999

Events **A & B** are mutually exclusive if $A \cap B = \phi$ (i.e. A and B are disjoint). We operate with events as with sets. *Random* experiment: “Toss a coin 3 times”. Here, S = {HHH, HHT, HTH, HTT, TTH, THT, TTT, THH}. ($2^3 = 8$ elements). Let A = “2nd throw is a head”. Then A = {HHH, HHT, THH, THT}. Let B = “Exactly 2 Heads”. Then B = {HHT, THH, HTH}. $A \cap B$ = {HHT, THH}. Let C = 2nd throw is a head **and** exactly 2 heads are thrown = $A \cap B$.

Relative Frequency Approach to Probability

Let N_A be the number of times event A **occurs** out of N repetitions of some random experiment. The relative *frequency* of A is $f_N(A) = N_A/N$. (1) $0 \le f_N(A) \le 1$. (2) $f_N(\phi) = 0$ ($N_\phi = 0$) and $f_N(S) = 1$ ($N_S = N$). (3) If events A and B are **mutually exclusive**, then $f_N(A \cup B) = f_N(A) + f_N(B)$ [$A \cap B = \phi$].

Relative Frequency Principle

If in a **large** number of repetitions, $f_N(A)$ **stabilises** at some number, (converges), then this number is called “the *probability* of A”. $P(A) = \lim_{N \to \infty} f_N(A)$. How **large** should N be? This is mathematically difficult to **handle** because of the “lim”. (The limit).

Axiomatic Definition of Probability (Kolmogorov)

Let S be a *sample space* and let ε be its Borel field. Let P be a **real** valued function defined on ε: P: $\varepsilon \to$ ‘. ($P(A) \in$ ‘, $A \in \varepsilon$). P is called the **probability measure**, and P(A) is called the **probability of event A**, if (i) $\forall A \in \varepsilon,\ 0 \le P(A) \le 1$; (ii) $P(S) = 1$, $P(\phi) = 0$; (iii) if A and B are *mutually exclusive events*, then $P(A \cup B) = P(A) + P(B)$. The triple (S,ε,P) is called a probability **space**.

Consequences. (1) $P(\bar{A}) = 1 - P(A)$, $\forall A \in \varepsilon$. $A \cup \bar{A} = S$, $A \cap \bar{A} = \phi$ ($\Rightarrow A, \bar{A}$ are mutually **exclusive** events). (3) $\Rightarrow P(A \cup \bar{A}) = P(A) + P(\bar{A})$. (2) $\Rightarrow P(S) = 1$. *Therefore*, $P(A) + P(\bar{A}) = 1$.

(2) For any A, B $\in \varepsilon$, $P(A \cup B) = P(A) + P(B \setminus A)$, **and** $P(B \setminus A) = P(B) - P(A \cap B)$. Proof: Consider a *Venn diagram*. $A \cup B = A \cup (B \setminus A)$; $A \cap (B \setminus A) = \phi$. **Then** $P(A \cup B) = P(A \cup (B \setminus A)) = P(A) + P(B \setminus A)$. Now $P(B) = P(B \setminus A) + P(A \cap B)$. **Prove** this. **We show that** $B = (B \setminus A) \cup (A \cap B)$; $(B \setminus A) \cap (A \cap B) = \phi$. These *imply that* $P(B) = P(B \setminus A) + P(A \cap B)$. (3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Corollary. $P(A) + P(B) \geq P(A \cup B)$. (4) If $A \subseteq B$, then $P(A) \le P(B)$. Proof. $A \cup B = B$. $P(B) = P(A) + P(B \setminus A)$ from **consequence** (2); and $0 \le P(B \setminus A) \le 1$. So $P(B) \geq P(A)$. If A = B, this *implies that* P(B) = P(A).

10th February 1999

Finite or Countable Probability Spaces

Let S be **countable** (or finite), S = {e_1, e_2, ...}. {e_i} = an **elementary** event, where i = 1,2,.... Assign to each *elementary event* a probability measure p_i such that (i) $p_i \geq 0$ for all i; (ii) $\sum_i p_i = 1$. The **probability** of an event A, P(A), is the sum of the probabilities of the points of A. Example: Let A = {e_3, e_7, e_1}. So A = $\{e_3\} \cup \{e_7\} \cup \{e_1\}$, with $\{e_i\} \cap \{e_j\} = \phi$. **Therefore**, (by axiom 3), P(A) = P({e_3}) + P({e_7}) + P({e_1}) = $p_3 + p_7 + p_1$.

Example: 4 Horses a, b, c and d run a race. Let {a} = “horse a wins”, and let S = {a,b,c,d}. (1) a is twice more likely to *win* than b; (2) b is **twice** more likely to win than c; (3) a is **three** times more likely to win than d. Q: Find the probability *measures* p_a, p_b, p_c and p_d. A: Let $p_c = p$, so that $p_b = 2p$, $p_a = 4p$, and $p_d = {}^{4p}/_3$. Now $p+2p+4p+{}^{4p}/_3 = 1$; so $p = {}^3/_{25}$. Therefore, $p_a = {}^{12}/_{25}$, $p_b = {}^6/_{25}$, $p_c = {}^3/_{25}$, and $p_d = {}^4/_{25}$.

Finite Equiprobable Spaces

Let S be finite with n points, S = {e_1,...e_n}. Assign to each **elementary** event {e_i} the probability $^1/_n$. If A is an event **containing** r points out of S, then P(A) = $^1/_n+{}^1/_n+{}^1/_n+$... (r times) = $^r/_n$ = $^{\text{number of points in A}}/_{\text{number of points in S}}$. P(A) = $^{\text{number of ways A can occur}}/_{\text{number of points in the sample space}}$. The classical **definition** of probability is *associated* with games of chance.

Example: We have a box of 12 balls: 8 white and 4 black. 2 balls are chosen at random without replacement. (i) What is the *probability* that the 2 balls have the same colour? (ii) What is the probability that at **least** 1 is black? A: (i) The sample space S has $(^{12}{}_2)$ equally likely outcomes = 66 outcomes. Let A = “All black pairs”. |A| = $(^4{}_2)$ = 6. So there are 6 ways to *choose* 2 black balls from 4.

Let B = “All **white** pairs”. |B| = $(^8{}_2)$ = 28. P(A) = $^{|A|}/_{|S|}$ = $^6/_{66}$. P(B) = $^{|B|}/_{|S|}$ = $^{28}/_{66}$. What is $P(A \cup B) = P(A)+P(B)$? [We have this formula as here, $A \cap B = \phi$]. $P(A \cup B) = {}^6/_{66}+{}^{28}/_{66} = {}^{34}/_{66}$. (This is P(Two balls are the same colour)). (ii) Find P(At **least** 1 is black). Now B = “*All white pairs*”, so $\bar{B}$ = “At least 1 is black”. Therefore, $P(\bar{B}) = 1-P(B) = 1-{}^{28}/_{66} = {}^{38}/_{66}$.

Q: Solve the above problem if repetitions *are* permitted. A: The sample space has $(^{12}{}_1)\times(^{12}{}_1)$ = 144 choices. Let A = two **balls** are white. |A| = $(^8{}_1)\times(^8{}_1)$ = 64 (**8** choices in 1st position; **8** choices in 2nd position). So P(A) = $^{|A|}/_{|S|} = {}^{64}/_{144}$. $P(\bar{A}) = 1-{}^{64}/_{144} = {}^{80}/_{144} = {}^5/_9$. And P(**same** colour balls) = $(^8{}_1)\times(^8{}_1)+(^4{}_1)\times(^4{}_1) = 64+16 = {}^{80}/_{144}$.

Tutorial

Q: **Using** $P(A \cup B) = P(A)+P(B)-P(A \cap B)$, find *an expression for* $P(A \cup B \cup C)$. A: $P(A \cup B \cup C) = P((A \cup B) \cup C) = P(A \cup B)+P(C)-P((A \cup B) \cap C) = P(A) + P(B) - P(A \cap B) + P(C) - (P(A \cap C) \cup P(B \cap C)) = P(A) + P(B) + P(C) - P(A \cap B) - ((P(A \cap C) + P(B \cap C)-P(A \cap C) \cap (B \cap C)) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B) \cap C = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.

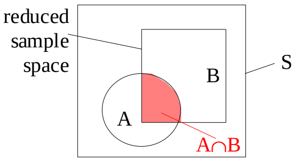
Q: **2 men & 3 women are in a race**. Those of the *same* sex have the same probability of winning, but each man is twice as likely to win as each woman. (a) Find P(particular man to win) and P(particular woman to win). (b) Find P(Race is won by a woman). (a) **Define** p = P(woman wins) and $2p$ = P(man wins). So $p+p+p+2p+2p = 1$; $7p = 1$, $p = {}^1/_7$. So P(woman wins) $= {}^1/_7$, and P(man wins) $= {}^2/_7$. (b) P(race is won by a woman) $= {}^3/_7$.

Q: **3 balls** are chosen without replacement from a box with *3 white, 4 black and 3 red* balls. What is the probability that the three balls have different colours? A: The number of **ways** to select 3 balls is $({}^{10}{}_3) = 120$. Let A = *"Different colours"*. We have 3 positions to choose from. Consider black, white, and red. For the **1st** position, we have 4 choices; 2nd position = 3 choices; 3rd position = 3 choices. But the order doesn't matter, so we have $4\times3\times3$ choices = 36 **choices**. So $|A| = 36$, $|S| = 120$, and $P(A) = {}^{36}/_{120} = {}^{3}/_{10}$.

16th February 1999

Conditional Probability

For the **events** A and B, $(P(B) > 0)$, the *conditional* probability of A given that B has occurred is $P(A|B) = {}^{P(A\cap B)}/_{P(B)}$. This is the measure of the *relative* probability of A with respect to the reduced space B. Recap: $P(B) = {}^{|B|}/_{|S|}$. (Finite). Using the definition, $P(A|B) = {}^{(|(A\cap B)|/|S|)}/_{(|B|/|S|)} = {}^{|A\cap B|}/_{|B|}$.



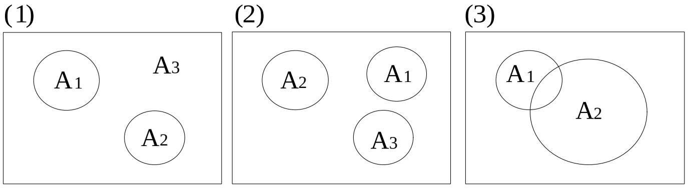
Example: A *pair* of fair dice are tossed. If the sum is 6, find the probability that one of the dice was two. Let B = "the **sum** is 6"; A = "one of the dice is 2". Now $|S| = 36$, $B = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$; $|B| = 5$. So $A\cap B = \{(2,4), (4,2)\}$; $|A\cap B| = 2$, so $P(A|B) = {}^{|A\cap B|}/_{|B|} = {}^2/_5$.

Multiplication Theorem for Conditional Probability

If $P(A|B) = {}^{P(A\cap B)}/_{P(B)}$, $(P(B) > 0)$, then $P(A\cap B) = P(A|B)P(B) = P(B|A)P(A)$, $(P(A) > 0)$. Let $\{A_1, A_2, \ldots, A_n\}$ be a set of events, with $P(A_1\cap\ldots\cap A_{n-1}) > 0$. Now $P(\cap^n{}_{i=1}) = P(A_1\cap A_2\cap\ldots\cap A_n) = P(A_1).P(A_2|A_1).P(A_3|A_1\cap A_2)\ldots P(A_n|A_1\cap A_2\cap\ldots\cap A_{n-1})$. **Note**: Try to prove the theorem by using the *formula* for P(A|B) and *induction*. Proving by induction: (1) Prove that the proposition **holds** for n = 1,2,3 (small n). (2) **Assume** that the proposition is true for n-1. (3) Prove that the proposition **holds** for n.

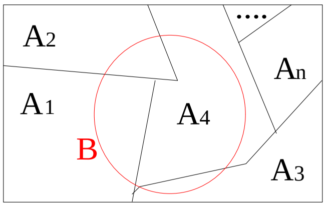
Partitions & Bayes Theorem

A **partition** of the sample space S is a set of events $\{A_1, \ldots, A_n\}$ such that (1) $\cup^n{}_{i=1} A_i = A_1\cup A_2\cup\ldots\cup A_n = S$; (2) $A_i\cap A_j = \phi$ for all $i, j \in \{1,\ldots,n\}$, $i \neq j$ (Note: $A\cap A = A$). Example: $\{A_1, A_2, A_3\}$ could be **represented** by diagram (1). Diagrams (2) and (3) are not *partitions*. Bayes Theorem: If $A_1, \ldots, A_n$ is a **partition** of the sample space S, and B is any other event, then for any $i \in \{1,2,\ldots,n\}$, $P(A_i|B) =$



$$\frac{P(A_i).P(B|A_i)}{P(A_1).P(B|A_1)+P(A_2).P(B|A_2)+\ldots+P(A_n).P(B|A_n)}.$$

Let B = $(A_1 \cap B) \cup (A_2 \cap B) \cup \ldots \cup (A_n \cap B)$. **Notice** $(A_i \cap B) \cap (A_j \cap B)$ = using *Associativity* and *Commutativity* = $(A_i \cap A_j) \cap (B \cap B) = \phi \cap B = \phi$. **Now** $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_n \cap B)$; $P(A_i|B) = P(A_i \cap B)/P(B)$ = using the red identity on the *previous* page = $\frac{P(B|A_i).P(A_i)}{P(B|A_1).P(A_1)+\ldots+P(B|A_n).P(A_n)}$.



17th February 1999

Conditional Probability: Examples

Q: A box **contains** 12 eggs, 4 of which are bad. We *select* eggs without replacing until we find a bad egg. What is the probability that we stop after 3 eggs? A: Let A_1 be "first egg is OK"; A_2 be "second egg is OK"; and A_3 be "third egg is *bad*". We want $P(A_1 \cap A_2 \cap A_3) = P(A_1).P(A_2|A_1).P(A_3|A_1 \cap A_2)$. **Now** $P(A_1) = {}^{8}/_{12}$; $P(A_2|A_1) = {}^{7}/_{11}$; and $P(A_3|A_1 \cap A_2) = {}^{4}/_{10}$. So $P(A_1 \cap A_2 \cap A_3) = {}^{8\times7\times4}/_{12\times11\times10} = {}^{28}/_{165}$.

Q: **Machine** A produces 50% of the *total* production, and 3% of its output is defective. B: 30%, 4%; C: 20%, 5%. If an item is selected at **random**, what is P(It is defective)? If an item selected at random *is* defective, what is the probability that it came from machine C? A: Let S = all produced items, {A,B,C}. This is a **partition** on S. Let X = "The item is *defective*". P(X) = ? Now $X = (A \cap X) \cup (B \cap X) \cup (C \cap X)$; $P(X) = P(A \cap X) + P(B \cap X) + P(C \cap X)$. We *know that* $P(X|A) = 0.03$; $P(X|B) = 0.04$; and $P(X|C) = 0.05$. Now using $P(A \cap X) = P(X|A).P(A)$ [$P(A) > 0$], we get $P(X) = P(X|A)P(A) + P(X|B)P(B) + P(X|C)P(C) = 0.03\times0.5 + 0.04\times0.3 + 0.05\times0.2$; $P(X) = 0.037$. Now we **want** $P(C|X) = {}^{P(X|C)P(C)}/_{P(X|A)P(A)+P(X|B)P(B)+P(X|C)P(C)}$. The *denominator* here is P(X); so $P(C|X) = {}^{P(X|C)P(C)}/_{P(X)} = {}^{0.05\times0.2}/_{0.037} = {}^{10}/_{37}$.

	White	Black	Red
A	3	2	5
B	4	1	3
C	5	2	4

Q: 3 boxes contain *coloured* balls as shown. A box is selected at **random** and 3 balls are drawn without replacement from it. Find the probability that the box was C, *given* that the 3 balls are (a) one of **each** colour; (b) all of the **same** colour. A: (a) Let E = "one of *each* colour". We want P(C|E). Here, $S = \{(b_1b_2b_3, \ldots, (b_ib_jb_k)\}$. This is **not** an *equiprobable* sample space. Now {A,B,C} is a partition on S. $P(A) = {}^{1}/_{3}$, $P(B) = {}^{1}/_{3}$, $P(C) = {}^{1}/_{3}$. Now $P(C|E) = {}^{P(E|C)P(C)}/_{P(E|A)P(A)+P(E|B)P(B)+P(E|C)P(C)}$. Now $P(E|A) = {}^{30}/_{120} = ¼$. [**Choices** of 3 balls out of $10 = ({}^{10}{}_{3}) = 120$; choices of 3 balls of different colours = 3×2×5 = 30]. So $P(E|B) = 4\times1\times3/({}^{8}{}_{3}) = \ldots$, and $P(E|C) = 5\times2\times4/({}^{11}{}_{3}) = \ldots$ We now have **enough** information to find an answer for the expression *shown* above.

18th February 1999

Set Theory

U is the **universal** set. ϕ is the **empty** set. A^c is the *complement* of A, where $A \subset U$. $\cup$, $\cap$, complement and \ are set **operations**. Rules: $A \cup A^c = U$ (The law of *excluded middle*). $A \cap A^c = \phi$ (The *non-contradiction* principle). $A \cup U = U$; $A \cup \phi = A$; $A \cap U = A$, $A \cap \phi = \phi$. [$A \cap A = A$, $A \cup A = A$ — **Idempotence**]. [$A \cap B = B \cap A$; $A \cup B = B \cup A$ — **Commutativity**]. **Associativity**: $(A \cap B) \cap C = A \cap (B \cap C)$ and $(A \cup B) \cup C = A \cup (B \cup C)$. Mutual **Distributivity**: $(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$ and $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$. $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$ are *de Morgan's laws*.

Past Exams: Questions & Answers

Remember that $P(A\backslash B)$ is $P(A\cap\bar{B})$. Q: **Three** die are tossed. What is the probability that all faces will *show* the same number? What is the probability that at least one side will show a 3? A: $|S| = 6\times6\times6 = 216$, an **equiprobable** space. Let A = "The same number" = {(1,1,1), (2,2,2), ... (3,3,3)}. $|A| = 6$. So $P(A) = {}^{|A|}/_{|S|} = {}^{6}/_{216} = {}^{1}/_{36}$. Now let B = "*none* of the faces is a 3". $\bar{B}$ = "At **least** one face is a 3", with $P(\bar{B}) = 1-P(B)$. $|B| = ?$ The number of *possible* elements of B is $5\times5\times5 = 125$ (5 **choices** in 1st position, ...) So $P(B) = {}^{125}/_{216}$, and $P(\bar{B}) = 1-{}^{125}/_{216} = {}^{91}/_{216}$.

	Nut	Cream	Caramel
A	1	4	3
B	4	3	1
C	3	4	1

Q: Bags **contain** sweets as shown. A bag is chosen at random; and 2 sweets are taken *without* replacement. Find P(Bag was C, given the choice was one nut and one caramel). And, if the sweets are replaced, find P(Bag chosen was A given the choice is **cream** and **caramel** centre). A: $P(A) = P(B) = P(C) = {}^{1}/_{3}$. (Chosen at *random*). Let X = "One nut & one caramel". $P(C|X) = ?$: from Bayes' Theorem: $P(C|X) = {}^{P(X|C)P(C)}/_{P(X|A)P(A)+P(X|B)P(B)+P(X|C)P(C)}$. (**No** replacement). The total number of choices of 2 out of 8 sweets (for any box A, B or C) is $({}^{8}{}_{2}) = 28$.

$P(X|A) = ?$ One **nut** and one **caramel** can be chosen from box A in 1×3 ways (each nut can be obtained with each caramel). $P(X|A) = {}^{3}/_{28}$. *Similarly*, $P(X|B) = {}^{4\times1}/_{28} = {}^{4}/_{28}$; and $P(X|C) = {}^{3\times1}/_{28} = {}^{3}/_{28}$. Then $P(C|X) = {}^{1/3\times3/28}/_{(1/3\times3/28)+(1/3\times4/28)+(1/3\times3/28)} = {}^{3}/_{10}$. Now let X = "One *cream* and one *caramel*, with replacement". Here, $P(X|A) = {}^{3\times4}/_{8\times8}$ (**with** replacement) $= {}^{12}/_{64}$. And $P(X|B) = {}^{3\times1}/_{8\times8} = {}^{3}/_{16}$; and $P(X|C) = {}^{4\times1}/_{8\times8} = {}^{4}/_{64}$. So $P(A|X) = {}^{(1/3\times12/64)}/_{(1/3\times12/64)+(1/3\times3/64)+(1/3\times4/64)} = {}^{12}/_{19}$.

Q: A **screening** procedure for a disease has +ve and -ve results. If an infected person is tested, P(+ve result) = 0.9. If a *non-infected* person is tested, P(+ve result) = 0.05. In a population where 20% of people are infected, what is P(A person whose test is +ve is in fact **not** infected)? A: $P(+|I) = 0.9$, $P(+|NI) = 0.05$, $P(I) = 0.2$, $P(NI|+) = ?$ Now from **Bayes'** Theorem, $P(NI|+) = {}^{P(+|NI)P(NI)}/_{P(+|NI)P(NI)\,+\,P(+|I)P(I)} = {}^{0.05\times(1-0.2)}/_{0.05\times(1-0.2)\,+\,0.9\times0.2} = {}^{0.04}/_{0.04+0.18} = {}^{2}/_{11}$.

23rd February 1999

Q: There are **3** die. The 1st one has *3 times more chance* to show an odd number than an even number. The 2nd and the 3rd have 3 times more chance to show a number greater than 4. If a die is taken at **random**, and the number shown is 5, what is P(Dice was 1st dice)? We *need* $P(\text{1st dice}|5) = {}^{P(5|\text{1st})P(\text{1st})}/_{P(5|\text{1st})P(\text{1st})+P(5|\text{2nd})P(\text{2nd})+\dots}$ Because P(1st) = P(2nd) = P(3rd), then we can **cancel** these, giving ${}^{P(5|\text{1st})}/_{P(5|\text{1st})+P(5|\text{2nd})+P(5|\text{3rd})}$.

1st dice: S = {1, 2, 3, 4, 5, 6}, **not** equiprobable. $3p+p+3p+p+3p+p = 1$; so $p = {}^{1}/_{12}$; and $P(5|\text{1st}) = {}^{3}/_{12}$. 2nd dice (And 3rd): S = {1, 2, 3, 4, 5, 6}, **not** equiprobable. $p+p+p+p+3p+3p = 1$, $p = {}^{1}/_{10}$. So $P(5|\text{2nd}) = {}^{3}/_{10}$ and $P(5|\text{3rd}) = {}^{3}/_{10}$. So now $P(\text{1st}|5) = {}^{3/12}/_{3/12+3/10+3/10} = {}^{5}/_{17}$.

Independence

$P(A|B) = {}^{P(A\cap B)}/_{P(B)}$ $[P(B) > 0]$. By intuition, $P(A|B) = P(A)$ if A and B independent, i.e. B makes no difference. **DEFINITION**: The events A and B in ε are independent iff $P(A\cap B) = P(A)P(B)$. So $P(A|B) = {}^{P(A\cap B)}/_{P(B)} = {}^{P(A)P(B)}/_{P(B)} = P(A)$ $[P(B) > 0]$. And, $P(B|A) = P(B)$. **Identities**: $P(A\cup B) = P(A)+P(B)-P(A\cap B)$ goes to $P(A\cup B) = P(A) + P(B) - P(A)P(B)$. So we only **need** P(A) & P(B) to calculate the union.

Example: A & B are on a shooting exercise. The probability of A *hitting the target* is 0.25. The probability of B hitting the target is 0.4. Q: If A and B fire **together**, what is P(target has been hit at least once)? A: Calculate $P(A\cup B) = P(A)+P(B)-P(A)P(B) = 0.25+0.4 - (0.25\times 0.4) = 0.55$. Q: Are **mutually exclusive** events *dependent*, *independent* or *cannot* say? A: Dependent.

Q: Let A = "A family has children of **both** sexes", and let B = "A family has children at most **one** of which is a boy". Assuming boys & girls are **equally** likely, show that (1) A & B are *independent* events if the family has 3 children; (2) A & B are *dependent* events if the family has 2 children. A: (1) S = {BBB, BBG, BGB, GBB, BGG, GBG, GGB, GGG}. This is **equiprobable**. Now A = {BBG, BGB, GBB, BGG, GBG, GGB}, and B = {GGB, GBG, BGG, GGG}, so $A\cap B$ = {GGB, GBG, BGG}. Therefore, $P(A) = {}^{|A|}/_{|S|} = {}^{6}/_{8}$, $P(B) = {}^{|B|}/_{|S|} = {}^{4}/_{8}$, and $P(A\cap B) = {}^{|A\cap B|}/_{|S|} = {}^{3}/_{8}$. Now $P(A)P(B) = P(A\cap B)$, (verify), so A and B are **independent**. (2) Here, S = {BB, BG, GB, GG}. A = {BG, GB}, so P(A) = ½. B = {BG, GB, GG}, so P(B) = ¾. And $A\cap B$ = {GB, BG}, so $P(A\cap B)$ = ½. But $P(A)P(B) \neq P(A\cap B)$, so events A and B are *not* independent.

In the previous example, Q: How many **times** should A fire so that $P(HIT) > 0.8$? $P(A\cup A\cup A\cup \ldots)$ — how many? P(A) = 0.25. What we *calculate* is $P(\bar{A}\cap\bar{A}\cap\bar{A}\cap\ldots)$ (n times). This **should** be less than 0.2. So $P(\bar{A})P(\bar{A})P(\bar{A})\ldots$ n times < 0.2; $(1-P(A))(1-P(A))(1-P(A)) < 0.2$; $0.75^n < 0.2$, which we **solve** to get n.

25th February 1999

Q: There are **9 one-dollar bills** and **1 ten-dollar bill** in one box; the other box contains **5 one-dollar bills, 3 ten-dollar bills and 7 five-dollar bills**. What is P(Draw a 10 dollar bill?). A: P(Draw a 10 dollar bill from *1st* box) = $^{1}/_{10}$. P(Draw a 10 dollar bill from **2nd** box) = $^{1}/_{5}$. So P(Draw a 10) = P(10|1st) + P(10|2nd) = $({}^{1}/_{10}\times{}^{1}/_{2})+({}^{1}/_{5}\times{}^{1}/_{2}) = {}^{1}/_{20}+{}^{1}/_{10} = {}^{3}/_{20}$.

Q: An **urn** contains *3 red marbles and 7 white* marbles. A marble is drawn, and one of the other colour is put back in. A second marble is then drawn. If two marbles are chosen, what is P(both are white)? A: **Think** of it as a tree diagram situation *as* shown. Then P(Both white) = ${}^{7}/_{10}\times{}^{6}/_{10} = {}^{21}/_{50}$.

P(1st W) = $^{7}/_{10}$ P(2nd W|1st W) = $^{6}/_{10}$
P(2nd R|1st W) = $^{4}/_{10}$
P(1st R) = $^{3}/_{10}$ P(2nd R|1st R) = $^{2}/_{10}$
P(2nd W|1st R) = $^{8}/_{10}$

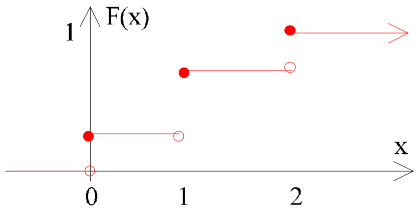
Q: In a **roll** of 5 dice, what is the probability of getting *exactly* 4 faces alike? A: Let "A" be getting 4 alike faces. $|S| = 6\times 6\times 6\times 6\times 6 = 7776$. We can have 4 alike faces with the faces 1's, 2's, etc. For **each** alike face, we have 5 choices for the non-alike face, and 5 **positions** to place them (Note: 1,1,1,1,2 ≠ 1,1,2,1,1). So $|A| = 6\times 5\times 5 = 150$, and $P(A) = {}^{|A|}/_{|S|} = {}^{150}/_{7776} = {}^{25}/_{1296}$.

2nd March 1999

Random Variables

Let <S, ε, P> be a *probability* space. A **One-dimensional random variable X** is a *real-valued* function defined on the sample space S. Example: Toss a coin twice: S = {HH, HT, TH, TT}. Let X = “The **number** of heads”, X: S → **R**; where X(HH) = 2, X(TH) = 1. X is a random variable on S. Let X be a *random* variable on S (<S, ε, P>). The cumulative distribution function F of X is defined for all real values by F(x) = P(X≤x), for all **real** x.

Example: the same **experiment** as before. S is an *equiprobable* space, where P({HH}) = P({HT}) = P({TH}) = P({TT}) = ¼. Let A_0 = “Number of heads < 0” = the **empty** set ϕ. $P(A_0)$ = 0. Let A_0^* = “Number of heads ≤ 0” = {TT}. $P(A_0^*)$ = ¼. But A_0^* = X ≤ 0; so $P(A_0^*)$ = P(X ≤ 0) = F(0) = ¼. Now **let** A_1 be “X ≤ 1” = {HT, TH, TT}. So $P(A_1)$ = P(X≤1) = F(1) = ¾. Note: F(0.82) = ¼; F(2) = P(X≤2) = 1.



F is a **function** mapping **R** onto [0,1] as shown. X: S → **R** F: **R** → [0,1]. Theorem: For all **real** a and b, where a < b, P(a<X≤b) = F(b)-F(a). This *gives* the probability of X being in the interval (a,b]. Proof: **Define** A = “X≤a”, B = “a<X≤b”. So A∪B = “X≤b”; A∩B = ϕ — so *mutually exclusive*. So P(A∪B) = P(A)+P(B); P(X≤b) = P(X≤a) + P(a<X≤b); P(a<X≤b) = F(b)-F(a). Proof **done**. **Example**: (The same situation): P(1≤X≤2) = $^{|\{HT,TH,HH\}|}/_4$ = ¾. P(1<X≤2) = F(2)-F(1) = $1-{}^3/_4$ = ¼.

Types of 1-Dimensional Variables

Discrete Random Variables

X takes only **finite** or a **countable** number of different values $\{x_1, x_2, x_3, \ldots\}$. *Denote* $P(x_i) = P(X = x_i)$. Theorem: (1) $P(x_i) \geq 0$ for i=1, ...; (2) $\sum_i P(x_i) = 1$. We call $P(x_i)$ the probability mass function of X (the pmf). F(x) = P(X≤x). Now “X ≤ x” is e.g. $\{x_1, x_2, x_3, x_4\}$, where P(X≤x) = $\sum_{x_i \leq x} P(x_i)$.

Example (the same as before): X ∈ {0,1,2}. P(0), P(1), P(2) = ¼, ½, ¼. The **sum** of these is 1. F(1.6) = P(0)+P(1) = ¾. Important *examples* of discrete random variables (r.v.): **Bernoulli**: A random experiment that has only *2 possible outcomes* — success (probability p) and failure (probability 1-p = q).

3rd March 1999

Binomial

We have n **repeated** Bernoulli trials. X is the number of successes, X = 0, 1, 2, ..., n. Assume the trials are *independent* and the probability p is *constant*. The probability of a single outcome is e.g. {s, f, f, s, f, s, ..., f}. For n **trials**, r **successes,** we have p.q.q.p......= p^rq^{n-r}. So $P(X=r) = {}^{n!}/_{(n-r)!r!} \times p^rq^{n-r}$; $P(X=r) = ({}^n{}_r)p^rq^{n-r}$. Is this a pmf? Now (i) $P(r) \geq 0$, so this is **correct**. (ii) Is $\Sigma^n_{r=0}\, p(r) = 1$? Calculating, $\Sigma^n_{r=0}\, p(r) = \Sigma^n_{r=0}({}^n{}_r)p^rq^{n-r} = (p+q)^n = 1^n = 1$. QED.

Example: the probability of a *successful* launch is 0.9. Assuming successive launchings are independent, with constant success of probability, find the probability of (a) exactly **4** successes in 7 trials; (b) 1 or more **failures** in 7 trials. [$n = 7$, $p = 0.9$]. A: (a) $P(X=4) = \binom{7}{4}0.9^{4}0.1^{7-4} = \ldots = 0.02296$. (b) $P(X\leq 6) = 1-P(X=7) = 1-\binom{7}{7}0.9^{7}0.1^{0} = \ldots = 0.5217$.

Negative Binomial

Negative Binomial = repeated Bernoulli trials until we reach a *fixed number of n successes*. Let X = the number of failures prior to the n^{th} success. Assume **independent** trials with **constant** probability of success. The last *Bernoulli* trial is a success, (the n^{th} success), and there are $n+r-1$ trials **prior** to the last success — so there must be $(n-1)$ *successes*.

The probability of $(n-1)$ **successes** prior to the n^{th} success (with r failures) is $\binom{n+r-1}{n-1}p^{n-1}q^{r}$. $P(X=r) = p(r) = \binom{n+r-1}{n-1}p^{n-1}q^{r}.p = \binom{n+r-1}{r}p^{n}q^{r}$ (Note: $\binom{n+r-1}{n-1} = \binom{n+r-1}{n+r-1-(n-1)} = \binom{n+r-1}{r}$). **Example**: A fair coin is tossed until we *obtain* 5 heads. What is the probability that we toss 7 tails before we stop? $P(X=7) = \binom{5+7-1}{7}(\frac{1}{2})^{5}(\frac{1}{2})^{7} = \binom{11}{7}(\frac{1}{2})^{12} = 0.08056$.

4th March 1999

Tutorial

Q: On an **exam**, there are 6 questions, each with 5 choices. If a student chooses an answer at random, with all the questions of *equal weight*, and the pass mark is $66\frac{2}{3}\%$, what is the probability that the student **passes** the exam? A: $p = \frac{1}{5}$, $q = \frac{4}{5}$, $n = 6$. The student needs to answer 4 questions **correctly** to pass. ($\frac{4}{6} = 66\frac{2}{3}\%$). So we *want* $P(X=4) = \binom{6}{4}(\frac{1}{5})^{4}(\frac{4}{5})^{2} = .. = 0.01536$.

Q: A die is **rolled** until a 1 occurs. What is the probability that *more than two tosses* are required? A: Could use the probability of getting (no one's, no one's) $= \frac{5}{6}\times\frac{5}{6} = \frac{25}{36}$, or use P(1 occurs on **1st** toss) $= \frac{1}{6}$; P(occurs on **2nd** toss) $= \frac{5}{36}$; (this comes from a tree diagram — 1st **not** a 6, 2nd a 6); so P(1 doesn't occur on 1st or 2nd toss) $= 1-\frac{1}{6}-\frac{5}{36} = \frac{25}{36}$.

Q: Two dice are rolled **repeatedly**. Find the probability that the third time a 9 is obtained, it occurs on the 7th roll of the two dice. A: This is "find the probability that we get 4 non 9's before we stop". Comparing to the *example* of the last lecture, we then get $P(X=4) = \binom{3+4-1}{4}p^{3}q^{4}$. Now what is p? When rolling **two** die, there are 36 outcomes, 4 of which are 9 (3+6, 4+5, 5+4, 6+3). So $P(X=4) = \binom{6}{4}(\frac{4}{36})^{3}(\frac{32}{36})^{4} = \ldots = 0.01285$.

Q: If a die is thrown **repeatedly**, what is the probability that a 6 is obtained before a 1 turns up? A: We want 6 before 1. P(6 before 1) $= \frac{1}{6} + \frac{4}{6}.\frac{1}{6} + (\frac{4}{6})^{2}.\frac{1}{6} + \ldots = \frac{1}{6}(1+\frac{2}{3}+(\frac{2}{3})^{2}+(\frac{2}{3})^{3}+\ldots)$. The thing in brackets is a **geometric series**; and so it is $\frac{1}{6}.\frac{1}{(1-(2/3))} = \frac{1}{6}.\frac{1}{(1/3)} = \frac{3}{6} = \frac{1}{2}$. Or, let $S = \{1,2,3,4,5,6\}$; $A = \{1 \text{ or } 6\}$. So $P(6|A) = \frac{P(A|6)P(6)}{P(A)} = \frac{1\times 1/6}{2/6} = \frac{1}{2}$.

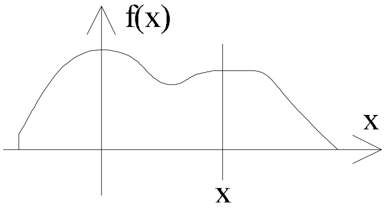
8th March 1999

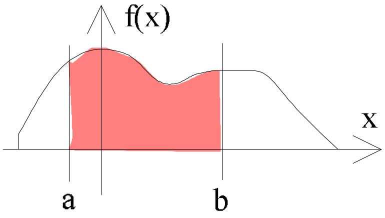
Poisson Distribution

For the Poisson Distribution, $P(X=r) = e^{-\lambda}\frac{\lambda^{r}}{r!}$ ($\lambda > 0$). (1) $P(X=r) \geq 0$. (2) $\Sigma^{\infty}_{r=0}\, e^{-\lambda}\, (\lambda^{r}/r!) = e^{-\lambda}\, \Sigma^{\infty}_{r=0}\, (\lambda^{r}/r!)$. Now $\Sigma^{\infty}_{r=0}\, (\lambda^{r}/r!)$ is e^{λ} (using Taylor series), so $e^{-\lambda}\, \Sigma^{\infty}_{r=0}\, (\lambda^{r}/r!) = 1$.

Continuous Random Variables

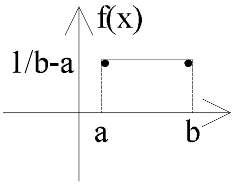
A continuous random variable takes **all** values in **R**. Example: *Human Height*. We use probability density functions, pdf: (1) $f(x) \geq 0$ for all real x; (2) $\int_{-\infty}^{\infty} f(x)dx = 1$. We write for this, $\int_{-\infty}^{\infty} f(x)dx = \lim_{b\to-\infty} \int_{b}^{0} f(x)dx + \lim_{a\to\infty} \int_{0}^{a} f(x)dx$. The p.d.f. defines the (cumulative) distribution function F, where $F(x) = \int_{-\infty}^{x} f(\xi)d\zeta = P(X \leq x)$. The area under the curve is 1. The area under the curve up to the x line is f(x). The p.d.f., f(x), is sometimes called **likelihood**.

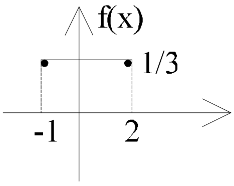




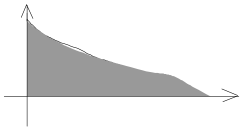
Theorem: $P(a<X\leq b) = F(b)-F(a) = \int_{-\infty}^{b} f(\zeta)d\zeta - \int_{-\infty}^{a} f(\zeta)d\zeta = \int_{a}^{b} f(\xi)d\zeta$. If f is *continuous* at b, then $\lim_{a\to b}$ (P(a<x≤b)) (*This* is P(X=b) = 0) = $\lim_{a\to b}$(F(b)-F(a)) = 0. And P(a<X≤b) = P(a≤X≤b) = P(a≤X<b). Important *examples* of continuous random variables:

Uniform Distribution. Let a and b be real, with a < b. f(x) = $^{1}/_{b-a}$ if x ∈ [a,b], and 0 *otherwise*. (1) $f(x) \geq 0$ ∀ real x. (2) $\int_{-\infty}^{\infty} f(x)dx = 1$. Here, $\int_{a}^{b} \frac{1}{b-a}dx = \frac{1}{b-a}\int_{a}^{b} dx = \frac{1}{b-a}[x]_{a}^{b} = \frac{b-a}{b-a} = 1$ Example: a = -1, b = 2. Find P(0<X≤1). A: P(0<X≤1) = $\int_{0}^{1} \frac{1}{3}dx = [\frac{x}{3}]_{0}^{1} = \frac{1}{3}$.

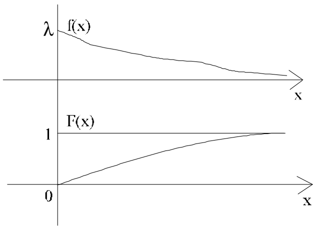


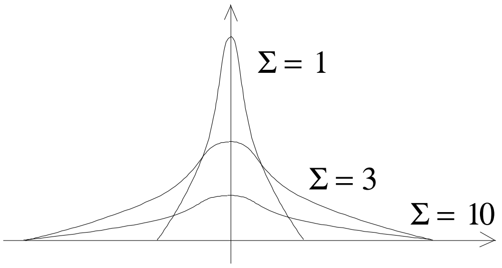


Exponential. $f(x) = \lambda e^{-\lambda x}$ if $x \geq 0$, and 0 otherwise. λ is a +ve parameter (1) $f(x) \geq 0$ for all *real* x. (2) $\int_{0}^{\infty} \lambda e^{-\lambda x} = \int_{0}^{\infty} -e^{-\lambda x}d(-\lambda x) = [-e^{-\lambda x}]_{0}^{\infty}$ = -(0-e^{0}) = 1. Example: the **life** of an electric bulb in hours x has the *following* pdf: f(x) = $^{1}/_{1000}e^{-(x/1000)}$ when x > 0, and 0 otherwise. Find the **distribution** function F(x) and use it to find the probability that the bulb *lasts* (i) > 1000 hours, (ii) **between** 1000 and 2000 hours.



A: F(x) = P(X≤x) = $\int_{-\infty}^{x} f(\zeta)d\zeta = \int_{0}^{x} \frac{1}{1000}e^{-(\zeta/1000)}d\zeta = -\int_{0}^{x} e^{-(\zeta/1000)}d(-\frac{\zeta}{1000})$ = $[-e^{-(\xi/1000)}]^{x}_{0}$ = $-e^{-(x/1000)}+1$ = $1-e^{-(x/1000)}$. Now P(X > 1000) = 1-P(X ≤ 1000) = 1-F(1000) = $1-(1-e^{-1000/1000})$ = $1-1+e^{-1}$ = e^{-1}. And P(1000≤X≤2000) = F(2000)-F(1000) = $1-e^{-2000/1000}$ - $(1-e^{-1000/1000})$ = $1-e^{-2}+e^{-1}-1$ = $e^{-1}-e^{-2}$.





Normal Distribution: $f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. μ and σ are parameters, with σ > 0.

N-Dimensional Random Variables

We form an n-dimensional random variable on a *sample* space S by assigning n real values to each point in S. [X = $(X_1,X_2,...,X_n)$; X: S→$\mathbf{R}^n$]. Example: Toss a coin *twice*. Let X = the number of heads in the 1st toss; Y = number in 2nd toss. (X,Y) is a **2-dimensional** random variable.

S	HH	HT	TH	TT
(X,Y)	(1,1)	(1,0)	(0,1)	(0,0)

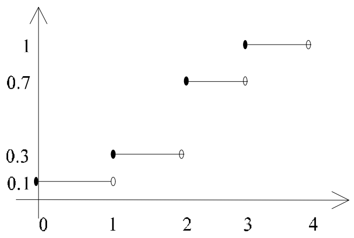
(1) For a 2-D **random** variable, (X,Y), the cumulative *distribution* function F(x,y) is defined as F(x,y) = P(X≤x, Y≤y) for all (x,y) in $\mathbf{R}^2$. F: $\mathbf{R}^n \to \mathbf{R}$, F: $\mathbf{R}^n \to [0,1]$. (2) **The** marginal distribution f^{ns} of X & Y are *defined* by marginal distribution f^{ns} of C. $F_1(x) = \lim_{y\to\infty} F(x,y)$ — get rid of the y. (F_1: the “1” means the **1st** argument). Marginal *distribution* f^n of Y: $F_2(y) = \lim_{x\to\infty} F(x,y)$ — get **rid** of the x. (3) In (X,Y), X and Y are *independent* iff $F(x,y) = F_1(x)F_2(y)$ for all (x,y) in $\mathbf{R}^2$. P(X≤x, Y≤y) = P(X≤x)P(Y≤y) = (No matter **what** y is)×(No matter **what** x is).

11th March 1999

Tutorial

Q: The **distribution** function of a discrete random variable X is *tabulated* on the right. (a) How many different values can X take? List them. (b) Find the p.m.f. of X, f(x). (c) What is the **probability** that X ≥ 3? What is the probability that X = 3? A: X can only take the values 0, 2, 5, 6 (This is when the “*jumps*” occur). Now P(X≥3) = $\Sigma_{x_i\geq3}\, p(x_i)$ = 0.1+0.5 = 0.6. And P(X = 3) = 0. (*No jump*).

x	(-∞,0)	[0,2)	[2,5)	[5,6)	[6,∞)
F(x)	0	0.1	0.4	0.5	1



Q: Let X be a *discrete random variable* with p.m.f. as shown on the right. (a) Find F(x) for the largest possible value of a. Draw it. (b) Find F(x) for the **largest** possible value of b. A: Now 0.1+a+0.4+0.3+b = 1; 0.8+a+b = 1; a+b = 0.2. **The largest** value of a is 0.2, when b = 0. **The largest** value of b is 0.2, when a = 0.

x	0	1	2	3	4
p(x)	0.1	a	0.4	0.3	b

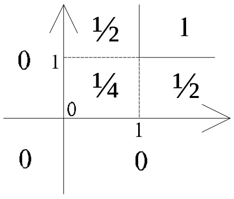
Q: A continuous random variable has p.d.f. given by $f(x) = Cxe^{-x^2}$ for x ≥ 0 and 0 otherwise. Find the value of the *constant* C. A: Integrating, $\int_0^\infty Cxe^{-x^2}dx = 1;\ C\int_0^\infty xe^{-x^2}dx = 1;$ $C[\frac{e^{-x^2}}{-2}]_0^\infty = 1;\ C(-\frac{1}{2})(0-1) = 1;\ \frac{1}{2}C = 1; C = 2.$

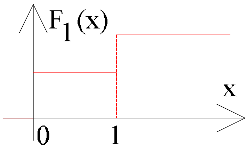
Discrete 2 Dimensional Variables

(X,Y) takes only a **finite**/*countable* number of values (x_i, y_j). Let $p(x_i,y_j) = P(X=x_i, Y=y_j)$. (1) $p(x_i, y_j) \geq 0$. (2) $\Sigma_i\Sigma_j\, p(x_i,y_j) = 1$. These two **say** that $p(x_i, y_j)$ is the pmf of the 2-D *random variable* (X,Y). $F(x,y) = \sum_{x_i\leq x}\sum_{y_j\leq y} p(x_i,y_j)$. **Marginal** Distribution Function: $F_1(x) = P(X\leq x, Y<\infty) = \Sigma_{x_i\leq x}\, \Sigma_{y_j}\, P(x_i,y_j)$. $F_2(y) = P(X<\infty, Y\leq y) = \Sigma_{x_i}\, \Sigma_{y_j\leq y}\, P(x_i,y_j)$. *Marginal* Probability mass function: $p_1(x_i) = \Sigma_j\, p(x_i, y_j)$; $p_2(y_j) = \Sigma_i\, p(x_j, y_i)$.

Example: Toss a coin twice. This is **shown** in the **table**. Note: In (X,Y), X is the number of heads in the 1st throw; Y = number in 2nd throw. i and j can take values 0 & 1. Now we want to *find* F(x,y) in [0,1]. Look at the diagram on the left. Notice that e.g. $F(-2.3, -0.8) = \Sigma_{x_i\leq-2.3}\, \Sigma_{y_j\leq-0.8}\, P(x_i,y_j) = 0$. **Generally**, $F(x_i < 0, y_j) = 0$. Another **example**: F(1.3, 0.5) = P(0,0)+P(1,0) = ½. Marginal function: $F_1(x) = \Sigma_{x_i\leq x}\, \Sigma_j\, p(x_i,y_j)$. Diagram shown on right.

S	HH	HT	TH	TT
(X,Y)	(1,1)	(1,0)	(0,1)	(0,0)
$p(x_i,y_j)$	¼	¼	¼	¼





Assignment 1

Q: In how many **ways** can a set of n elements be *split into 2* non empty subsets? A: Let $U = \{u_1,...,u_n\}$. The number of **all** possible subsets is 2^n. Except U and ϕ, all *others* have a non-empty complement. Each subset $A \subset U$ defines an $\bar{A} \subset U$, and this **partition** is the same as taking $\bar{A} \subset U$ and $\bar{\bar{A}} = A \subset U$. Therefore, the number of **different** splits is $(2^n-2)/2 = 2^{n-1}-1$.

Q: There are N_1 students in Year 1, N_2 in Year 2, and N_3 in Year 3. 2 **students** won the lottery, which everybody tried. Knowing that B has studied for *a longer time* than A, what is the probability that B is in year 3? A: S = {(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)}. $|S| = 3^2 = 9$. S *is not* equiprobable. Let X = “B studies longer than A”. X = {(1,2), (1,3), (2,3)}. Let Y = “B is in year 3”. Y = {(1,3), (2,3), (3,3)}.

We want to **find** $P(Y|X) = P(X\cap Y)/P(X)$. Now $P(X) = P(1,2)+P(1,3)+P(2,3)$. And $P(X\cap Y) = P(1,3)+P(2,3)$. Let $N = N_1+N_2+N_3$. Example: P(1,2) = P(A is in 1)×P(B is in 2) = $(N_1/N)\times(N_2/N)$. So $P(Y|X) = \frac{(N_1N_3/N^2)+(N_2N_3/N^2)}{(N_1N_2/N^2)+(N_2N_3/N^2)+(N_1N_3/N^2)} = \frac{N_1N_3+N_2N_3}{N_1N_2+N_2N_3+N_1N_3}$.

Q: Prove by induction the **multiplication** theorem: for a set of events $\{A_1,...,A_n\}$ and $P(A_1\cap...\cap A_{n-1}) > 0$, $P(\cap^n_{i=1} A_i) = P(A_1)P(A_2|A_1)...P(A_n|A_1\cap...\cap A_{n-1})$. A: For n = 2, $P(A_1\cap A_2) = P(A_1).P(A_2|A_1)$. This is correct. Assume that the formula *holds* for n-1 i.e. $P(A_1\cap A_2\cap...\cap A_{n-1}) = P(A_1).P(A_2|A_1)...P(A_{n-1}|A_1\cap A_2\cap...\cap A_{n-2})$ (*).

Now $P((A_1\cap A_2\cap...\cap A_{n-1})\cap A_n) = P(\cap^n_{1=i} A_i)$. So $P((A_1\cap...\cap A_{n-1})\cap A_n)$ = (by (*)) = $P(A_1\cap...\cap A_{n-1}).P(A_n|A_1\cap...\cap A_{n-1}) = P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)...P(A_n|A_1\cap...\cap A_{n-1})$. Therefore, the formula *holds* for n, which **completes** the proof.

Q: *5 Russians and 3 Germans* compete in 2 semi-final heats, with 4 swimmers in each heat. (a) What’s the number of possible corridor arrangements for the 1st semi final? (b) Do likewise for the 2nd semi final. (c) What is the number of possible **8-corridor arrangements** if all 8 swimmers are to start in one series? (d) Is the *sum* of the first two ((a) and (b)) equal to the number you obtain in (c). Explain **why** you arrive at your answer.

A: Consider the **1st** semi-final. For RRRR, we have 1 *combination*. For RRRG, we have $^{4!}/_{3!}$ = 4 combinations. For RRGG, 6 combos. For RGGG, 4 **combos**. So for the 1st semi-final, there are 1+3+6+4 = 15 different *arrangements*. Now consider the 2nd semi-final. RGGG gives $^{4!}/_{3!}$ = 4 combos. RRGG gives $^{4!}/_{2!2!}$ = 6 combos. RRRG gives 4 combos *while* RRRR gives 1 combo. So there are 4+6+4+1 = 15 different arrangements for the 2nd semi-final (The same number because the **possible** combination of 4 and 4 are symmetric. Now for (c), we have $^{8!}/_{5!3!}$ = 56 possible *arrangements*.

(d) We have (a)+(b) = 30, and (c) = 56. Therefore the sum is *not equal* to (c). The difference comes from the fact that for EACH 1st semi final arrangement, all *possible* permutations of the 2nd semi final arrangements will count as **different** corridor arrangements. So in (a)+(b), the 4 *arrangements* in the first semi final (RRRG) and the 6 in the 2nd semi final (RRGG) count as 24 different arrangements.

17th March 1999

Continuing the example where we *tossed* the two coins (X = number of heads in the 1st throw; Y = number in 2nd throw). $F(x,y) = 0$ for $-\infty<x<0$ OR $-\infty<y<0$. It is a ¼ for $0\leq x<1$ AND $0\leq y<1$. It is a ½ for $1\leq x<\infty$ AND $0\leq y<1$. It is a ½ for $0\leq x<1$ AND $1\leq y<\infty$. **And** it is 1 for $1\leq x<\infty$ AND $1\leq y<\infty$. Try to visualise it from the *pictures* on page 12.

$F_1(x) = 0$ for $-\infty<x<0$. It is a ½ for $0\leq x<1$. It is 1 for $1\leq x<\infty$. Similarly for $F_2(y)$, it is 0 for $-\infty<y<0$; ½ for $0\leq y<1$, and 1 for $1\leq y<\infty$. Now $p_1(x_i) = \Sigma_j\, p(x_i,y_j)$ — the argument is *discrete*. $p_1(0) = p(0,1)+p(0,0) = ½$. $p_1(1) = p(1,0)+p(1,1) = ½$. $p_2(y_i)$ is similar.

If (X,Y) is a 2-D r.v., then X & Y are *independent* iff $F(x,y) = F_1(x).F_2(y)$. Let event $A = (X\leq x)$. Let event $B = (Y\leq y)$. Now $F(x,y) = P(X\leq x, Y\leq y)$. (Note: The *comma* means "and"). So $F(x,y) = P(A\cap B)$. Now $F_1(x) = P(X\leq x) = P(A)$ and $F_2(y) = P(Y\leq y) = P(B)$. Now we *know that* $P(A\cap B) = P(A)P(B)$.

Theorem: If (X,Y) is a 2-D discrete r.v., and $p(x_i, y_j) = p_1(x_i).p_2(y_j)$ for all x_i, y_j, then X & Y are *independent*. We want to prove that $F(x,y) = F_1(x).F_2(y)$. Now $F(x,y) = \Sigma_{x_i\leq x}\Sigma_{y_j\leq y}\, p(x_i,y_j) = \Sigma_{x_i\leq x}\Sigma_{y_j\leq y}\, p_1(x_i).p_2(y_j) = \Sigma_{x_i\leq x}\, p_1(x_i)\, \Sigma_{y_j\leq y}\, p_2(y_j) = F_1(x).F_2(y)$.

18th March 1999

Q: Prove that F(x), the *cumulative distribution function*, is monotonically non-decreasing. Let a and b be such that a and b are Real, and b > a. X can be **discrete** or **continuous**. In F(x), the argument x is continuous. But F(x) can contain *discontinuities* ($F(x)$: $\mathbf{R}\rightarrow[0,1]$). By definition, $F(a) = P(X\leq a)$, and $F(b) = P(X\leq b)$. From the *theorem* about F(x), $P(a<x\leq b) = F(b)-F(a)$. This is a probability. Therefore, it is non-negative — a property of probability. So $F(b) - F(a) \geq 0$, and therefore $F(b) \geq F(a)$. This *completes* the proof.

Alternative method: $F(a) = P(X\leq a)$ and $F(b) = P(X\leq b)$ by definition. *Define* $A = (X\leq a)$ and $B = (a<X\leq b)$. It follows that $A\cap B = \phi$ and $A\cup B = X\leq b$. Now $P(A\cup B) = P(X\leq b) = F(b)$ by *definition*. $P(A\cup B) = P(A)+P(B)-\phi$. **(Mutually exclusive**). $F(b) = F(a)+P(B)$. As $P(B) \geq 0$, *then* $F(b) \geq F(a)$.

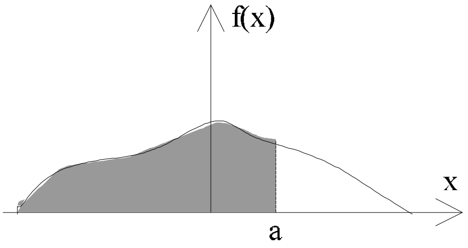
Q: Toss a *biased* coin twice, where P(H) = 0.4 and P(T) = 0.6. Let X = the number of heads obtained; Y = the number of tails obtained. Find $p(x_i,y_j)$, $F_1(x)$, $F_2(y)$, $p_1(x_i)$, and $p_2(y_j)$. A: $x_i, y_j \in \{0,1,2\}$. (X,Y) cannot take *all* combinations. $p(x_i,y_j)$ can be seen in the table. From this, we can **calculate**: $\mathbf{F_1(x)} = 0$ for $x < 0$, 0.36 for $0\leq x<1$, 0.84 for $1\leq x<2$, 1 for $x \geq 2$. $\mathbf{F_2(y)}$ is 0 for $y < 0$, 0.16 **for** $0\leq y<1$, 0.64 for $1\leq y<2$ *and* 1 for $y \geq 2$. Now $p_1(x_i)$ is $p_1(0) = 0.36$, $p_1(1) = 0.48$, and $p_1(2) = 0.16$. **And** $p_2(y_j)$ is $p_2(0) = 0.16$, $p_2(1) = 0.48$, and $p_2(2) = 0.36$.

(X,Y)	(0,2)	(1,1)	(2,0)
	TT	HT, TH	HH
$p(x_i,y_j)$	0.36	0.48	0.16

23rd March 1999

Continuous Valued 2-D (n-D) Random Variables

$(X,Y) \in \mathbf{R}^2$. *Definition*: The (joint) probability density function of (X,Y) is f(x,y), with the following **properties**: (1) $f(x,y) \geq 0$ for all (x,y) in $\mathbf{R}^2$. (2) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$. The p.d.f. determines the *distribution function* F(x,y), where $F(x,y) = (X\leq x, Y\leq y)$.

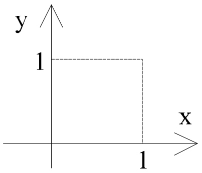


Reminder: $P(X \le a)$, the area, is $\int_{-\infty}^{a} f(x)dx = F(a)$. And $F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdx$. Marginal p.d.f.: $f_1(x) = \int_{-\infty}^{\infty} f(x,y)dy$ (We get **rid** of y). And $f_2(y) = \int_{-\infty}^{\infty} f(x,y)dx$ (We get **rid** of x).

<u>Marginal Distribution functions</u>. $F_1(x) = P(X \le x, \text{any } Y) = \int_{-\infty}^{x} f_1(\xi)d\xi$. $F_2(y) = \int_{-\infty}^{y} f_2(\xi)d\xi$. *Theorem*: Let (X,Y) be a 2-D r.v. If $f(x,y) = f_1(x)f_2(y)$ for all (x,y) in $\mathbf{R}^2$, then X and Y are called *independent*. So to be independent, we must show that $p(x_i,y_j) = p_1(x_i)p_2(y_j)$ *holds* for all x_i, y_j.

Conditional Probability Distributions

(a) Let (X,Y) be a **discrete** 2-D r.v., where the p.d.f. and the p.m.f. are known. The conditional p.m.f. of X given Y is given by $P(x_i|y_j) = P(X=x_i \mid Y=y_j) = P(x_i,y_j)/p_2(y_j)$. Here, $P(x_i,y_j) = P(X=x_i \text{ and } Y=y_j)$. Let y_j be fixed. Then $P(x_i|y_j)$ is a **legitimate** p.m.f. of X if (1) $P(x_i,y_j) \geq 0$; so that $P(x_i,y_j)/p_2(y_j) \geq 0$; ($p_2(y_j) > 0$ so this is all right); (2) $\Sigma_i\, p(x_i|y_j) = 1$. Now $\Sigma_i\, p(x_i|y_j) = \Sigma_i\, P(x_i,y_j)/p_2(y_j) = [1/p_2(y_j)]\Sigma_i\, P(x_i,y_j) = {}^{1}/_{p_2(y_j)}.p_2(y_j) = 1$. **And** $P(y_j|x_i) = P(x_i,y_j)/p_1(x_i)$. *Similarly*, $p(y_j|x_i)$ is a <u>legitimate</u> p.m.f. of Y for **fixed** x_i. (b) *Continuous* valued (X,Y). $f(x|y) = F(x,y)/f_2(y)$ ($f_2(y) > 0$); $f(y|x) = F(x,y)/f_1(x)$ ($f_1(x) > 0$).



<u>Example</u>: Let (X,Y) be a *continuous* valued r.v.; where $F(x,y) = Cx(1-y)$ for $x \in [0,1]$, $y \in [0,1]$; and 0 elsewhere. (i) <u>Find</u> C. Now $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$ so we have $\int_0^1\int_0^1 Cx(1-y)dxdy = 1.C\int_0^1 xdx\int_0^1 (1-y)dy = 1$. So $C[{}^{x^2}/_2]_0^1[y-{}^{y^2}/_2]_0^1 = 1$; $C[{}^1/_2][1-{}^1/_2] = 1$; ${}^C/_4 = 1$, C = 4.

(ii) Find the **marginal** p.d.f.'s $f_1(x)$, $f_2(y)$. A: $f_1(x) = \int_{-\infty}^{\infty} F(x,y)dy$; $f_1(x) = \int_0^1 4x(1-y)dy$ for all $x \in [0,1]$, and 0 otherwise. So $f_1(x) = \int_0^1 4x(1-y)dy = 4x\int_0^1 (1-y)dy = 4x[y-{}^{y^2}/_2]_0^1 = 4x[1-{}^1/_2] = 2x$. So $f_1(x) = 2x$ for $x \in [0,1]$ and 0 *elsewhere*. Similarly, $f_2(y) = \int_{-\infty}^{\infty} F(x,y)dx = \int_0^1 4x(1-y)dx$ for all $y \in [0,1]$, and 0 otherwise. So $f_2(y) = \int_0^1 4x(1-y)dx = 4(1-y)\int_0^1 x = 4(1-y)[{}^{x^2}/_2]_0^1 = 2(1-y)$.

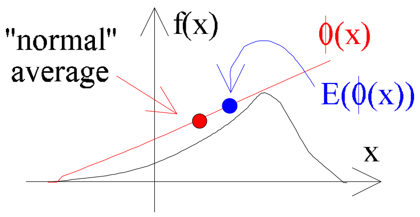
(iii) Are X & Y **independent**? A: $F(x,y) = 4x(1-y) = 2x.2(1-y) = f_1(x).f_2(y)$ for all x, y. This holds for <u>all</u> points. For example, the point $(2, {}^1/_2)$ gives $F(x,y) = 0$; $f_1(x) = 0$; $f_2(y) = 2(1-y)$. So $F(x,y) = f_1(x)f_2(y)$ holds here. Conclusion: X and Y *are* independent.

<u>24th March 1999</u>

Expectations & Moments

<u>Definition</u>: Let (X,Y) be a 2-D r.v., and let $\phi(X,Y)$ be any *function* of X and Y. The ***expected value*** $E(\phi(X,Y))$ is given by (a) <u>discrete</u> (X,Y): $E(\phi(X,Y)) = \Sigma_i\Sigma_j\, \phi(x_i,y_j).p(x_i,y_j)$; (b) *Continuous* valued (X,Y): $E(\phi(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi(x,y).f(x,y)dxdy$. We can extend this definition for an n-D r.v. $X(x_1,\ldots,x_n)$, and a *function* $f(x_1,\ldots,x_n)$.

1-D case: **discrete** x, $\phi(x)$: $E(\phi(X)) = \Sigma_i\phi(x_i)p(x_i)$. Continuous valued X, $\phi(X)$: $E(\phi(X)) = \int_{-\infty}^{\infty} \phi f(x)dx$.



Properties of Expectations

Let $\phi_1(X_1,\ldots,X_n)$, ..., $\phi_m(X_1,\ldots,X_n)$ be m *functions* of the n-D r.v. $(X_1,\ldots,X_n)$, and let a_1, ..., a_m be real coefficients. Then $E(\Sigma^m_{i=1}\ a_i\phi(X_1,\ldots,X_n)) = \Sigma^m_{i=1}\ a_iE(\phi_i(X_1,\ldots.,X_n))$. Proof: (For 2-D discrete (X,Y)): $E(\Sigma^m_{i=1}\ a_i\phi_1(X,Y)) = \Sigma_j\Sigma_k(\Sigma^m_{i=1}\ a_i\phi_i(x_j,y_k)).p(x_j,y_k) = \Sigma^m_{i=1}\ a_i(\Sigma_j\Sigma_k\ \phi_i(x_j,y_k).p(x_j,y_k))$. Now we **know that** $\Sigma_j\Sigma_k\ \phi_i(x_j,y_k).p(x_j,\ y_k)) = E(\phi_i(X,Y))$, so we *have* the result $E(\Sigma^m_{i=1}\ a_i\phi_1(X,Y)) = \Sigma^m_{i=1}\ a_iE(\phi_1(X,Y))$.

A **special** case: $\phi_i(X_1,\ldots,X_n) = X_i$. So here, $E(\Sigma^m_{i=1}\ a_iX_i) = \Sigma^m_{i=1}\ a_iE(X_i)$. *REMEMBER THIS RESULT*. Another case: E(constant). Let constant = a, so that $E(a) = \int_{-\infty}^{\infty} af(x)dx = a\int_{-\infty}^{\infty} f(x)dx$. **Now** $\int_{-\infty}^{\infty} f(x)dx = 1$, so have $E(a) = a$.

25th March 1999

Tutorial

Q: Two die are thrown. Let X be calculated as the **number** on the 2nd die - the number on the 1st die. (a) Find the p.m.f. of X; (b) What is the probability that the *difference* between the 2 faces is 2 or less; (c) Calculate E(X). A: |S| = 36; X takes values from -5 to 5. We have p.m.f. as shown in the table. (b) Let A = "The *difference is* $\leq$ 2". $P(A) = {}^2/_3$ from the table. (c) We have $E(X) = \Sigma^5_{i=-5}\ i.p(i) = -{}^5/_{36}-{}^8/_{36}-{}^9/_{36}-{}^8/_{36}-{}^5/_{36}+0+{}^5/_{36}+{}^8/_{36}+{}^9/_{36}+{}^8/_{36}+{}^5/_{36} = 0$.

X	-5	-4	-3	-2	-1	0	1	2	3	4	5
P(X=x)	$^1/_{36}$	$^2/_{36}$	$^3/_{36}$	$^4/_{36}$	$^5/_{36}$	$^6/_{36}$	$^5/_{36}$	$^4/_{36}$	$^3/_{36}$	$^2/_{36}$	$^1/_{36}$

Q: Let X be a **random** variable with p.d.f. $f(x) = k(x+4)$ for $x \in [1,2]$ and $f(x) = 0$ elsewhere. (a) Find k. (b) Calculate E(X). (c) For which value x of X is $P(X\geq x) = 0.7$? A: $\int_{-\infty}^{\infty} f(x)dx = 1$. So $\int^2_1 k(x+4)dx = 1$; $k[{}^{x^2}/_2+4x]^2_1 = 1$; ... $k = {}^2/_{11}$. (b) $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int^2_1 x.{}^2/_{11}(x+4)dx = {}^2/_{11}[{}^{x^3}/_3+2x^2]^2_1 = .. = {}^{50}/_{33}$. (c) We want $\int_x^{\infty} f(x)dx = 0.7$; ${}^2/_{11}\int^2_x (u+4)du = 0.7$; ${}^2/_{11}[{}^{u^2}/_2+4u]^2_x = 0.7$; ...$20u^2+80u+21 = 0$, from which *we get 2 values* — 1 of which is in the interval.

Q: (X,Y), a discrete 2-D r.v. has p.m.f. as shown in the **table**. (a) Find the *marginal* p.m.f.'s; (b) Are X & Y dependent?; (c) Find $E(X^2+Y^2)$. A: (a) For $p_1(x_i)$, $p_1(1) = 0.3+0.1+0.2 = 0.6$; $p_1(2) = 0.4$. And for $p_2(y_j)$, $p_2(1) = 0.3+0.1 = 0.4$; $p_2(2) = 0.3$; $p_2(3) = 0.3$. (b) As P(1,1) is **not** equal to $p_1(1).p_2(1)$, then X and Y are not *independent*. (c) $E(X^2+Y^2) = 2\times0.3 + 5\times0.1 + 10\times0.2 + 5\times0.1 + 8\times0.2 + 13\times0.1 = 6.5$.

(X,Y)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)
$p(x_i,y_j)$	0.3	0.1	0.2	0.1	0.2	0.1

Easter Holiday.

20th April 1999

Assignment 2

Q: **Three** people fire at a target. If A, B & C fire *together*, P(Target not hit) = ¼. If A & B fire *together*, P(Target hit) = $^5/_8$. If A & C fire together, then P(Target hit) = ½. Find the **individual** probabilities of hitting the target. A: P(Target has not been hit) = (1-P(A))(1-P(B))(1-P(C)) = ¼.

Also we know that (1-P(A))(1-P(B)) = 1-$^5/_8$ = $^3/_8$. And (1-P(A))(1-P(C)) = ½. Now $^3/_8$.(1-P(C)) = ¼; 1-P(C) = $^2/_3$; P(C) = $^1/_3$. And $^2/_3$.(1-P(A)) = ½; 1-P(A) = ¾; P(A) = ¼. And ¾.(1-P(B)) = $^3/_8$; P(B) = 1-$^1/_2$ = ½. So P(A) = ¼, P(B) = ½, P(C) = $^1/_3$.

Q: Find the probability that at **least** 3 of the first 5 persons *encountered* on a given day were born on a Saturday. A: *Binomial* r.v.: p = $^1/_7$; n = 5. Now P(X = 3) = $(^5{}_3)p^3q^2 = 360/7^5$. P(X = 4) = $(^5{}_4)p^4q^1 = 30/7^5$. And P(X = 5) = $(^5{}_5)p^5q^0 = 1/7^5$. So P(X = 3 or 4 or 5) = P(X = 3) + P(X = 4) + P(X = 5) = 360 + 30 + $(1/7^5) = 391/7^5 \approx 0.02326$.

Q: A bag contains **5 white** and **4 black** balls. A and B take turns in drawing one ball at a time *without* replacement. A draws first. What is the probability that A will be the **first** to draw a black ball? A: A draws a black ball = {B} ∪ {WWB} ∪ {WWWWB}. Now P({B}) = $^4/_9 = {}^{28}/_{63}$. P({WWB}) = $^5/_9\times{}^4/_8\times{}^4/_7 = {}^{10}/_{63}$. And P({WWWWB}) = $^5/_9\times{}^4/_8\times{}^3/_7\times{}^2/_6\times{}^4/_8 = {}^2/_{63}$. So we have P({B})+P({WWB})+P({WWWWB}) = $^{28+10+2}/_{63} = {}^{40}/_{63}$.

Expectation of a 1-D Random Variable

(1) The **r^{th}** moment of X about the origin is $\mu_r' = E(X^r)$. (2) $\mu_1' = \mu = \mu_x = E(X)$. The **mean** of X, the measure of *central* tendency, is $\mu_x = \Sigma_i x_ip(x_i)$ for **Discrete** valued X and $\int_{-\infty}^{\infty} xf(x)dx$ for **Continuous** valued x. (3) The r^{th} moment of X about the *mean is* $\mu_r = E((X-\mu)^r)$. (4) $\mu_2 = \sigma^2 = \sigma_x^2 = E(X-\mu)^2$, the *variance* of X. And $+\sqrt{\sigma^2}$ is the *standard deviation* of X. These two are measures of **variability**.

Theorem: $E((X-\mu)^2) = \sigma_x^2 = E(X^2)-(E(X))^2 = \mu_2'-(\mu)^2$. Proof: $E((X-\mu)^2) = E(X^2-2\mu X+\mu^2) = E(X^2) - E(2\mu X) + E(\mu^2) = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - 2\mu\mu + \mu^2 = E(X^2) - (E(X))^2$. This result gives an *easy* formula to **calculate** σ_x^2.

Theorem: Let $Y = X-\mu_x/\sigma_x$. Then E(Y) = 0, $E((Y-0)^2) = 1$. Proof: $E(Y) = E(X-\mu_x/\sigma_x) = (1/\sigma_x)\times E(X-\mu_x) = (1/\sigma_x)\times(E(X)-E(\mu_x)) = (1/\sigma_x)\times(\mu_x-\mu_x) = 0$. And $E((Y-0)^2) = E((X-\mu_x/\sigma_x)^2) = (1/\sigma_x^2)\times(E(X-\mu_x)^2)$. Now as $(E(X-\mu_x)^2)$ is σ_x^2, we have $E((Y-0)^2) = 1$. Any *random* variable can be standardised to a random variable with mean 0 and variance 1.

Other Measures of Central Tendency

(1) The **median** of X, m, is the value *satisfying* $P(X \geq m) \geq$ ½ and $p(X \leq m) \geq$ ½.

(1a) Continuous valued X. We know that $P(X \leq a)$ or $P(X<a)$ is the *area under the graph* before the point $x = a$. In this case, we **want** $P(X \geq m) = \frac{1}{2}$ and $P(X \leq m) = \frac{1}{2}$. (Because $P(X = m) = 0$).

(1b) Discrete X. **Two** examples are shown. For the first one, $m = 1$. This is because $P(X \leq 1) = 0.3+0.6 = 0.9 \geq \frac{1}{2}$; and $P(X \geq 1) = 0.6+0.1 = 0.7 \geq \frac{1}{2}$. In the **second** example, m can be anything in [1,2] — work out some examples for *yourself*.

x_i	0	1	2
$p(x_i)$	0.3	0.6	0.1

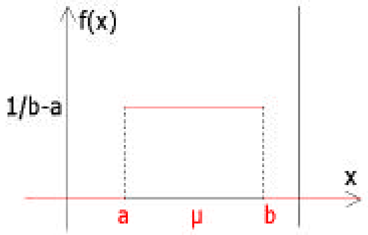
x_i	0	1	2	3
$p(x_i)$	0.3	0.2	0.1	0.4

(2) **Mode**. A mode of X is a value of X for which $p(x_i)$ or $f(x)$ has a *maximum*. Distributions with one mode are called **unimodal**. Distributions with 2 modes are **bimodal**. Any more, and the distribution is called **multimodal**.

21st April 1999

Measures of central tendency: mean, median, mode. *Measures of variability:* variance and s.d. *Measure of unsymmetry*: the coefficient of skewness. *Measure of flatness:* the coefficient of kurtosis. The coefficient of skewness, B_1, is $\mu_3/\sigma^3 = E((X-\mu_x)^3)/\sigma^3$. When the graph is **more** skewed, B_1 is higher. The coefficient of *Kurtosis*, B_2, is $\mu_4/\sigma^4 = E((X-\mu_x)^4)/\sigma^4$. When the graph is **more** flat, B_2 is higher.

In examples, **remember** that μ_x is the *expectation* or *mean*, $E(X)$. μ_2' is $E(X^2)$, and μ_2 is the variance, σ^2, or $E((X-\mu)^2)$. *Example* Let $a < b$, with $a,b \in \mathbf{R}$. Define $f(x) = {}^1/_{b-a}$ for $x \in [a,b]$ and **otherwise**.



$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_a^b \frac{x}{b-a}dx = [\frac{x^2}{2(b-a)}]_a^b = {}^{b^2-a^2}/_{2(b-a)} = {}^{(b-a)(b+a)}/_{2(b-a)} = {}^{(b+a)}/_2$. Now $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_a^b \frac{x^2}{b-a}dx = {}^1/_{b-a}[{}^{x^3}/_3]^b_a = {}^{b^3-a^3}/_{3(b-a)} = {}^{(b-a)(b^2+ab+a^2)}/_{3(b-a)} = {}^{a^2+ab+b^2}/_3$. So $\mathrm{Var}(X) = E(X^2)-(E(X))^2 = {}^{a^2+ab+b^2}/_3 - {}^{(b+a)^2}/_4 = {}^{a^2+ab+b^2}/_3 - {}^{b^2+2ab+a^2}/_4 = {}^{4a^2+4ab+4b^2-3b^2-6ab-3a^2}/_{12} = {}^{b^2-2ab+a^2}/_{12} = {}^{(a-b)^2}/_{12}$.

Inequalities for Expectations

Theorem: Let Y be a 1-D r.v. which takes only *nonnegative* values, and let $E(Y) = \mu_Y$ exist. Then for any $a > 0$, $P(a \leq Y) \leq {}^{E(Y)}/_a = \mu_Y/a$. Proof: $\mu_Y = \int_{-\infty}^{\infty} yf(y)dy = \int_0^{\infty} yf(y)dy \geqslant \int_a^{\infty} yf(y)dy \geqslant \int_a^{\infty} af(y)dy = a\int_a^{\infty} f(y)dy = aP(a \leqslant Y)$. So $\mu_Y \geq aP(a \leq Y)$, i.e. $P(a \leq Y) \leq \mu_Y/a$.

22nd April 1999

Tutorial

Q: Let X be a d.r.v. **taking** values 3 or 5, each with probability ½. Find Var(X). Calculate the coefficients of *Skewness* and *Kurtosis* of X. A: $E(X) = (3 \times {}^1/_2)+(5 \times {}^1/_2) = 4$. $E(X^2) = (9 \times {}^1/_2)+(25 \times {}^1/_2) = 17$. So $\mathrm{Var}(X) = E(X^2)-(E(X)^2) = 17-4^2 = 1$. Now **Skewness**, $B_1 = \mu_3/\sigma^3 = E((X-\mu_x)^3)/\sigma^3$. **Kurtosis**, $B_2 = \mu_4/\sigma^4 = E((X-\mu_x)^4)/\sigma^4$. For X, $\sigma^3 = (\sqrt{4})^3 = 8$. Similarly, $\sigma^4 = 4^2 = 16$. Now $E((X-\mu_x)^3) = E((X-4)^3) = (3-4)^3.{}^1/_2+(5-4)^3.{}^1/_2 = 0$. And $E((X-\mu_x)^4) = E((X-4)^4) = (3-4)^4.{}^1/_2 + (5-4)^4.{}^1/_2 = 1$. So $B_1 = {}^0/_8 = 0$, and $B_2 = {}^1/_{16}$.

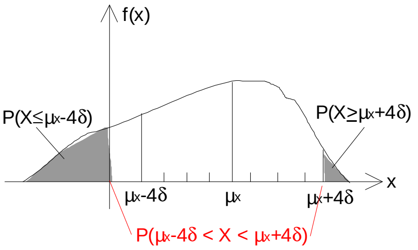
Q: Let Y be a c.r.v. with p.d.f. f(y) = ½ for $y \in [0,1]$ and $y \in [2,3]$; and 0 **elsewhere**. Find Var(Y). A: $E(Y) = \int_0^1 \frac{y}{2}dy + \int_2^3 \frac{y}{2}dy = [{}^{y^2}/_4]^1_0 + [{}^{y^2}/_4]^3_2 = \ldots = {}^3/_2$. $E(Y^2) = \int_0^1 \frac{y^2}{2}dy + \int_2^3 \frac{y^2}{2}dy = [{}^{y^3}/_6]^1_0 + [{}^{y^3}/_6]^3_2 = \ldots {}^{10}/_3$. Hence $Var(X) = E(X^2)-(E(X)^2) = {}^{10}/_3-({}^3/_2)^2 = {}^{13}/_{12}$.

Q: Let f(x) be $f(x) = {}^{3(1-x^2)}/_4$ for $x \in [-1,1]$ and 0 *elsewhere*. Prove that f is a valid p.d.f.. **Find** Var(X). Let g(x) = 3x-2. Find the <u>mean</u> and <u>variance</u> of g(x). A: Is it valid? Yes: All *probabilities* are +ve as $(1-x^2) \geq 0$ for $x \in [-1,1]$. And $\int_{-1}^{1} 3(1-x^2)/4\,dx = \ldots = 1$ as *expected*. Calculate Var(X) as before, to get $Var(X) = {}^1/_5$. Now g(x) = 3x-2, so $E(g(x)) = \int_{-1}^{1} ¾(3x-2)(1-x^2)dx = \ldots = -{}^7/_2$. Now $Var(g(x)) = E((g-\mu_g)^2) = E((3x-2-(-{}^7/_2))^2) = E((3x+{}^3/_2)^2) = E(9x^2+9x+{}^9/_4) = \int_{-1}^{1} {}^3/_4.(9x^2+9x+{}^9/_4)(1-x^2)dx = \ldots = {}^{81}/_{20}$.

<u>27th April 1999</u>

<u>Theorem: Chebyshev Inequality</u>

If X is a 1-D r.v., then for all k > 0, $P(k\sigma_x \leq |X-\mu_x|) \leq {}^1/_{k^2}$. Proof: Let $Y = |X-\mu_x|^2$, with $a = \sigma_x^2k^2$. Then using the *Theorem* $P(\sigma_x^2k^2 \leq |X-\mu_x|^2) \leq E(Y)/\sigma_x^2k^2$. Now $E(Y) = E(|X-\mu_x|^2) = \sigma_x^2$. So $P(\sigma_xk \leq |X-\mu_x|) \leq \sigma_x^2/\sigma_x^2k^2$. (Note: we have taken the *square* root inside each side of the probability). **QED. Application**: Looking at the *graph*, $P(X \leq u_x-4\delta) + P(X \geq \mu_x+4\delta) \leq {}^1/_{16}$.



Corollary 2: *Markov's* Inequality: $P(k \leq |X|) \leq E(|X|^r)/k^r$.

<u>Moments of some Standard Distributions (Discrete Variables)</u>

	Distribution	**Parameters**	μ_x	σ^2_x
1	*Binomial* Distribution. $P(X=r) = ({}^n{}_r)p^rq^{n-r}$. r can go *from 0 to n.*	n = the number of trials, p = P(success), q = 1-p	np	npq
2	*Bernoulli* trial(s)	p	p	pq
3	*Negative* Binomial. $P(X=r) = ({}^{n+r-1}{}_r)p^nq^r$. r can go from *0 to ∞*.	n = the number of successes, p = the probability of success	${}^{nq}/_p$	${}^{nq}/_{p^2}$
4	Poisson. $P(X=r) = e^{-\lambda} \lambda^r/r!$. r can go *from 0 to ∞.*	$\lambda > 0$	λ	λ

<u>Relationship Between Poisson and Binomial Distributions</u>

If $n \rightarrow \infty$ and $p \rightarrow 0$ in a *binomial distribution*, so that the mean np tends to λ, where λ is a <u>fixed positive constant</u>, then the Binomial r.v. tends to a **Poisson** random variable with *parameter* λ.

Example: Suppose that **300** misprints are distributed at random throughout a book with 500 pages. Find the probability that a *randomly* chosen page contains (a) exactly 2 misprints; (b) 2 or more misprints. A: Here, n, the number of <u>letters</u> on the page, is large; and p is small. This implies a *Poisson* distribution. λ is the expected number of errors per page = ${}^{300}/_{500}$ = 0.6. (a) $P(X=2) = e^{-0.6}.{}^{0.6^2}/_{2!} = 0.098$. (b) $P(X \geq 2) = 1-P(X=1)-P(X=0) = 1-e^{-0.6}0.6-e^{-0.6} = 0.122$.

28th April 1999

Continuous Valued Random Variables

	Distribution	**Parameters**	**μ_x**	**σ_x^2**
1	*Exponential*. $f(x) = \lambda e^{-\lambda x}$ when $x \geq 0$, and 0 when $x < 0$.	$\lambda>0$	$1/\lambda$	$1/\lambda^2$
2	*Normal* Distribution. $f(x) = \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(X-\mu_X)^2}{2\sigma_X^2}}$	μ_x, σ_x ($X \sim N(\mu_x, \sigma_x^2)$)	μ_x	σ_x^2
3	*Standardised* Normal. $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$	N(0,1)	0	1

What is the **link** between the Normal and the Standardised normal? $P(a \leq X \leq b) = \int_a^b f(x)dx = \int_a^b \frac{1}{\sqrt{2\pi}\,\sigma_x} e^{-\frac{(X-\mu_X)^2}{2\sigma_X^2}} dx$. Now $Z = X-\mu_x/\sigma_x$; $dz = (1/\sigma_x)dx$. So $dx = \sigma_x dz$. Changing *limits*, $x = a$ implies $z = a-\mu_x/\sigma_x$, and $x = b$ implies $z = b-\mu_x/\sigma_x$. So we have $\int_{a-\mu X/\sigma_X}^{b-\mu_X/\sigma_X} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = P(a-\mu_x/\sigma_x \leq Z \leq b-\mu_x/\sigma_x)$. Z has a *standardised* normal distribution, so $P(a-\mu_x/\sigma_x \leq Z \leq b-\mu_x/\sigma_x) = \Phi(b-\mu_x/\sigma_x) - \Phi(a-\mu_x/\sigma_x)$. ($\Phi(t)$ is the *cumulative* distribution function of Z).

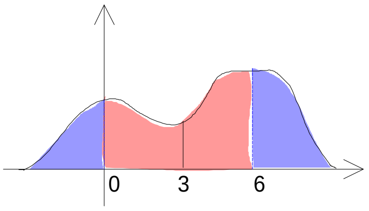
Properties of Φ: (1) $\Phi(t) \geq 0$ for all t, and $\Phi(t) \leq 1$ for all t. (2) F is *monotonically* non decreasing. (3) $\Phi(a) = 1-\Phi(-a)$, because of the symmetry of the **standardised** normal p.d.f. about the origin. $\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-a}^{\infty} e^{-\frac{z^2}{2}} dz = 1-P(Z \geq -a)$. So $\Phi(a) = 1-\Phi(-a)$.

Exercise: the temperature in June is *normally distributed* with mean 68ºF; standard deviation 6ºF. Find the probability that the temperature is (1) between 70º and 80º; (ii) Greater than 60º. A: $X \sim N(68, 36)$, so let $Z = {}^{X-68}/_{6}$. (i) $P(70 \leq X \leq 80) = P({}^{70-68}/_{6} \leq Z \leq {}^{80-68}/_{6}) = P(0.33 \leq Z \leq 2) = \Phi(2)-\Phi(0.333) = 0.9772-0.6293$ (from *tables)* = 0.348. (ii) $P(X \geq 60) = P(60 \leq X) = P({}^{60-68}/_{6} \leq Z) = P(-1.333 \leq Z) = 1-P(Z \leq -1.333) = 1-\Phi(-1.333) = \Phi(1.333) = 0.9082$ from *tables*.

29th April 1999

Tutorial

Q: Let X be a **random** variable of *unknown* distribution but of mean 3 and variance 4. Find a lower bound for the probability that X is between 0 and 6.



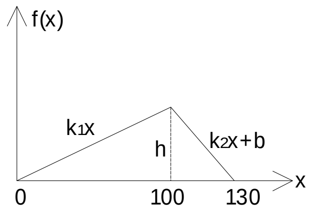
A: In $P(k\sigma_x \leq |X-\mu_x|) \leq {}^1/_{k^2}$, we have $P(2k \leq |X-3|) \leq {}^1/_{k^2}$. We want $k\sigma_x = 3$, i.e. $2k = 3$, $k = {}^3/_2$. So $P(3 \leq |X-3|) \leq {}^4/_9$. This is the area shown in blue. To get what we want, the red area, use $P(0 \leq X \leq 6) = 1-P(3 \leq |X-3|)$. So $1-P(3 \leq |X-3|) \geq 1-{}^4/_9$; $P(0 \leq X \leq 6) \geq {}^5/_9$. THINK about the **change** in inequality and what it *means*. So the lower bound is ${}^5/_9$.

Q: What is the **expected** number of sixes in 30 rolls of a fair die? What is the *standard* deviation? A: Apply the Binomial Theorem, with $n = 30$ and $p = {}^1/_6$, so that $q = {}^5/_6$. So $E(X) = np = 5$. And $Var(x) = npq = {}^{25}/_6$. So the standard deviation is $\sqrt{({}^{25}/_6)} = {}^5/_{\sqrt{6}}$.

Q: X is a **discrete** variable with $P(X = -1) = {}^1/_8$, $P(X = 0) = ¾$ and $P(X = 1) = {}^1/_8$. Calculate the mean, the standard deviation and the probability $P(|X-\mu| \geq 2\sigma)$. Compare with the **Chebyshev** estimate. A: (see over)....

$E(X) = (-1\times{}^1/_8)+(0\times{}^3/_4)+(1\times{}^1/_8) = 0$. $Var(X) = E(X^2)-(E(X)^2) = \ldots = {}^2/_8$. So Mean = 0, s.d. = ½. Now $P(|X-\mu| \geq 2\sigma)$ as requested is $P(|X-0| \geq 2\times{}^1/_2) = P(|X| \geq 1) = {}^1/_8+{}^1/_8 = ¼$. The **chebyshev** estimate is as follows: $P(k\sigma_x \leq |X-\mu_x|) \leq {}^1/_{k^2}$. We want $k\sigma_x = 1$, so ${}^1/_2 k = 1$, $k = 2$. Now $P(1 \leq |X|) \leq {}^1/_{2^2}$, so $P(1 \leq |X|) \leq ¼$.

Assignment 3

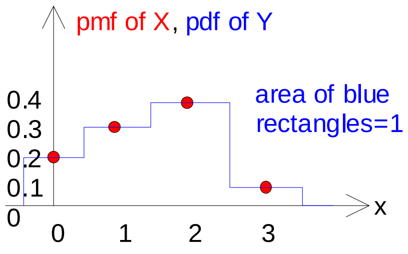


Q: The demand in **kg** of fruit you are selling is a random variable with the following p.d.f.: $f(x) = k_1x$ for $x \in [0,100]$; k_2x+b for $x \in [100,130]$, and 0 elsewhere. The price is fixed and is either £10 with probability 0.2, £15 with probability 0.5, and £19 with probability 0.2. You gain 20% of the total sale. Find the **constants** and calculate the *expected* profit.

A: From the **graph**, we see that $h = k_1.100$ or $h = k_2.100+b$. So $100k_1 = 100k_2+b$. We know that the area under the graph is 1, so the area of the **triangle** is 1: $1 = 130.h/2$. Using $h = 100k_1$, we get $1 = 130.100.k_1/2$; $k_1 = {}^1/_{6500}$. Substituting in the *other* formula we found, we get $k_2 = -{}^1/_{1950}$. And so $b = {}^{130}/_{1950}$ by taking the values at $x = 130$. The *expected* profit is $E(X)\times E(Y)\times 0.2$, where 0.2 = 20%. Calculate E(X) and E(Y) in the **normal** way (X uses integration, Y uses cross multiplication). Then the *expected* profit is about £219.

4th May 1999

Normal Approximation to the Binomial Distribution



For large n, (> 30), and **with** both np and nq $\geq$ 5, we can *approximate* the Binomial random variable X with parameters n and p by a normal r.v. Y with mean $\mu_Y = np$ and variance $\sigma^2_Y = npq$, i.e $Y \sim N(np, npq)$, provided a *continuity* condition is applied as follows: $P(a\leq X\leq b) = P(a-½\leq Y\leq b+{}^1/_2)$. For example, in the *graph*, $P(X=0) = P(-{}^1/_2\leq Y\leq ½)$. *Also*, $P(a<X<b) = P(a+{}^1/_2\leq Y\leq b-{}^1/_2)$.

Example: In 100 throws of a *fair* coin, what is the *probability* that the number of heads is (i) between 40 and 55; (ii) more than 36. A: We have $n = 100 > 30$, and $p = ½$, so $np = 50 = nq > 5$. Therefore we can use e.g. $P(40\leq X\leq 55) = P(39.5\leq Y\leq 55.5)$.

Standardise: $P(39.5 \leq Y \leq 55.5) = P({}^{39.5-50}/_5 \leq Z \leq {}^{55.5-50}/_5)$. ($npq = 25$ so the s.d. is $\sqrt{25} = 5$). So $P(-2.1 \leq Z \leq 1.1) = \Phi(1.1)-\Phi(-2.1) = 0.8643-(1-\Phi(2.1)) = \ldots = 0.8644$. *Similarly*, $P(X \geq 36) = P(Y \geq 35.5) = P(Z \geq {}^{35.5-50}/_5) = P(Z \geq -2.9) = 1-P(Z\leq -2.9) = 1-\Phi(-2.9) = 1-(1-\Phi(2.9)) = 0.091$.

Moments of 2 r.v.'s

The *bivariable* moment of order r in X and order s in Y, about the origin, is defined by $\mu_{r,s}' = E(X^r.Y^s) = \{\Sigma_{xi}\Sigma_{yj}\, \mathbf{x_i}^r\mathbf{y_j}^s\, p(x_i, y_j); \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^ry^sf(x,y)dxdy\}$. In **this** notation, the mean of X is $\mu_x = E(X^1Y^0) = \mu_{1,0}'$; the mean of Y is $\mu_Y = E(X^0.Y^1) = \mu_{0,1}'$. Now $E(X^1,Y^0) = \{\Sigma_{xi}\Sigma_{yj}\, \mathbf{x_i}\, p(x_i, y_j) = \Sigma_{xi}\, x_i\, p_1(x_i); \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x\, f(x,y)dxdy = \int_{-\infty}^{\infty} x\, f_1(x)dx\}$. Note: $\Sigma_{xi}\Sigma_{yj}\, \mathbf{x_i}\, p(x_i, y_j) = \Sigma_{xi}\, x_i(\Sigma_{yj}\, p(x_i,y_j))$. (The red bit is $p_1(x_i)$).

CENTRAL = about the mean. The *bivariable* central moment of order r in X and order s in Y is $\mu_{r,s} = E((X-\mu_x)^r(Y-\mu_Y)^s)$. In this *notation*, $\sigma_x^2 = \mu_{2,0}$ and $\sigma_y^2 = \mu_{0,2}$.

5th May 1999

Covariance

$\sigma_x^2 = \mu_{2,0}$; $\sigma_y^2 = \mu_{0,2}$. The covariance between X and Y is $Cov(X,Y) = \mu_{1,1} = E((X-\mu_X)(Y-\mu_Y))$. The *correlation* between X and Y (a dimensionless measure) is $\rho_{XY} = Cov(X,Y)/\sigma_X\sigma_Y$. Note: The *correlation* between X and Y is the covariance between their standardised versions. So $\rho_{XY} = E((X-\mu_X)(Y-\mu_Y))/\sigma_X\sigma_Y = E((X-\mu_X/\sigma_X)(Y-\mu_Y/\sigma_Y)) = E((X_{st}-\mu_{st})(Y_{st}-\mu_{st})) = Cov(X_{st}, Y_{st})$. [Where $X-\mu_X/\sigma_X = X_{st}$; $\mu_{st} = 0$; $\sigma_{st} = 1$].

Theorem: $\mu_{1,1} = Cov(X,Y) = E(XY)-E(X)E(Y)$. This is an easy way to *calculate* Cov(X,Y). **Proof**: $Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = E(XY-\mu_XY-\mu_YX+\mu_X\mu_Y) = E(XY)-\mu_XE(Y)-\mu_YE(X)+\mu_X\mu_Y = E(XY)-\mu_X\mu_Y-\mu_Y\mu_X+\mu_X\mu_Y = E(XY)-\mu_Y\mu_X = E(XY)-E(X)E(Y)$.

Example: Let (X,Y) be a *discrete* 2-D r.v. as shown in the table. Calculate the *covariance* and correlation. To do this, we **need** μ_x, μ_y, E(XY), σ_x, and σ_y. Now $\mu_x = \Sigma_{xi}\Sigma_{yj} = x_ip(x_i,y_j) = 2.{}^{2}/_{15} + 1.{}^{8}/_{15} + 0.{}^{5}/_{15} = {}^{12}/_{15} = {}^{4}/_{5}$. And $\mu_y = \Sigma_{xi}\Sigma_{yj} = y_jp(x_i,y_j) = 0.{}^{2}/_{15}+1.{}^{8}/_{15}+2.{}^{5}/_{15} = {}^{18}/_{15} = {}^{6}/_{5}$. And $E(XY) = \Sigma_{xi}\Sigma_{yj}\, x_iy_jp(x_i,y_j) = 0.{}^{2}/_{15} + {}^{8}/_{15} + 0.{}^{5}/_{15} = {}^{8}/_{15}$.

(x_i,y_j)	(2,0)	(-1,1)	(0,2)
$p(x_i,y_j)$	${}^{2}/_{15}$	${}^{8}/_{15}$	${}^{5}/_{15}$

Now $\sigma_x^2 = E(X^2)-[E(X)^2]$. $E(X^2) = 2^2.{}^{2}/_{15}+{}^{8}/_{15} = {}^{16}/_{15}$. $E(Y^2) = {}^{8}/_{15}+{}^{20}/_{15} = {}^{28}/_{15}$. So $\sigma_x^2 = {}^{16}/_{15}-({}^{4}/_{5})^2 = {}^{32}/_{75}$. And $\sigma_y^2 = {}^{28}/_{15}-({}^{6}/_{5})^2 = {}^{32}/_{75}$. So $Cov(X,Y) = E(XY)-E(X)E(Y) = {}^{8}/_{15}-({}^{4}/_{5})({}^{6}/_{5}) = -{}^{32}/_{75}$. Note: covariance **CAN** be -ve, but variance can never be -ve. Now $\rho_{XY} = Cov(X,Y)/\sigma_x\sigma_y = {}^{-32/75}/_{\surd(32/75)\surd(32/75)} = -1$. So X and Y are *correlated*. (ρ_{XY} is in the range $-1 \le \rho_{XY} \le 1$; if $\rho_{XY} = -1$ it is negatively *correlated*; if $\rho_{XY} = 1$ it is positively *correlated*).

Theorem: If X and Y are *independent*, then $Cov(X,Y) = 0 = \rho_{XY}$. Proof (Discrete Case): X and Y are *independent*, therefore $p(x_i,y_j) = p_1(x_i).p_2(y_j)$ for all x_i,y_j. Now $Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = \Sigma_{xi}\Sigma_{yj}\,(x_i-\mu_x)(y_j-\mu_y)p(x_i,y_j) = \Sigma_{xi}\Sigma_{yj}(x_i-\mu_x)(y_j-\mu_y)p_1(x_i)p_2(y_j) = (\Sigma_{xi}(x_i-\mu_x)p_1(x_i))(\Sigma_{yj}(y_j-\mu_y)p_2(y_j)) = (\Sigma_{xi}\,x_ip_1(x_i)-\mu_x\Sigma_{xi}\,p_1(x_i))(\Sigma_{yj}\,y_jp_2(y_j) - \mu_y\Sigma_{yj}\,p_2(y_j))$. Now **because** $\Sigma_{xi}x_ip_1(x_i) = \mu_x$ and $\Sigma_{xi}p_1(x_i) = 1$ (Similarly for y), **then** $Cov(X,Y) = (\mu_x-\mu_x.1)(\mu_y-\mu_y.1) = 0$. Hence $\rho_{XY} = 0/? = 0$.

Counter Example. Let (X,Y) be a 2-D r.v. with *p.m.f.* as shown in the table. This is not an independent r.v., as $p(-1,1) = 0$, *which* is not equal to $p_1(-1)p_2(1) = {}^{1}/_{3}.{}^{1}/_{3}$.

$y_j \setminus x_i$	-1	0	1	$p_2(y_j)$
0	${}^{1}/_{3}$	0	${}^{1}/_{3}$	${}^{2}/_{3}$
1	0	${}^{1}/_{3}$	0	${}^{1}/_{3}$
$p_1(x_i)$	${}^{1}/_{3}$	${}^{1}/_{3}$	${}^{1}/_{3}$	

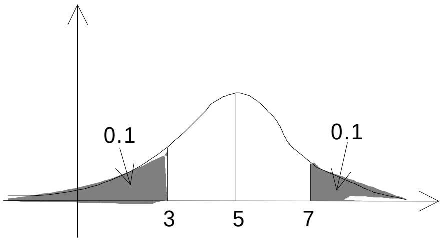
To **get** Cov(XY), we *need* $E(X) = -{}^{1}/_{3}+{}^{1}/_{3} = 0$; $E(Y) = 0.{}^{2}/_{3} + {}^{1}/_{3} = {}^{1}/_{3}$; $E(XY) = 0.-1.{}^{1}/_{3} + 0.0.0 + 0.1.{}^{1}/_{3} + 1.-1.{}^{1}/_{3} + 1.-1.0 + 1.0.{}^{1}/_{3} + 1.-1.0 = 0$. So $Cov(XY) = E(XY)-E(X)E(Y) = 0-0.{}^{1}/_{3} = 0$. Hence the **converse** to the theorem is *false*.

6th May 1999

Tutorial

Q: Find how many **throws** with one die are required so that the *probability* of getting 9 or more sixes is ½. A: Here $n = ?$, $p = {}^{1}/_{6}$, so $q = {}^{5}/_{6}$. Approximation: $X \sim N({}^{1}/_{6}n, {}^{5}/_{36}n)$. We want $P(X \geq 9) = ½$, so $P(X \leq 9) = 1-{}^{1}/_{2}$. Change 9 to 8.5 because of the *continuity* condition. So $P(X \leq 8.5) = ½$. Now $Z = (8.5-{}^{1}/_{6}n)/(\sqrt{({}^{5}/_{36})n})$. We want $P(Z) = ½$. This *implies* that $Z = 0$ from tables. So $0 = (8.5-{}^{1}/_{6}n)/(\sqrt{({}^{5}/_{36})n})$; $n = 51$.

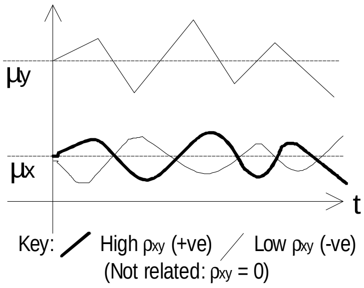
Q: Let X be a **continuous-valued** random variable of unknown distribution which has mean 5, and is *symmetric* abou the mean. Knowing that the probability of X being greater than 7 is 0.1, find a **lower** bound for the variance of X. A: We can **create** the graph from the information. So $P(2 \leq |X-5|) \leq 0.2$ Now from *Chebyshev*, $0.2 = {}^{1}/_{k^2}$, so $k = \sqrt{5}$. And $2 = \sigma k$; so $\sigma = {}^{2}/_{\sqrt{5}}$; $\sigma = 0.894$; or $\sigma^2 = {}^{4}/_{5}$.



Q: A manufacturer of light bulbs finds that *on average*, 2% are defective. What is the probability that (a) out of 1000 such bulbs selected at random, 15 or more are defective?; (b) exactly 5 are **defective** from a box of 100? A: $n = 1000$, $p = {}^{2}/_{100}$, so $q = {}^{98}/_{100}$. So $X \sim N(20, {}^{98}/_{5})$. We want $P(X \geq 15) = 1-P(X \leq 15) = 1-P(Z \leq {}^{15-20}/_{\sqrt{(98/5)}}) = 1-P(Z \leq {}^{14.5-20}/_{\sqrt{(98/5)}})$ by the continuity *condition*. So $1-P(Z \leq {}^{14.5-20}/_{\sqrt{(98/5)}}) = 1-P(Z \leq -1.24) = 1-\Phi(-1.24) = 1-(1-\Phi(1.24)) = 0.8925$. (b) $n = 100$, $p = {}^{2}/_{100}$, so $q = {}^{98}/_{100}$. $P(X = 5)$ by the *binomial* theorem $= ({}^{100}{}_{5})({}^{98}/_{100})^{95}({}^{2}/_{100})^{5} = 0.184 \times 10^{-3}$.

11th May 1999

Theorem: Let X and Y be r.v.'s. Define $Z = \alpha+aX+bY$ (α, a, b are constants). Then $\mu_Z = \alpha+a\mu_X+b\mu_Y$ and $\sigma_Z^2 = a^2\sigma_x^2+b^2\sigma_y^2+2ab\text{Cov}XY$. *Proof*: $\mu_Z = E(Z) = E(\alpha+aX+bY) = E(\alpha)+E(aX)+E(bY) = \alpha+aE(X)+bE(Y) = \alpha+a\mu_X+b\mu_Y$. And $\sigma_Z^2 = E((Z-\mu_Z)^2) = E((\alpha+aX+bY-\alpha-a\mu_X-b\mu_Y)^2) = E((a(X-\mu_X)+b(Y-\mu_Y))^2 = E(a^2(X-\mu_x)^2 + b^2(Y-\mu_y)^2 + 2ab(X-\mu_x)(Y-\mu_y)) = a^2E(X-\mu_x)^2 + b^2E(Y-\mu_y)^2 + 2ab((X-\mu_x)(Y-\mu_y)) = a^2\sigma_x^2+b^2\sigma_y^2+2ab\text{Cov}XY$. **Corollary**: If X & Y are *independent*, then $\sigma_z^2 = a^2\sigma_x^2+b^2\sigma_y^2$. **Now** we know that $|\rho_{XY}| \leq 1$. *Analysis*: $\rho_{XY} = 1$ = totally positively correlated. 0 = not correlated. -1 = totally negatively correlated.



Regression Functions

Let (X,Y) be a r.v. Then the **regression** function Y on X ("Y given X") is *defined* as: For **discrete** variables, $E(Y|x) = \Sigma_{y_j} y_j.p(y_j|x_i) = \Sigma_{y_j} y_j.{}^{p(x_i,y_j)}/_{p_1(x_i)}$. For **random** variables, $E(Y|x) = \int_{-\infty}^{\infty} y.f(y|x)dy = \int_{-\infty}^{\infty} y.{}^{f(x,y)}/_{f_1(x)}dy$. For each **particular** X, the regression function of Y on X gives *the* mean of Y for that X. So $E(Y|X) = \phi(x)$.

Similarly, the regression *function* of X on Y is $E(X|y) = \psi(y)$. *Discrete*: $E(X|y) = \Sigma_{x_i} x_i\, p(x_i|y_j) = \Sigma_{x_i} x_i.{}^{p(x_i,y_j)}/_{p_2(y_j)}$. Continuous: $E(X|y) = \int_{-\infty}^{\infty} x.f(x|y)dx = \int_{-\infty}^{\infty} x^{f(x,y)}/_{f_2(y)}dx$. If $f(x) = aX+b$, we say that the regression f^n of Y on X is *linear*. Note: a and b are called the regression coefficients.

Example: Let (X,Y) be a r.v. with *pmf* as shown in the table. Find $\phi(x)$ and $\psi(y)$. A: As in the table. Note: **Example** calculations: $\phi(0) = {}^{18}/_{2}(-{}^{3}/_{18}) = -{}^{3}/_{2}$. And $\phi(1) = {}^{18}/_{4}(-{}^{4}/_{18}+{}^{8}/_{18}) = 1$, ...

$Y \backslash X$	**0**	**1**	**2**	**3**	$p_2(y_j)$	$\psi(y_j)$
-3	$^1/_{18}$	0	$^2/_{18}$	0	$^3/_{18}$	$^4/_3$
-2	0	$^2/_{18}$	0	$^3/_{18}$	$^5/_{18}$	$^{11}/_5$
0	$^1/_{18}$	0	$^1/_{18}$	0	$^2/_{18}$	1
4	0	$^2/_{18}$	0	$^1/_8$	$^3/_{18}$	$^5/_3$
8	0	0	$^1/_{18}$	$^4/_{18}$	$^5/_{18}$	$^{14}/_5$
$p_1(x_i)$	$^2/_{18}$	$^4/_{18}$	$^4/_{18}$	$^8/_{18}$		
$\phi(x_i)$	$-^3/_2$	1	$^1/_2$	$^{15}/_4$		

The Central Limit Theorem (CLT)

If $X_1, ..., X_n$ are *independent* r.v.'s, with the **same** distribution — mean μ, variance σ^2, and if $\bar{X} = {}^1/_n \Sigma^n_{i=1} x_i$, then $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ (the *standardised* form of $\bar{X}$) is *approximately* distributed as N(0,1) for large n ($n \geq 30$).

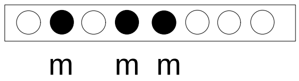
12th May 1999

Tutorial

Q: What is the **probability** that a randomly picked integer N produces a number *ending* with a 1, if (a) N is squared; (b) N is multiplied by another arbitrary integer? A: To get a square **ending** in 1, the original number must end in a 1 or a 9. Each *last* digit can take 10 values, 2 of which give a 1 at the end when squared. So the probability is $^2/_{10} = {}^1/_5$. Two numbers multiplied **together** resulting in a number ending with a 1 can *only* be obtained when the last numbers of the two numbers are one of (1,1), (3,7), (7,3) and (9,9). So the probability is $^4/_{100} = {}^1/_{25}$.

Q: You have 10 cards with the **digits** 0 to 9 on them. Find the *probability* that an arbitrary 2-digit number made out of *two* of these cards divides by 18? A: We have a sample **space** of size 81 — we must have a TWO digit number, so cannot have 00, 11, 22, etc. The numbers divisible by 18 are 18, 36, 54, 72 and 90. So $P(A) = {}^{|A|}/_{|S|} = {}^5/_{81}$.

Q: There are n+k seats in a theatre. If n spectators come in, what is the probability that some fixed number of seats, $m \leq n$, are *taken*? A **All** possibilities: $\binom{n+k}{n}$. Of these, the number of *possibilities* where the m fixed **seats** are taken is $\binom{n+k-m}{n-m}$. So P(m seats are *taken*) = $\binom{n+k-m}{n-m}/\binom{n+k}{n}$.

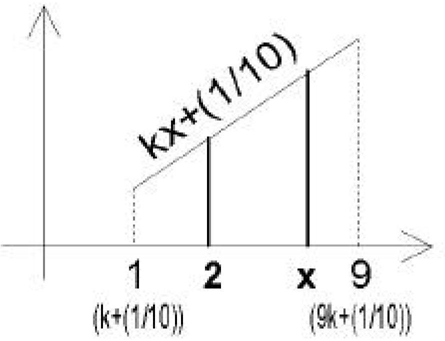


HERE, n = 4, k = 4, m = 3

Q: 3 cards are drawn from a **deck** of 52 cards. Find the probability that **these three** cards are *3, 7 and an Ace*. A: Let A = "Draw 5, 7, A". $|A| = 4\times4\times4 = 64$. The number of *possibilities* of taking 3 cards out of 53, regardless of the order, is $\binom{52}{3} = {}^{52\times51\times50}/_{1\times2\times3} = 22100$ possibilities. So $P(A) = {}^{|A|}/_{|S|} = {}^{64}/_{22100}$.

Q: Let A and B be events on some **sample** space S. If $P(A) = {}^3/_5$, $P(B) = {}^1/_8$, and $P(\bar{A}|\bar{B}) = {}^3/_7$, find $P(A \cup B)$ and $P(A \cap B)$. A: $P(\bar{A}|\bar{B}) = \frac{P(\bar{A}\cap\bar{B})}{P(\bar{B})} = \frac{P(A\cup B)^c}{1-P(B)} = \frac{1-P(A\cup B)}{1-P(B)}$. So $1-P(A\cup B) = P(\bar{A}|\bar{B}) \times (1-P(B)) = {}^3/_7(1-{}^1/_8) = {}^3/_8$. So $P(A\cup B) = 1-{}^3/_8 = {}^5/_8$. We obtain $P(A\cap B)$ by *using* $P(A\cup B) = P(A)+P(B)-P(A\cap B)$.

Q: Let X be a **random** variable with a p.d.f. $f(x) = kx+{}^1/_{10}$ for $x \in [1,9]$, and 0 elsewhere. Find k, and the *value* x such that $P(2 \le X \le x) = 0.3$. A: We know that the area under the graph is 1. Using the *trapezium* rule, $1 = ({}^1/_2)((k+{}^1/_{10})+(9k+{}^1/_{10}))(9-1)$; $1 = 4((k+{}^1/_{10})+(9k+{}^1/_{10}))$; ... $k = {}^1/_{200}$. Now find $P(2 \le X \le x) = 0.3$.



Two methods: (a) use $\int_2^x ({}^1/_{200}X+{}^1/_{10})dX$ which gives a quadratic which we *calculate* x from. (b) At X = 2, the height is ${}^2/_{200}+{}^1/_{10}$. At X = x, the height is ${}^x/_{100}+{}^1/_{10}$. Using the *trapezium* rule again, we get $0.3 = ({}^1/_2)(({}^2/_{100}+{}^1/_{10})+({}^x/_{100}+{}^1/_{10}))(x-2)$, ..., x = 4.4.

Q: You try in a £1 lottery with 10 numbers. Two numbers are *drawn* from the urn. If you match them, you win £100. If you match them in the reverse order, you win £10. (a) Calculate your **expected** profit in 1 game; (b) What is the least number of games you have to play so that P(Get £100) > 0.04; (c) What is the *expected* number of games you have to play so that you get £10 3 times; (d) What is the probability that you played *exactly* 5 games in part (c)?

X = Profit	p
-1	${}^{88}/_{90}$
9	${}^1/_{90}$
99	${}^1/_{90}$

A: Think about the situation, and you arrive at the *table* shown. So $E(X) = -{}^{88}/_{90}+{}^9/_{90}+{}^{99}/_{90} = {}^{20}/_{90} = {}^2/_9$. In (b), use **Negative** Binomial.

Exam Paper: June 1999

SECTION 1 (Compulsory)

(1) (a) Two different integers are chosen at random from the integers from 1 to 11 inclusive. If their sum is even, find the probability that both numbers are odd.

[5 marks]

(b) Let S be an equiprobable sample space with N elements, and let A and B be events on S with $P(A) = {}^{1}/_{3}$ and $P(B) = {}^{1}/_{5}$. What is $P(A \cup B)$ if

(i) A and B are independent events?
(ii) A and B are mutually exclusive events?
(iii) $(B \backslash \bar{A})$ contains exactly 4 points from S?
(iv) A and B are independent events and S is *not* equiprobable? **[5 marks]**

(c) Let S be a sample space and A and B be events on S. Prove that $\forall$ A, B, such that $A \subset B$, $P(B) \geq P(A)$. **[5 marks]**

(d) You play a chance game by taking one card from a deck of 52 cards without replacement and keep the drawn cards in a pile. When you draw a card so that in your pile you have either 2 Kings, a King and a Queen, or 2 Queens, you win and the game is over.

(i) What is the probability that you win by drawing exactly 10 cards?
(ii) What is the probability that you need to draw at most 3 cards to win?

[5 marks]

SECTION 2 (Answer 2 out of 4 questions)

(2) A biased coin with P(B) = 0.3 is tossed 3 times. Let (X, Y) be a two-dimensional random variable with X = the number of heads thrown and Y = the number of tails in the first throw.

(a) Give the probability mass function (p.m.f.) of (X, Y).

(b) Find the marginal p.m.f.'s.

(c) Find the marginal distribution functions.

(d) Are X and Y dependent?

(e) Find the expected value of (X, Y).

[15 marks]

(3) Six married couples are standing in a room. If two people are chosen at random, find the probability that

(a) They are married to each other.

(b) One is male and the other is female.

(c) If a committee of 5 is chosen at random, what is the probability that the committee does not include two people who are married to each other?

[15 marks]

(4) Let X be a continuous-valued random variable with a probability density function f(x) > 0 for x ∈ [a, b] and 0, elsewhere.

(a) Prove that the cumulative distribution function F(x) is strictly increasing in (a, b).

(b) Find the value of the constant k for

$$f(x) = \begin{cases} kxe^{-x^2/4}, & x \in [a,b] \\ 0, & elsewhere. \end{cases}$$

(c) For a = 0, b = 1, find x such that $P(X \geq x) = 0.7$.

[15 marks]

(5) (a) Of 100 boxes of fuses, 5 fuses per box, 20 boxes contain fuses from factory A, 30 boxes from factory B, and 50 boxes from factory C. Fuses from factory A are, on average, 5% defective; from factory B, 4%, and from factory C, 2%. The fuses and boxes all look alike and are piled without regard of place of manufacture. A box is selected at random and a fuse in it is tested and found to be defective. What is the probability that it was produced at factory B?

(b) Five red books and 3 green books are placed at random on a shelf. Find the probability that the green books will all be together.

[15 marks]

(Questions done: 1, 2, 5)