30th January 2001

# Section 1: Probability Space

*Probability Theory* provides Mathematical Tools for dealing with **experiments** whose outcomes cannot be predicted with certainty (*random* experiments). A random experiment has more than one outcome, and, on each performance, one, and **only** one, outcome occurs. The sample space S is the set of *all possible outcomes* of a random experiment.

S may be **finite**, **countably infinite,** or **non-countable**. *Sample Points* — e, w, etc., elements of S (i.e. outcomes) Subsets of S are denoted by A, B, ... Probability Model: (1) S, (2) A measure of *likelihood* of *occurrence* of an outcome or a set of outcomes — the **probability** of the outcome.

## 1.1: Borel Field

Associated with S, there is a **non-empty** class $\mathcal{F}$ of subsets $A_1, A_2, \dots, A_n, \dots$ of S with the *following* properties: **(1)** if $A \in \mathcal{F}$, then $A' \in \mathcal{F}$ (A' is the *complement* of A in S ($=\bar{A}$ or $A^C$)); and **(2)** if $\{A_n\}$ is an *infinite* sequence of subsets of $\mathcal{F}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$. Note that $\mathcal{F}$ is called a **Borel** field (or a $\sigma$-field), and its elements are called *events*. So an event A is a **subset** of S such that $A \in \mathcal{F}$. (In *measure theory*, elements of $\mathcal{F}$ are called measurable sets, so that an event is a **measurable** set).

**Note**: If S is *countable*, then every subset of S is an event. If S is *non-countable*, then we can, by complicated mathematical techniques, construct subsets of S which do not satisfy the *axioms* of probability. These subsets are called non-measurable sets, and are so different from our events, measurable sets, that we want to **exclude** them from our study of events. We do this by requiring our events to *satisfy* the *axioms* of a Borel field.

**Theorem 1.1**: Let $\{A_n\}$ be an *infinite sequence of events* which are elements of $\mathcal{F}$. Then **(i)** $S \in \mathcal{F}$ **(ii)** $\phi \in \mathcal{F}$ **(iii)** $\cap_{n=1}^{\infty} A_n \in \mathcal{F}$ **(iv)** $\cup_{n=1}^{p} A_n \in \mathcal{F}$ **(v)** $\cap_{n=1}^{p} A_n \in \mathcal{F}$ **(vi)** $A_i \backslash A_j \in \mathcal{F}$ (where i, j = 1, 2, 3, ...., and p is a *finite +ve integer*). *Proofs:* (i) $\exists\, A \in \mathcal{F}$, $A' \in \mathcal{F}$ by (1) ($\mathcal{F}$ is non-empty). *Put* in property (2): $A_1 = A$, $A_2 = A'$, $A_3 = A_4 = \dots = \phi$ — *then* $\cup_{n=1}^{\infty} A_n = A \cup A' \cup \phi \cup \phi \cup \dots = A \cup A' = S \in \mathcal{F}$.

(ii) (1) $\Rightarrow S' = \phi \in \mathcal{F}$. (iii) $A_n \in \mathcal{F}$, n = 1, 2, ... $\Rightarrow$ by (1) $A_n' \in \mathcal{F}$, n = 1, 2, ... **By** de Morgan's law, and *property* (2), $[\cap_{n=1}^{\infty} A_n]' = \cup_{n=1}^{\infty} A_n' \in \mathcal{F}$. The result *follows* from property (1). **Aside**: de Morgan's Laws are (i) $[\cup_{n=1}^{\infty} A_n]' = \cap_{n=1}^{\infty} A_n'$; *and* (ii) $[\cap_{n=1}^{\infty} A_n]' = \cup_{n=1}^{\infty} A_n'$. (iv) Put in *property* (2) $A_{p+1} = A_{p+2} = \dots = \phi$. (v) *Put* in (iii) $A_{p+1} = A_{p+2} = \dots = S$. (vi) $A_i \backslash A_j = A_i \cap A_j' \in \mathcal{F}$ using *property* (1) and (v).

1st February 2001

**Definition**: An event A *occurs* if outcome $e \in A$. Event A *does not occur* if $e \notin A$. S is the **sure** event (sure to occur), and $\phi$ is the **impossile** event (impossible to occur). $A \subset B$ if *the occurrence* of A $\Rightarrow$ the occurrence of B. A = B if $A \subset B$ *and* $B \subset A$. ($A \subset B$ does **not** exclude A = B}. $A \cup B = \{e \in S \mid e \in A \text{ or } e \in B \text{ or both}\}$, the *event* which occurs if at least **1** of A or B occurs. *Similarly* for $\cup_n A_n$.

$A\cap B = \{e \in S \mid e \in A \text{ and } e \in B\}$, the event which occurs if **both** A and B occur. *Similarly for* $\cap_n A_n$. $A\backslash B = \{e \in S \mid e \in A \text{ and } e \notin B\}$, the event which **occurs** if A occurs and B does not. $A' = \{e \in S \mid e \notin A\}$ *occurs* if A does not occur. **Clearly**, $A\backslash B = A\cap B'$, and $A' = S\backslash A$. A and B are *exclusive* if $A\cap B = \phi$ — can **not** occur together.

$\{A_n\}$ is a *pairwise exclusive* sequence if $A_i\cap A_j = \phi$ when $i \neq j$. $\{A_n\}$ forms a *partition* of S if $A_i\cap A_j = \phi$, $i \neq j$, and $\cup_n A_n = S$. **Cartesian Product:** $\times_n S_n = \{(e_1, e_2, ..., e_n) \mid e_1 \in S_1, e_2 \in S_2, ...\}$. *Upper Limit of* $\{A_n\}$: $\limsup_{n\to\infty} A_n = \bar{A} = \cap_{N=1}^{\infty}\{\cup_{n=N}^{\infty} A_n\}$. *Lower Limit of* $\{A_n\}$: $\liminf_{n\to\infty} A_n = \underline{A} = \cup_{N=1}^{\infty}\{\cap_{n=N}^{\infty} A_n\}$.

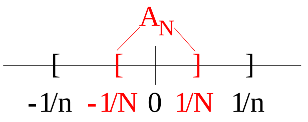
**Notes**: (1) $a \in \bar{A}$ iff "a belongs to *infinitely many* of the $A_n$". (2) $a \in \underline{A}$ iff "a **belongs** to all but a **finite** number of the $A_n$". **Proofs**. (1) $\bar{A} = \cap_{N=1}^{\infty}\{\cup_{n=N}^{\infty} A_n\}$. Call the red bit $B_N$. So $B_1 \supset B_2 \supset B_3$...., and $a \in \bar{A} \Leftrightarrow a \in B_N\ \forall N$. (i) $a \in B_N\ \forall N \Rightarrow$ "a belongs to *infinitely* many of the $A_n$"; otherwise $a \notin A_n\ \forall n \geq n_1$ for some $n_1 \Rightarrow a \notin B_{n_1}$, *. (ii) "a belongs to *infinitely many* of the $A_n$" $\Rightarrow a \in B_N\ \forall N$ (*otherwise*, if $a \notin B_{N'}$, for some N', then $a \notin A_n\ \forall n \geq N' \Rightarrow$ "a does not belong to *infinitely* many of the $A_n$" *. Hence $a \in \bar{A} \Leftrightarrow a \in B_N\ \forall N \Leftrightarrow$ "a belongs to *infinitely* many of the $A_n$".

(2) $\underline{A} = \cup_{N=1}^{\infty}\{\cap_{n=N}^{\infty} A_n\}$ (= $\cup_{N=1}^{\infty} C_N$, where the blue bit is $C_N$). Here, $C_1 \subset C_2 \subset C_3$...., and $a \in \underline{A} \Leftrightarrow a \in C_{N'}$ for *some* N'. (i) $a \in C_{N'}$ for some N' $\Rightarrow$ "a belongs to all **but** a finite number of the $A_n$". (ii) "a belongs to all but a **finite** number of the $A_n$" $\Rightarrow a \in A_n\ \forall n \geq N'$ for some N' $\Rightarrow a \in C_{N'}$. Hence $a \in \underline{A} \Leftrightarrow a \in C_{N'}$ for some N' $\Leftrightarrow$ "a belongs to *all but a finite number* of the $A_n$".

**Note (3)**: From (1) and (2) *above*, $\underline{A} \subset \bar{A}$ ($a \in \underline{A} \Rightarrow$ "a belongs to *all but a finite number* of $A_n$" $\Rightarrow$ "a belongs to an **infinite** number of the $A_n$" $\Rightarrow a \in \bar{A}$). If $\underline{A} = \bar{A}$, then the event is *called* the limit of $\{A_n\}$, and is **written as** $\lim_{n\to\infty} A_n$ — and the *sequence* $\{A_n\}$ is called **convergent**. **Theorem 1.2:** If $\underline{A} \in F$, if $\bar{A} \in F$, and if $\{A_n\}$ is *convergent*, then $\lim_{n\to\infty} A_n \in F$. **The proof** comes from *Theorem 1.1*.

6th February 2001

**Put** $A_n = \{x \in \mathbf{R} \mid |x| \leq {}^1/_n\}$. Here, $\cup_{n=N}^{\infty} A_n = A_N$, so that $\limsup_{n\to\infty} A = \bar{A} = \cap_{N=1}^{\infty}(\cup_{n=N}^{\infty} A_n) = \cap_{N=1}^{\infty} A_N = \{0\}$. Now $\liminf_{n\to\infty} A_n = \underline{A} = \{0\}$ for all N $\Rightarrow \bar{A} = \{0\} = \underline{A}$, i.e. $\{A_n\}$ is *convergent*, and $\lim_{n\to\infty} = \{0\}$. **Example**: Put $A_n = \{(-1)^n\}$, e.g. $A_2 = \{1\}$, and $A_5 = \{-1\}$. *Here*, $\cup_{n=N}^{\infty} A_n = \{-1,1\}\ \forall N$. **Hence** $\bar{A} = \cap_{N=1}^{\infty}\{-1,1\} = \{-1,1\}$. *However*, $\cap_{n=N}^{\infty} A_n = \phi$. Therefore, $\underline{A} = \cup_{N=1}^{\infty}\phi = \phi$. Thus $\bar{A} \neq \underline{A}$, so that $\{A_n\}$ is *not convergent* — and so does **not** possess a limit.



**Theorem 1.3:** If $\{A_n\}$ is *monotonic*, then it is *convergent*. (The **converse** is false). ($\{A_n\}$ is monotonic *increasing* (denoted by $\uparrow$) if $A_i \subset A_{i+1}\ \forall i$; and $\{A_n\}$ is monotonic *decreasing* ($\downarrow$) if $A_i \supset A_{i+1}\ \forall i$). (i) If $\{A_n\} \uparrow$, then $\lim_{n\to\infty} A_n = \cup_{N=1}^{\infty} A_N$. (ii) If $\{A_n\} \downarrow$, then $\lim_{n\to\infty} A_n = \cap_{N=1}^{\infty} A_N$. **Proof:** If $\{A_n\} \uparrow$, then $A_i \subset A_{i+1}\ \forall i$. So $\cap_{n=N}^{\infty} A_n = A_N\ \forall N$, and $\liminf_{n\to\infty} A_n = \cup_{N=1}^{\infty}\{\cap_{n=N}^{\infty} A_n\} = \cup_{N=1}^{\infty} A_N$. *Further*, $\cup_{N=1}^{\infty} A_N \supset \cup_{n=N}^{\infty} A_n\ \forall N$. So $\liminf_{n\to\infty} A_n = \cup_{N=1}^{\infty} A_N \supset \cap_{N=1}^{\infty}\{\cup_{n=N}^{\infty} A_n\} = \limsup_{n\to\infty} A_n = \bar{A}$, i.e. $\underline{A} \supset \bar{A}$. From *note 3 before*, $\underline{A} \subset \bar{A}$, so that $\underline{A} = \bar{A} = \cup_{N=1}^{\infty} A_N$.

(ii) If $\{A_n\} \downarrow$, then $A_i \supset A_{i+1}$ $\forall$ i, and so $\cup_{n=N}^{\infty} A_n = A_N$ $\forall$ N. Next, $\overset{\text{lim sup}}{}_{n\to\infty} A_n = \cap_{N=1}^{\infty}\{\cup_{n=N}^{\infty} A_n\} = \cap_{N=1}^{\infty} A_N$. **Further,** $\cap_{N=1}^{\infty} A_N \subset \cap_{n=N}^{\infty} A_n$ for **any** N, so that $\overset{\text{lim sup}}{}_{n\to\infty} A_n = \cap_{N=1}^{\infty} A_N \subset \cup_{N=1}^{\infty}\{\cap_{n=N}^{\infty} A_n\} = \underline{A}$. But by *note 3 again,* $\underline{A} \subset \bar{A}$ , **so that** $\underline{A} = \bar{A} = \cap_{N=1}^{\infty} A_n$. c.f. the **example** $A_n = [-{}^{1}/_{n}, {}^{1}/_{n}]$.

8th February 2001

**Theorem 1.4:** Let C be any class of events in a *sample space, S*. Then ∃ a unique smallest Borel field containing C, called the **Borel field generated by C**. **Proof:** The class of all events in S is *clearly* a Borel field containing C. Let $\mathcal{F}_0$ be the *intersection* of all the Borel fields of subsets of S *containing* C. Then $\mathcal{F}_0$ **is** a Borel field, *contains* the class C, and is, by definition, contained in *every* Borel field containing C, i.e. $\mathcal{F}_0$ is the **smallest** Borel field containing C. It's *unique*, since if $\mathcal{F}'$ is also a *smallest* Borel field containing C, then $\mathcal{F}_0 \subset \mathcal{F}'$ and $\mathcal{F}' \subset \mathcal{F}_0 \Rightarrow \mathcal{F}' = \mathcal{F}_0$.

**Example**: Toss a coin: S = {H, T}. The class of all *events* in S is $\mathcal{F}$, where $\mathcal{F} = \{\phi, \{H\}, \{T\}, \{H,T\}\}$ (4 **members)**. $\mathcal{F}$ is a Borel field (check that the *properties* of a Borel field are satisfied). Also, it is a Borel Field $\mathcal{F}_0$ generated by {H}. For C, *which* is {{H}}, $\{H\} \in \mathcal{F}_0 \Rightarrow \{H'\} = \{T\} \in \mathcal{F}_0$. Therefore, $\{H\} \cup \{T\} = S \in \mathcal{F}_0 \Rightarrow S' = \phi \in \mathcal{F}_0$, i.e. $\mathcal{F}_0 = \mathcal{F}$.

**Example**: Let S be finite, with n elements, then the *class of all events* in S is a Borel field $\mathcal{F}$ with $2^n$ members. $\mathcal{F}$ may be generated by the *class* of all 1-sample point events. Let $S = \{e_1, e_2, ..., e_n\}$. A list of all events in S: The impossible event, $\phi$; *1-sample* point events: $\{e_1\}, \{e_2\}, ..., \{e_n\}$ (1 event); *2-sample* point events: $\{e_1,e_2\}$ .... $\{e_{n-1},e_n\}$ ($({}^{n}{}_{2})$ events); ...; *n-sample* point events: $\{e_1, ..., e_n\}$ ($({}^{n}{}_{n}) = 1$ event). **Total** number of events $= 1 + ({}^{n}{}_{1}) + ({}^{n}{}_{2}) + ... + ({}^{n}{}_{n-1}) + 1 = (1+1)^n = 2^n$.

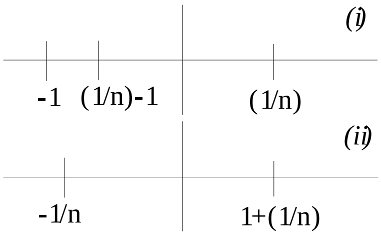
Let $\mathcal{F} = \{\phi, \{e_1\},...,\{e_n\}, \{e_1,e_2\}, ... ,\{e_1,e_2,...,e_n\}\}$. Clearly, $\mathcal{F}$ satisfies the *definition of a Borel field*. Further, $\mathcal{F}$ can be **generated** by $C = \{\{e_1\}, \{e_2\}, ..., \{e_n\}\}$, since on taking **finite unions** of the form $\cup_{k=1}^{r} \{e_{i_k}\} = \{e_{i_1}, e_{i_2}, ..., e_{i_r}\}$ for $1 \le i < i_2 < ... < i_r \le n$ (r = 2, 3, ..., n), we can obtain all the *r-sample* point events. Also, $\phi = S'$. If S is *countable,* then the class of all events in S is a Borel field which may be generated by the class of all *1-sample* point events. [**Notes**: $\{e_1\}, \{e_2\}, ... \in \mathcal{F}_0$; all events in S are obtained by taking **countable unions** of the type $\cup_k \{e_{i_k}\}$; and $\phi = S'$].

**Example**: Draw a ticket from a bag containing *4 tickets* numbered 1 to 4. Here, S = {1, 2, 3, 4}. The *Borel field* generated by the class C = {{1}, {3,4}} of events is given by $\mathcal{F}_0$ = {{1}, {3,4}, {2,3,4}, {1,2}, {1,3,4}, {1,2,3,4}, {2}, $\phi$} ($A_1, A_2, A_3 = A_1', A_4 = A_2', A_5 = A_1 \cup A_2, A_6 = A_1 \cup A_3, A_7 = A_5', A_8 = A_6'$).

**Example**: Let $\mathcal{F}_0$ be the Borel field generated *by the class of intervals* (a,b] on the real line. The **elements** of $\mathcal{F}_0$ are called Borel sets on **R** (the same $\mathcal{F}_0$ is *generated* if $^{*}$ (a,b] is replaced by the *class* of [a,b] or (a,b) or [a,b); or more generally by the class of **open** or **closed** sets on **R**). If, in a *random experiment*, we take the sample space S to be **R**, then real numbers will be *sample points* — and these Borel sets will be events. $^{*}$: Define $A_n = (b-{}^{1}/_{n}, b]$ for n = 1, 2, ..., ∞. It follows that $A_n \in \mathcal{F}_0$, and that $\cap_{N=1}^{\infty} A_N = b \in \mathcal{F}_0$. It can be shown that $\mathcal{F}_0$ does *not* contain all subsets of **R**!

# Tutorial

**Exercise 2**: Find the *upper and lower* limits of the following sequences of events: (i) $\{A_n\} = \{[{}^1/_n-1,\ {}^1/_n]\}$; and (ii) $\{A_n\} = \{[-{}^1/_n,\ 1+ {}^1/_n]\}$. **Exercise 3**: Given that $\lim_{n\to\infty} (1+{}^x/_n)^n = e^x$, find the *upper and lower limits* of the following sequence of sets: (i) $\{A_n\} = \{[-{}^1/_n,\ (1-{}^1/_n)^n)\}$; and (ii) $\{A_n\} = \{[({}^1/_n-1)^n,\ 1]\}$.



A: (i) $\cup_{n=N}^{\infty} A_n = (-1, {}^1/_N]$. $\limsup_{n\to\infty} A_n = \bar{A} = \cap_{N=1}^{\infty} (-1, {}^1/_N] =$ $(-1, 0]$. And $\cap_{n=N}^{\infty} A_n = [{}^1/_N-1, 0]$. Therefore, $\liminf_{n\to\infty} A_n = \underline{A} = \cup_{N=1}$ $[{}^1/_N-1, 0] = (-1, 0]$. Conclusion: $\underline{A} = \bar{A}$, so that the *limit* is $(-1,0]$. (ii $\cup_{n=N}^{\infty} A_n = [-{}^1/_N, 1+{}^1/_N]$. Now $\limsup_{n\to\infty} A_n = \bar{A} = \cap_{N=1}^{\infty} [-{}^1/_N, 1+{}^1/_N] =$ $[0, 1]$. And $\cap_{n=N}^{\infty} A_n = [0, 1]$. Therefore, $\liminf_{n\to\infty} A_n = \underline{A} = \cup_{N=1}^{\infty}$ [0, 1 $= [0, 1]$. So (*again)*, $\underline{A} = \bar{A}$.

(i) $n = 1 \Rightarrow [-1, 0]$. $n = 2 \Rightarrow [-{}^1/_2, \tfrac14]$. $n = 4 \Rightarrow [-{}^1/_4, ({}^3/_4)^4]$. So $\cup_{n=N}^{\infty} A_n = [-{}^1/_N, {}^1/_e)$. It follows that $\limsup_{n\to\infty} A_n = \bar{A} = \cap_{N=1}^{\infty}[-{}^1/_N, {}^1/_e) = [0, {}^1/_e)$. Now $\cap_{n=N}^{\infty} A_n = [0, {}^1/_e)$. So $\liminf_{n\to\infty} A_n = \underline{A} = \cup_{N=1}^{\infty}[0, {}^1/_e) = [0, {}^1/_e)$. So $\underline{A} = \bar{A} = [0, {}^1/_e]$, and *therefore* $\lim_{n\to\infty} A_n = [0, {}^1/_e)$. (ii) $n = 1 \Rightarrow [0, 1]$. $n = 2 \Rightarrow [{}^1/_4, 1]$. $n = 3 \Rightarrow [(-{}^2/_3)^3 = -{}^8/_{27}, 1]$. $n = 5 \Rightarrow [-0.32768, 1]$. So $\cup_{n=N}^{\infty} A_n = [-{}^1/_e, 1]$. It follows that $\limsup_{n\to\infty} A_n = \bar{A} = \cap_{N=1}^{\infty}[-{}^1/_e, 1] = [-{}^1/_e, 1]$. Further, $\cap_{n=N}^{\infty} A_n = [{}^1/_e, 1]$. So $\liminf_{n\to\infty} A_n = \underline{A} = \cup_{N=1}^{\infty} [{}^1/_e, 1] = [{}^1/_e, 1]$.

**Example 3**: Consider $B_N = \cap_{n=N}^{\infty} A_n$. Here, $B_n$ is $\Uparrow$, i.e. $B_1 \subset B_2 \subset B_3 \subset \ldots$, and $B_n$ is the set of *sample points* $\omega$, which, for some $N \geq 1$, belong to **all** $A_n$, where $n \geq N$. Now $\lim B_N = \cup_{N=1}^{\infty} B_N = \underline{A} = \cup_{N=1}^{\infty} \cap_{n=N}^{\infty} A_n = \liminf_{n\to\infty} A_n$, i.e. $\underline{A}$ is the *set of points* which belong to all except a finite number of events $A_1, A_2, \ldots$ Repeat the analysis of example 3 *when* $B_N = \cup_{n=N}^{\infty} A_n$.

**Theorem 2**: If $\{A_n\}$ is a *monotonic sequence of events*, then $\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n)$. **Theorem 3:** (The *Borel-Cantelli* lemma): If $\{B_n\}$ is a sequence of events, then (i) $\Sigma_{n=1}^{\infty} P(B_n) < \infty$ implies that $P(\limsup_{n\to\infty} B_n) = 0$; (ii) if $\{B_n\}$ are *mutually independent*, and if $\Sigma_{n=1}^{\infty} P(B_n) = \infty$, then $P(\limsup_{n\to\infty} B_n) = 1$.

(**Q1**) Describe a *suitable sample space* for the following random experiments: (i) tossing a coin (a) 4 times, and (b) n times; (ii) recording the number of *defectives* in a set of 1000 items; and (iii) testing the life span of a piece of *electronic* equipment. A: (i) S = {HHHH, HHHT, HHTH, ..., TTTT}, 16 *events*. (ii) S = {0, 1, ..., 1000}, the number of *defective* items. (iii) S = {0, 1, 2, ..., ∞}, the number of *seconds* of operation.

(**Q2**) Let A and B be *any* 2 events. Show, algebraically, that (i) $A \cap B$ and $A \backslash B$ are **mutually exclusive** — and that their union is A; (ii) show that B and $A \backslash B$ are *mutually exclusive* — and that their union is $A \cup B$; and (iii) show that $A \cap B$, $A \backslash (A \cap B)$, and $B \backslash (A \cap B)$ are *pairwise exclusive* — and that their union is $A \cup B$. A: (i) $(A \cap B) \cap (A \backslash B) = (A \cap B) \cap (A \cap B') = (A \cap A) \cap (B \cap B') = A \cap \phi = \phi$. **QED**. *Now* $(A \cap B) \cup (A \backslash B) = (A \cap B) \cup (A \cap B') = A$. **QED**.

(ii) $B\cap(A\backslash B) = B\cap(A\cap B') = (B\cap A)\cap(B\cap B') = (B\cap A)\cap\phi = \phi$. **QED**. *Now* $B\cup(A\backslash B) = B\cup(A\cap B') = (B\cup A)\cap(B\cup B') = (B\cup A)$. **QED**. (iii) Pair 1: $(A\cap B)\cap(A\backslash(A\cap B)) = (A\cap B)\cap(A\cap(A\cap B)') = [(A\cap B)\cap A]\cap[(A\cap B)\cap(A\cap B)'] = [(A\cap B)\cap A]\cap\phi = \phi$. **QED**. *Now* $(A\cap B)\cup(A\backslash(A\cap B)) = (A\cap B)\cup(A\cap(A\cap B)') = [(A\cap B)\cup A]\cap[(A\cap B)\cup(A\cap B)'] = [(A\cap B)\cup A] = A$. **QED**.

Pair 2: $(A\cap B)\cap(B\backslash(A\cap B)) = [(A\cap B)\cap B]\cap[(A\cap B)\cap(A\cap B)'] = [(A\cap B)\cap B)\cap\phi] = \phi$. **QED**. *Now* $(A\cap B)\cup(B\backslash(A\cap B)) = (A\cap B)\cup(B\cap(A\cap B)') = [(A\cap B)\cup B]\cap[(A\cap B)\cup(A\cap B)'] = (A\cap B)\cup B = B$. **QED**. Pair 3: $[A\backslash(A\cap B)]\cap[B\backslash(A\cap B)] = [A\cap(A\cap B)']\cap[B\cap(A\cap B)'] = [A\cap(A'\cup B')]\cap[B\cap(A'\cup B')]$ by *de Morgan* $= [(A\cap A')\cup(A\cap B')]\cap[(B\cap A')\cup(B\cap B')] = (A\cap B')\cap(B\cap A') = (A\cap A')\cap(B\cap B') = \phi\cap\phi = \phi$. **QED**. *Now* $[A\backslash(A\cap B)]\cup[B\backslash(A\cap B)] = [A\cap(A\cap B)']\cup[B\cap(A\cap B)'] = [A\cap(A'\cup B')]\cup[B\cap(A'\cup B')] = (A\cap B')\cup(B\cap A') = [A\cup(B\cap A')]\cap[B'\cup(B\cap A')] = [(A\cup B)\cap(A\cup A')]\cap[(B'\cup B)\cap(B'\cup A')] = (A\cup B)\backslash(A\cap B)$. **QED**.

(**Q3**) If two *unbiased dice* are thrown n times together, what is the probability of obtaining a double six at least **once**? What is the smallest value of n such that this probability *exceeds* 0.5? A: P(Double Six) = $^{1}/_{36}$. Thrown *n times*. **Binomial**. So $P(X\geq 1) = 1-P(X=0)$, where $P(X=r) = (^{n}{}_{r})p^{r}q^{n-r}$, so that $P(X=0) = (^{n}{}_{0})(^{1}/_{36})^{0}(^{35}/_{36})^{n} = (^{35}/_{36})^{n}$. Therefore, $P(X\geq 1) = 1-(^{35}/_{36})^{n}$. We *want* $1-(^{35}/_{36})^{n} > 0.5$, so that $½ > (^{35}/_{36})^{n}$; $\log(^{1}/_{2}) > n\log(^{35}/_{36})$; $^{\log(1/2)}/_{\log(35/36)} < n$; $24.605 < n$; $n > 24.605$, i.e. $n \geq 25$. So the **smallest** n is n = 25.

(**Q4**) A device fails if all its three essential components (**a, b and c)** fail. It is known from experience that, within a *specified period of time*, the probabilities of a, b and c failing are ½, ¼ and $^{1}/_{5}$ respectively, and that the *probabilities* of failure together of a and b, b and c, and c and a are $^{1}/_{9}$, 0 and $^{1}/_{15}$ respectively. Find the *probability* that within the period, (i) the device does not fail, and (ii) at least **one** component fails. **Note**: it is incorrect to assume that the event “b and c fail together” is the *impossible* event, ϕ.

A: $P(A) = ^{1}/_{2}$, $P(B) = ^{1}/_{4}$, $P(C) = ^{1}/_{5}$, $P(A\cap B) = ^{1}/_{9}$, $P(B\cap C) = 0$, and $P(A\cap C) = ^{1}/_{15}$. (i) P(Device *does not fail*) = 1 - P(device fails) = $1 - P(A\cap B\cap C)$. Now $A\cap B\cap C \subset B\cap C$, so that $P(A\cap B\cap C) \leq P(B\cap C)$. **But** we know that $P(B\cap C) = 0$.

Therefore, $P(A\cap B\cap C) \leq 0$. But, *any probability must be* $\geq 0$, so it follows that $P(A\cap B\cap C) = 0$. Conclusion: P(Device *does not fail*) = $1 - P(A\cap B\cap C) = 1$. (ii) At least *one* component fails: this is $P(A\cup B\cup C)$. Now $P(A\cup B\cup C) = P(A) + P(B) + P(C) - P(A\cap B) - P(A\cap C) - P(B\cap C) + P(A\cap B\cap C) = ½ + ¼ + ^{1}/_{5} - ^{1}/_{9} - 0 - ^{1}/_{15} + 0$; $P(A\cup B\cup C) = ^{139}/_{180}$.

(**Q5**) A *random* experiment consists of choosing an integer from the set of all positive integers in such a way that the probability of obtaining the integer n is **proportional** to $^{1}/_{n!}$. Find the *constant* of proportionality. A: $P(n) \propto ^{1}/_{n!}$; $P(n) = ^{k}/_{n!}$. **Now** $1 = P(1)+P(2)+P(3)+\ldots$, so that $^{1}/_{k} = \Sigma_{i=1}^{\infty}\, ^{1}/_{i!}$. **Now** $e^{x} = \Sigma_{i=0}^{\infty}\, x^{i}/i! = 1+x+^{x^2}/_{2!}+\ldots$, so that $e^{1} = \Sigma_{i=0}^{\infty}\, ^{1}/_{i!} = 1+\Sigma_{i=1}^{\infty}\, ^{1}/_{i!}$. Therefore, $^{1}/_{k} = e-1$; $k = ^{1}/_{e-1}$.

(**Q7**) Let $S = \{\omega \mid 0 \le \omega \le 1\}$, with the *probabilities of interval sets* equal to their length. If $A_n = \{\omega \mid 0 \le \omega \le 1-{}^1/_{3^n}\}$, and if $B_n = A_n^c \cap A_{2n}$ ($n = 1, 2, \ldots$), show that $P(\limsup_{n\to\infty} B_n) = 0$. A: $A_n = [0, 1-{}^1/_{3^n}]$. $n = 1 \Rightarrow [0, {}^2/_3]$. $n = 2 \Rightarrow [0, {}^8/_9]$. $n = 3 \Rightarrow [0, {}^{26}/_{27}]$. $A_n^c = [1-{}^1/_{3^n}, 1]$. $n = 1 \Rightarrow [{}^2/_3, 1]$. $n = 2 \Rightarrow [{}^8/_9, 1]$. $n = 3 \Rightarrow [{}^{26}/_{27}, 1]$. $A_{2n} = [0, 1-{}^1/_{3^{2n}}]$. $n = 1 \Rightarrow [0, {}^8/_9]$. $n = 2 \Rightarrow [0, {}^{80}/_{81}]$. $n = 3 \Rightarrow [0, {}^{728}/_{729}]$. *Intersection*: $n = 1 \Rightarrow [{}^2/_3, {}^8/_9]$. $n = 2 \Rightarrow [{}^8/_9, {}^{80}/_{81}]$. $n = 3 \Rightarrow [{}^{26}/_{27}, {}^{728}/_{729}]$. *Generally*, $[1-{}^1/_{3^n}, 1-{}^1/_{3^{2n}}] = B_n$. Now $\cup_{n=N}^{\infty} B_n = [1-{}^1/_{3^N}, 1] = C_N$. Further, $\limsup_{n\to\infty} B_n = \cap_{N=1}^{\infty} C_N = \{1\}$. So $P(\limsup_{n\to\infty} B_n)$ = the *length* of "interval" $\{1\} = 0$. **QED**.

# Assignment 1

Q: Find the *upper and lower limits* of the following sequence of events (on the sample space **R**): (i) $\{A_n\} = \{({}^1/_n-1, {}^1/_n)\}$, (ii) $\{A_n\} = \{(-{}^1/_n, 1+{}^1/_n)\}$, and (iii) $\{A_n\}$, *where* $\{A_n = [0, 1]$ for when n is **even**, and $\{A_n\} = \{0\}$ for when n is **odd**$\}$. A: (i) $n = 1 \Rightarrow (0, 1)$. $n = 2 \Rightarrow (-{}^1/_2, {}^1/_2)$. $n = 3 \Rightarrow (-{}^2/_3, {}^1/_3)$. Let $B_N = \cup_{n=N}^{\infty} A_n = (-1, {}^1/_N)$. Then $\limsup_{n\to\infty} A_n = \bar{A} = \cap_{N=1}^{\infty} B_N = \cap_{N=1}^{\infty}(-1, {}^1/_N) = (-1,0]$.

Now *let* $C_N = \cap_{n=N}^{\infty} A_n = ({}^1/_N-1, 0]$. Then $\liminf_{n\to\infty} A_n = \underline{A} = \cup_{N=1}^{\infty} C_N = (-1,0]$. Therefore, $\underline{A} = \bar{A} = (-1, 0]$, and so $\lim_{n\to\infty} A_n = (-1, 0]$. (ii) $n = 1 \Rightarrow (-1, 2)$. $n = 2 \Rightarrow (-{}^1/_2, {}^3/_2)$. $n = 3 \Rightarrow (-{}^1/_3, {}^4/_3)$. Let $B_N = \cup_{n=N}^{\infty} A_n = (-{}^1/_N, 1+{}^1/_N)$. Then $\limsup_{n\to\infty} A_n = \bar{A} = \cap_{N=1}^{\infty} B_N = [0, 1]$. Now *let* $C_N = \cap_{n=N}^{\infty} A_n = [0, 1]$. Then $\liminf_{n\to\infty} A_n = \underline{A} = \cup_{N=1}^{\infty} C_N = [0, 1]$. So $\underline{A} = \bar{A} = [0, 1]$, so that $\lim_{n\to\infty} A_n = [0, 1]$.

(iii) **Let** $B_N = \cup_{n=N}^{\infty} A_n = [0, 1]$. It follows that $\limsup_{n\to\infty} A_n = \bar{A} = \cap_{N=1}^{\infty} B_N = [0, 1]$. Now *let* $C_N = \cap_{n=N}^{\infty} A_n = \{0\}$. Then $\liminf_{n\to\infty} A_n = \underline{A} = \cup_{N=1}^{\infty} C_N = \{0\}$. Note *here* that $\underline{A} \neq \bar{A}$.

Q: State the *axioms* of a Probability Space, and show that these axioms are **satisfied** by a probability measure which is a *conditional* probability. A: A probability space is a triplet $(S, \mathcal{F}, P)$, where S is the *sample space*, $\mathcal{F}$ is the *Borel field of events* in S, and P is a *probability measure* on $\mathcal{F}$ (a **set** function). The three axioms of a probability space are as follows: (i) $P(A) \geq 0$ for every $A \in \mathcal{F}$; (ii) $P(S) = 1$; and (iii) if $\{A_n\}$ is a sequence of *pairwise exclusive events* in $\mathcal{F}$, then $P(\cup_n A_n) = \Sigma_n P(A_n)$.

We must show that the *above axioms* are satisfied by a probability measure which is a **conditional** probability. Let A and B be two events in $\mathcal{F}$ of a given probability space $(S, \mathcal{F}, P)$. The **conditional** probability of event A given event B, *denoted* by $P(A|B)$, is given by $P(A|B) = {}^{P(A\cap B)}/_{P(B)}$ if $P(B) > 0$; and is *undefined* if $P(B) = 0$.

When dealing with **conditional** probabilities, we are conditioning on a given event B (assuming that the experiment has *resulted in some outcome* in B). B, in effect, becomes our "new" sample space. To show that for a given event B (for which $P(B) > 0$), that $P(A|B)$ is a **probability function** having $\mathcal{F}$ as its domain, we must check the **three** axioms.

(i) $P(A|B) = {}^{P(A\cap B)}/_{P(B)} \geq 0$ for every $A \in \mathcal{F}$ because $P(B) > 0$ and $P(A\cap B) \geq 0$. (ii) $P(S|B) = {}^{P(S\cap B)}/_{P(B)} = {}^{P(B)}/_{P(B)} = 1$. (iii) If $A_1, A_2, \ldots$ is a *sequence of mutually exclusive events* in $\mathcal{F}$, and if $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$, then $P[\cup_{i=1}^{\infty} A_i|B] = P[(\cup_{i=1}^{\infty} A_i)\cap B]/P(B) = P[\cup_{i=1}^{\infty} (A_i\cap B)]/P(B) = [\Sigma_{i=1}^{\infty} P(A_i\cap B)]/P(B)$ (using the *fact* that $(S, \mathcal{F}, P)$ **originally** was a probability space; *and using* $\{A_n\}$ exclusive pairwise $\Rightarrow$ $\{A_n\cap B\}$ also *pairwise exclusive*) $= \Sigma_{i=1}^{\infty} P(A_i|B)$. Hence $P(A|B)$ for a *given* B (where $P(B) > 0$) **is** a probability function.

13th February 2001

## 1.2: An Axiomatic Approach to Probability

Let S be the *sample space* of some random experiment, and let $\mathcal{F}$ be a Borel field of events in S. Then a *probability measure* (P on $\mathcal{F}$) is defined as a **set** function which assigns to every event $A \in \mathcal{F}$, a *real* number P(A), called the **probability** of the event A, which satisfies the **following** axioms: (1) $P(A) \geq 0$ for every $A \in \mathcal{F}$; (2) $P(S) = 1$; and (3) if $\{A_n\}$ is a *sequence of pairwise exclusive* events in $\mathcal{F}$, then $P(\cup_n A_n) = \Sigma_n P(A_n)$.

The *triple* $(S, \mathcal{F}, P)$ is called the probability space. Its existence is assumed from now on. **Note**: If S is countable, then $\mathcal{F}$ is usually taken to be the *class of all subsets* of S, in which case every subset of S is an event to which a probability measure may be **assigned** according to the above axioms.

If S is **non-countable**, then $\mathcal{F}$ may be taken to be the *Borel field* generated by some simple class of events that are suitable (e.g. the class of intervals in the case of **R**). $\mathcal{F}$ cannot be taken to be the class of all subsets of S, since *certain of these subsets* are non-measurable — or we cannot assign to them a probability measure consistent to the above axioms.

**Theorem 1.5:** For every event A, $P(A') = 1-P(A)$. *Proof:* $A\cup A' = S$, and $A\cap A' = \phi$. So $P(A\cup A') = (3) = P(A)+P(A')$. **But** $P(A\cup A') =$ (*because* $A\cup A' = S$) $= P(S) = 1$. So $P(A)+P(A') = 1 \Leftrightarrow P(A') = 1-P(A)$. **Theorem 1.6:** $P(\phi) = 0$. *Proof:* $A\cap\phi = \phi$, $A\cup\phi = A$. So $P(A\cup\phi) = P(A)+P(\phi)$; $P(A) = P(A)+P(\phi)$; $P(\phi) = 0$. **Note**: The *converse* is not true, i.e. $P(A) = 0 \not\Rightarrow A = \phi$.

*Proof:* Let $A = \{c\}$, where $c \in (a,b]$. Now $P(a<X\leq b) = \int_a^b f(x)dx$, where f(x) is the *density function* of X, a continuous random variable. It follows that $P(X = c) = \int_c^c f(x)dx = 0$. But $\{C\} \neq \phi$. *Generally*, if A is a **countable** subset of a **non-countable** set **R**, then $P(x \in A) = \int_A f(x)dx = 0$. *Similarly*, $P(A) = 1$ does not imply that $A = S$. Take $S = \mathbf{R}$, then $A = \mathbf{R}\backslash\{a\} \Rightarrow P(A) = 1$.

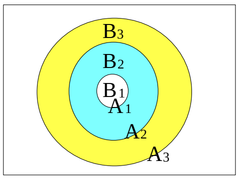
**Theorem 1.7:** For any *event* A, we have $0 \leq P(A) \leq 1$. *Proof:* Axiom (1): $P(A) \geq 0$. Theorem 1.5 $\Rightarrow P(A) = 1-P(A') \leq 1$, i.e. by $P(A') \geq 0$, $0 \leq P(A) \leq 1$. **Theorem 1.8:** For any *two events* A and B, we have $P(A\backslash B) = P(A)-P(A\cap B)$. *Proof:* $(A\cap B)\cap(A\backslash B) = \phi$; $(A\cap B)\cup(A\backslash B) = A \Rightarrow$ (by (3)) $P(A) = P(A\cap B)+P(A\backslash B) \Rightarrow P(A\backslash B) = P(A) - P(A\cap B)$.

**Theorem 1.9:** If $B \subset A$, then (i) $P(A\backslash B) = P(A)-P(B)$, and (ii) $P(B) \le P(A)$. *Proof:* $B \subset A \Rightarrow A \cap B = B$. **Theorem 1.8** $\Rightarrow P(A\backslash B) = P(A)-P(B)$. (ii) *follows* from (i), as $P(A\backslash B) \geq 0$. **Theorem 1.10:** For any *two events* A and B, $P(A\cup B) = P(A)+P(B)-P(A\cap B)$. *Proof:* $B\cap(A\backslash B) = \phi$; $B\cup(A\backslash B) = A\cup B$. Now Axiom 3 $\Rightarrow P(A\cup B) = P(B)+P(A\backslash B)$. *But* $P(A\backslash B) =$ (Theorem 1.8) $= P(A)-P(A\cap B)$, so that $P(A\cup B) = P(B)+P(A)-P(A\cap B)$.

**Corollary:** $P(A\cup B) \le P(A)+P(B)$ [$P(A\cap B) \geq 0$]. Also, for a *finite sequence of events* $A_1$, $A_2$, ..., $A_n$, $P(\cup_{i=1}^{n} A_i) \le \Sigma_{i=1}^{n} P(A_i)$ (proof *by induction*). **Theorem 1.11:** If $\{A_n\}$ is an *infinite* sequence of events, then $P(\cup_{n=1}^{\infty} A_n) \le \Sigma_{n=1}^{\infty} P(A_n)$. *Proof:* Put $B_1 = A_1$, and put $B_n = A_n\backslash\cup_{i=1}^{n-1} A_i$ (n = 2, 3, ...) — then the $\{B_n\}$ is a *sequence of pairwise exclusive* events, and $\cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n$; $B_n \subset A_n \ \forall\ n$. **Hence** $P(\cup_{n=1}^{\infty} A_n) = P(\cup_{n=1}^{\infty} B_n) =$ (*axiom 3*) $= \Sigma_{n=1}^{\infty} P(B_n) \le \Sigma_{n=1}^{\infty} P(A_n)$ (from Theorem 1.9).

15th February 2001

**Theorem 1.12:** If $\{A_n\}$ is a *monotonic sequence of events*, then $\lim_{n\to\propto} P(A_n) = P(\lim_{n\to\infty} A_n)$ (= $P(\cup_{n=1}^{\infty} A_n)$ if $\uparrow$, and = $P(\cap_{n=1}^{\infty} A_n)$ if $\downarrow$). **Proof:** Let $A_n \uparrow$ Let $B_n = A_n\backslash A_{n-1}$, with $B_1 = A_1$. Then $\{B_n\}$ is a *pairwise exclusive sequence* with $A_n = \cup_{i=1}^{n} B_i$. *Hence* $\lim_{n\to\infty} A_n =$ (Theorem 1.3) $= \cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty}(\cup_{n=1}^{N} B_i) = \cup_{i=1}^{\infty} B_i$. *Hence* $P(\lim_{n\to\infty} A_n) = P(\cup_{i=1}^{\infty} B_i) =$ (*axiom 3*) $= \Sigma_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \Sigma_{i=1}^{n} P(B_i) =$ (*axiom 3*) $= \lim_{n\to\infty} P(\cup_{i=1}^{n} B_i) = \lim_{n\to\infty} P(A_n)$. (ii) Let $A_n \downarrow \Rightarrow A_n' \uparrow$. So (i) $\Rightarrow P(\lim_{n\to\infty} A_n') = \lim_{n\to\infty} P(A_n') =$ (Theorem 1.5) $= 1-\lim_{n\to\infty} P(A_n)$. But $\lim_{n\to\infty} A_n' =$ (Theorem 1.3) $= \cup_{n=1}^{\infty} A_n'$. So $P(\lim_{n\to\infty} A_n') = P(\cup_{n=1}^{\infty} A_n') =$ (de Morgan) $= P(\cap_{n=1}^{\infty} A_n)' =$ (Theorem 1.5) $= 1-P(\cap_{n=1}^{\infty} A_n) =$ (Theorem 1.3) $= 1-P(\lim_{n\to\infty} A_n)$ (N.B. $A_n \downarrow$). So $1-\lim_{n\to\infty} P(A_n) = 1-P(\lim_{n\to\infty} A_n) \Rightarrow \lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n)$, which *completes the proof*.



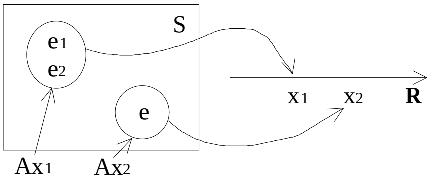
## The Borel-Cantelli Lemmas (Heathcote, Page 219)

Let $\{A_n\}$, for n = 1, 2, 3,...., be a *sequence of events.* (i) If $\Sigma_{n=1}^{\infty} P(A_n) < \infty$, then with **probability 1,** only a finite number of events $A_n$ occur ($P(\limsup_{n\to\infty} A_n) = 0$). (ii) If the $A_n$'s are *independent*, and if $\Sigma_{n=1}^{\infty} P(A_n) = \infty$, then with probability 1, *infinitely* many of the $A_n$'s occur. ($P(\limsup_{n\to\infty} A_n) = 1$). **Proof:** Let A be the event that *infinitely many of the* $A_n$ occur, then $A = \cap_{N=1}^{\infty}\{\cup_{n=N}^{\infty} A_n\} = \limsup_{n\to\infty} A_n = \bar{A}$. So, for *every N*, we **have** $A \subset \cup_{n=N}^{\infty} A_n$.

Therefore, $P(A) \le$ (by Theorem 1.9) $P(\cup_{n=N}^{\infty} A_n) \le \Sigma_{n=N_\varepsilon}^{\infty} P(A_n) < \varepsilon$, where $\varepsilon > 0$ for *sufficiently large* $N_\varepsilon$, since $\Sigma_{n=1}^{\infty} P(A_n) < \infty$. **Hence** $P(A) = 0$, i.e. *the probability of infinitely many events* $A_n$ occurring = 0 $\Rightarrow$ the probability of *finitely* many $A_n$'s occurring is 1. (ii) Let A' be the **complement** of A defined above. Hence A' occurs iff at most a *finite number of the* $A_n$ occur, i.e. iff all but a **finite** number of the $A_n$' occur. Hence $A' = \cup_{N=1}^{\infty}\{\cap_{n=N}^{\infty} A_n'\}$ (**Note:** $\liminf_{n\to\infty} A_n'$).

So $P(A') = P(\cup_{N=1}^{\infty}\{\cap_{n=N}^{\infty} A_n'\}) \le$ (by Theorem 1.11) $\Sigma_{N=1}^{\infty} P(\cap_{n=N}^{\infty} A_n') = \Sigma_{N=1}^{\infty} \Pi_{n=N}^{\infty}(1-P(A_n))$ (because the *independence* of the $A_n$'s $\Rightarrow$ *the independence* of the $A_n$''s. Note also that the red bit is $P(A_n')$). Now consider $\Pi_{n=N}^{M}(1-P(A_n))$, with $M > N$. This is $\le \exp(-\Sigma_{n=N}^{M} P(A_n))$ (which *follows* from $1-x \le e^{-x}$, for $0 \le x \le 1$). So for each *fixed* N, let $M \to \infty$, and so $\Pi_{n=N}^{\infty}(1-P(A_n)) \le \exp[-\Sigma_{n=N}^{M} P(A_n)] = 0$, since $\Sigma_{n=N}^{\infty} P(A_n) = \infty$. So $P(A') = 0 = 1-P(A) \Rightarrow P(A) = 1$. Hence we obtain the *result*.

# Section 2: Random Variables and Probability Distributions



Let (S, $\mathcal{F}$ , P) be a *probability space*, then a real r-dimensional r.v. X is a **real** valued function whose domain is S, whose co-domain is **R**, and which assigns to *each element* e of S a real number X(e) = x, such that the set $A_x = \{e \in S \mid x(e) = x\} \in \mathcal{F}$ , i.e. $A_x$ is an event (where X is a *measurable function* w.r.t. $\mathcal{F}$ ). Define the **Range** of X, $R_X$, by $R_X = \{x \in \mathbf{R} \mid x = X(e)$; for all $e \in S\}$. r.v.'s will be denoted by X, Y, ... (upper case letters); and their *values* by x, y, ... (lower case letters)

Write $(X \le x)$ for the *event* $\{e \in S \mid X(e) \le x\}$, and write $(X \in \mathbf{R})$ for the event $\{e \in S \mid X(e) \in \mathbf{R}\}$. **Note**: the 1-dim r.v. X induces *from the basic probability space* (S, $\mathcal{F}$, P) a **probability** space (**R**, β, P), where **R** is the real line; β is the class of borel sets in **R**; and P is a *probability measure*. **Example**: release 2 rats through a T maze. Here, S = {(RR), (RL), (LR), (LL)}. Let X be the number of rats *turning right minus the number turning left.* X is a r.v., with values X(RR) = 2, X(RL) = 0 = X(LR), and X(LL) = -2. **Domain:** $D_X = S$; **Range:** $R_X = \{-2, 0, 2\}$.

## 2.2: (Cumulative) Distribution Function (cdf)

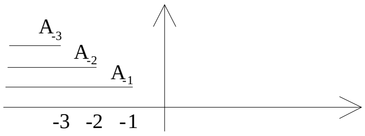
Let X be a 1-dimensional r.v. A **function** F defined for *every real x* by $F(x) = P(X \le x)$ is called the cdf of X. **Clearly**, F(x) is a non-negative and *real valued function* of x. **Theorem 2.1:** Let F be the cdf of a r.v. X. Then, *for any reals* a and b, where a < b, we have $P(a<X\le b) = F(b)-F(a)$. **Proof:** Events $(X\le a)$ and $(a<X\le b)$ are *exclusive*, and $(X\le a)\cup(a<X\le b) = (X\le b)$.

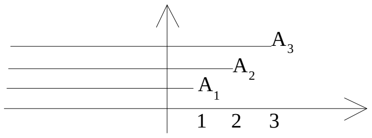
So $P((X\le a)\cup(a<X\le b))$ = (Axiom 3) = $P((X\le a))+P((a<X\le b)) = P(X\le b)$. So $P((a<X\le b)) = F(b)-F(a)$. **Exercise**: show that (i) $P(a\le X\le b) = F(b)-F(a)+P(X=a)$, (ii) $P(a\le X<b) = F(b)-F(a)+P(X=a)-P(X=b)$, and (iii) $P(a<X<b) = F(b)-F(a)-P(X=b)$. *Proof* of (i): Events (X=a) and $(a<X\le b)$ are *exclusive*, since $(X=a)\cup(a<X\le b) = (a\le X\le b)$. So $P((X=a)\cup(a<X\le b))$ = (*Axiom 3*) = $P(X=a)+P(a<X\le b)$. So *Theorem 2.1* ⇒ $P(a\le X\le b) = F(b)-F(a)+P(X=a)$. *(ii) and (iii) are proved in a similar manner.*

**Theorem 2.2** (properties of cdf): If F is the cdf of X, then (i) $F(x_2) \ge F(x_1)$ if $x_2 > x_1$ (F is *non decreasing*); (ii) $F(-\infty) = 0$ ($\lim_{x\to-\infty} F(x) = 0$); (iii) $F(\infty) = 1$ ($\lim_{x\to\infty} F(x) = 1$); and (iv) $\lim_{h\to0+} F(x+h) = F(x+) = F(x)$ for all x (F is *continuous* on the right). **Conversely**, any function F with these properties is a cdf of a r.v. X, i.e. F *uniquely* determines a probability space (**R**, β, P) such that $P(X\le x) = F(x)$ for all $x \in \mathbf{R}$.

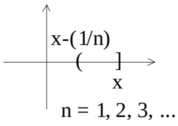
20th February 2001

**Proofs**. (i) $(X \le x_2) \supset (X \le x_1)$ ⇒ (Theorem 1.9 of Section 1) $P(X \le x_2) \ge P(X \le x_1)$ ⇒ $F(x_2) \ge F(x_1)$. (ii) Let $A_r = (X\le r)$, then the *sequence* $A_{-1}, A_{-2}, A_{-3}, \ldots$ is ↓ $(A_{-1} \supset A_{-2} \supset A_{-3} \supset \ldots)$, and so Theorem 1.3 of Section 1 ⇒ $\lim_{r\to\infty} (A_{-r}) = \cap_{r=1}^{\infty} A_{-r} = \phi$. Hence $F(-\infty) = \lim_{r\to\infty} F(-r) = \lim_{r\to\infty} P(A_{-r})$ = (Theorem 1.12 of Section 1) = $P(\lim_{r\to\infty} A_{-r}) = P(\phi) = 0$.





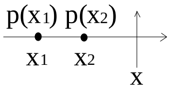
(iii) *Similarly*, the sequence $A_1, A_2, \ldots$ is $\uparrow$ $(A_1 \subset A_2 \subset A_3 \ldots)$ So $\lim_{r\to\infty} (A_r)$ = (Theorem 1.3) = $\cup_{r=1}^{\infty} A_r = \mathbf{R}$. Hence $F(\infty) = \lim_{r\to\infty} (F(r)) = \lim_{r\to\infty}(P(A_r))$ = (Theorem 1.12 of Section 1) = $P(\lim_{r\to\infty} A_r) = P(\mathbf{R}) = 1$. (iv) Let $h > 0$, then $F(x+h)-F(x)$ = (Theorem 2.1) = $P(x<X\le(x+h))$ (so $x \in (x,x+h]$). **But** $(x<X\le x+h) \to \phi$ as $h \to 0+$, so that $\lim_{h\to0+}(F(x+h)-F(x)) = \lim_{h\to0+}P(x<X\le x+h)$ = (Theorem 1.12) = $P(\lim_{h\to0+}(x<X\le x+h)) = P(\phi) = 0$. Therefore, $\lim_{h\to0+} F(x+h) = F(x)$.



The proof of the **converse** of 2.2. is omitted! **Example:** If F, the cdf of X, has a *discontinuity* or *jump* of size p $(0 < p < 1)$ at a point x, i.e. if $F(x)-F(x-) = p$, then $p = P(X=x)$. **Proof:** Theorem 2.1 $\Rightarrow$ $P(x-{}^1/_n < X \le x) = F(x)-F(x-{}^1/_n)$. As $n \to \infty$, $(x-{}^1/_n < X \le x) \to (X=x)$, and so $F(x-{}^1/_n) \to F(x-)$. **Hence** $P(X=x) = F(x)-F(x-) = p$.

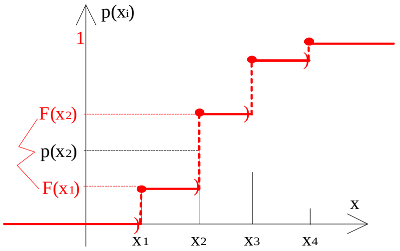
## 2.3: Types of Random Variables

### 2.3.1: Discrete Random Variables



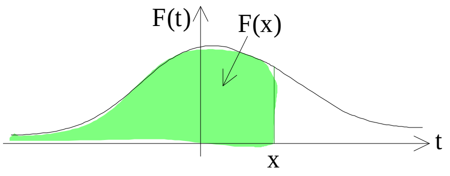
X is called a *discrete random variable* if its range is a **countable** set of reals $\{x_1, x_2, x_3, \ldots, x_n, \ldots\}$, and *such that* $P(X=x_i) = p(x_i) \geq 0$ for all i, and such that $\Sigma_i\, p(x_i) = 1$. Therefore, p = 0 for all **other** points on the real line. (Think of a *unit mass* distributed among the $x_i$'s). $p(x_i)$ is called a probability (*mass*) function of X. The cdf F of X is given by $F(x) = P(X\le x) = \Sigma_{x_i \le x}\, p(x_i)$, and is called a **discrete** cdf.

F is a *step function*, with jumps of sizes $P(X=x_i) = p(x_i)$ at $x_i$, and remains **constant** in the open interval $(x_i, x_{i+1})$. Further, $p(x_i) = F(x_i)-F(x_i-)$. See the printout for a *list of important discrete distributions.*

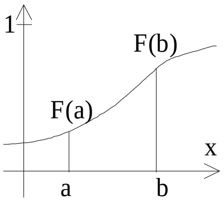


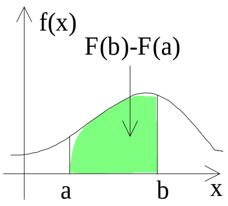
22nd February 2001

### 2.3.2: Continuous Random Variables



A r.v. X is called *continuous* if $\exists$ a function $f(x) \geq 0$, and such that for *every* real x, the c.d.f. F of X is given by $F(x) = P(X\le x) = \int_{-\infty}^{x} f(t)dt$ (---(a)). f is called the *probability density function* (pdf) of the r.v. X. F(x) is a **continuous** function of x. Also, ${}^{dF}/_{dx}$ *exists*, and ${}^{dF}/_{dx} = f(x)$ at all *continuity* points of f(x).





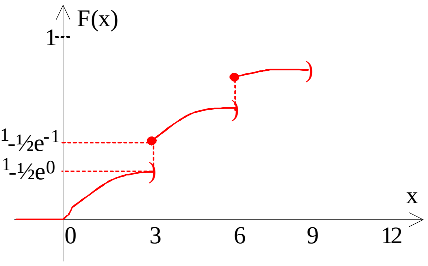
(a) $\Rightarrow$ f must *satisfy* $\int_{-\infty}^{\infty} f(x)dx = 1$ $(F(\infty) = 1)$. A *function* f(x) is a p.d.f. of some **random** variable X iff $f(x) \geq 0$, and iff $\int_{-\infty}^{\infty} f(x) = 1$. Also, *for any 2 reals* a and b, where $a < b$, we have $P(a<X\le b)$ = (Theorem 2.1) = $F(b)-F(a) = \int_a^b f(x)dx$. In *particular*, the probability that X lies in the range $(x, x+dx]$ is given by $P(x<X\le x+dx) = \int_x^{x+dx} f(t)dt \approx f(x)dx = dF(x)$. In **general**, for any Borel set B in **R**, we have $P(X \in B) = \int_B f(x)dx$.

**Note 1:** For continuous X, $P(X=x_0) = \int_{x_0}^{x_0} f(x)dx = 0$ (true *also* for a countable set A: $P(x \in A) = \int_A f(x)dx = 0$), i.e. an *event* with probability zero is not necessarily the impossible event, but is highly **unlikely** to occur. Similarly, an event with probability 1 is not necessarily S, since $P(S\backslash\{x_0\}) = 1$. Also, $P(a<X\le b) = P(a\le X\le b) = P(a\le X<b) = P(a<X<b) = \int_a^b f(x)dx$.

**Note 2:** *Decomposition Theorem*. It can be shown that any cdf F can be decomposed as $F = c_1F_1+c_2F_2+c_3F_3$, where $c_i \geq 0$, and $\Sigma_i c_i = 1$. $F_1$ is *discrete*, $F_2$ is *continuous*, and $F_3$ is called a *continuous singular* cdf ($F_3$ is continuous everywhere, and its **derivative** vanishes almost everywhere). $F_3$ rarely occurs in *practice*, so that $c_3 = 0$, and $F = c_1F_1+c_2F_2$. If $c_2 = 0$, then $F = F_1$, **discrete**. If $c_1 = 0$, then $F = F_2$, **continuous**. But if $c_1 \neq 0$, and if $c_2 \neq 0$, then F is called a *mixed* cdf.

**Note 3:** The distribution of a *continuous r.v.* is called a continuous distribution, and it may be specified by a pdf and giving the range of the r.v. **Example**: Any function f(x) such that $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$ is a p.d.f. of some *continuous random variable* X, i.e. f(x) determines uniquely a **continuous** cdf $F(x) = \int_{-\infty}^{x} f(t)dt$.

**Proof**. $F(\infty)$ = (Theorem 2.2) = $\int_{-\infty}^{\infty} f(x)dx = 1$; $F(-\infty) = 0$; F(x) *continuous* ⇒ continuous on right; and for $a < b$, $F(b)-F(a) = \int_a^b f(x)dx \geq 0$ *since* $f(x) \geq 0$. **Example 2**: The duration in *minutes* of a long distance telephone call from a certain town is a r.v. X with cdf $F(x) = 1-{}^1/_2e^{-x/3}-{}^1/_2e^{-[x/3]}$ for $x \geq 0$, and $F(x) = 0$ *for* $x < 0$. ([x] = the integer part of x).



Find the **percentage** number of calls each having *duration* of: (i) 3 minutes, (ii) less than 3 minutes, (iii) between 3 and 6 minutes inclusive, and (iv) at least 3 minutes, given that it is less than 8 minutes. (i) $P(X=x) = p(x) = F(x)-F(x-)$. **So** $p(3) = F(3)-F(3-) = (1-\frac{1}{2}e^{-1}-\frac{1}{2}e^{-1}) - (1-\frac{1}{2}e^{-1}-\frac{1}{2}e^{0}) = 1-e^{-1}-\frac{1}{2}(1-e^{-1}) = \frac{1}{2}(1-e^{-1}) = 0.32$. (ii) $P(X<3) = P(X\le 3)-P(X=3) = F(3)-p(3) = (1-\frac{1}{2}e^{-1}-\frac{1}{2}e^{-1}) - \frac{1}{2}(1-e^{-1}) = \frac{1}{2}(1-e^{-1}) = 0.32$. (iii) $P(3\le X\le 6) = F(6)-F(3)+p(3) = \ldots = 0.54$. (iv) $P(X\geq 3 \mid X<8) = {}^{P(3\le X<8)}/_{P(X<8)} = {}^{F(8)-F(3)+p(3)-p(8)}/_{F(8)-p(8)} = \ldots = 0.64$ (taking *note* that $p(8) = 0$ as F is **continuous** at $x = 8$).

## Tutorial

Using the **same** F(x) as above, find the probability that a call has **duration** of (i) 3 mins, (ii) *less* than 4 mins, (iii) *more* than 5 mins, (iv) *more* than 5 mins *given* that it is *less* than 9 mins, and (v) *less* than 9 mins *given* that it is *more* than 5 mins. A: (i) done above. (ii) $P(X<4) = P(X\le 4)-P(X=4) = F(4)-p(4) = F(4)-0$ (*no jump*). So $P(X<4) = 1-\frac{1}{2}e^{-4/3}-\frac{1}{2}e^{-1} = 0.68$.

(iii) $P(X>5) = 1-P(X\le 5) = 1-F(5) = 1-(1-\frac{1}{2}e^{-5/3}-{}^1/_{2e}) = 0.278$. (iv) $P(X>5 \mid X<9) = {}^{P(5<X<9)}/_{P(X<9)} = {}^{F(9)-F(5)-p(9)}/_{F(9)-p(9)}$. Now $p(9) = F(9)-F(9-)$, so we *have* ${}^{-F(5)+F(9-)}/_{F(9-)}$. Now $F(5) = 0.722$, and $F(9-) = 1-\frac{1}{2}e^{-3}-\frac{1}{2}e^{-2} = 0.907$, so that $P(X>5 \mid X<9) = {}^{-0.722+0.907}/_{0.907} = 0.203$. (v) $P(X<9 \mid X>5) = {}^{P(5<X<9)}/_{P(X>5)} = {}^{0.185}/_{0.278}$ = (using *previous values*) = 0.665.

| | Red | White | Blue |
|---|---|---|---|
| A | 2 | 3 | 5 |
| B | 4 | 1 | 3 |
| C | 3 | 4 | 3 |

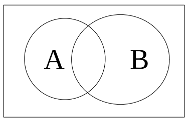
(**Q8**) *Coloured balls* are distributed in three boxes as shown in the table. A box is selected at random, and a ball is **randomly selected** and found to be red. What is the probability that box *C* was chosen? A: P(A) = P(B) = P(C) = $^{1}/_{3}$. P(R|A) = $^{2}/_{10}$. P(R|B) = ½. P(R|C) = $^{3}/_{10}$. P(C|R) = $^{P(R|C)P(C)}/_{P(R|A)P(A)+P(R|B)P(B)+P(R|C)P(C)}$ = $[^{3}/_{10}\,^{1}/_{3}]/[(^{2}/_{10}+^{1}/_{2}+^{3}/_{10})\,^{1}/_{3}] = [^{3}/_{10}]/1 = ^{3}/_{10}$.

(**Q9**) In Question 8, three balls are selected at random, and it is observed that they are *one of each colour*. What is the probability that box C was chosen? Consider both cases in which (i) balls are **not** replaced, and (ii) balls **are** replaced. A: P(C|E) = $^{P(E|C)P(C)}/_{P(E|A)P(A)\,+\,P(E|B)P(B)\,+\,P(E|C)P(C)}$. Draw a *tree diagram* to work out the values for P(E|?), for both (i) $^{64}/_{159}$, and (ii) $^{7}/_{72}$.

(**Q11**) A card is drawn at random from a pack. If A is the event that a card is an **ace,** and B is the event that the card is a **spade**, show that A and B are *independent*. A: P(Ace) = $^{4}/_{52}$. P(Spade) = ¼. *Independent* if P(A∩B) = P(A)P(B). Now P(Ace and Spade) = $^{1}/_{52} = {}^{4}/_{52}\times{}^{1}/_{4}$, so they are independent.

(**Q12**) A ticket is drawn at random from a bag containing four tickets, numbered 000, 011, 101 and 110. If $A_i$ is the event that on a ticket drawn, 0 appears as the i*th* digit from the right, show that the $A_i$'s, though **pairwise independent**, are **not** independent. A: $A_1$ = ½ = $A_2$ = $A_3$. P($A_1$ and $A_2$) = ¼ = P($A_1$ and $A_3$) = P($A_2$ and $A_3$) so *pairwise independent.* P(($A_1$ and $A_2$) and $A_3$) = ¼ ≠ P($A_1$ and $A_2$)P($A_3$) = ¼×½ = $^{1}/_{8}$. So they are not independent.

(**Q13**) If two independent events are *exclusive,* show that at least one of them must have probability zero. A: *Independent* ⇒ P(A∩B) = P(A)P(B). *Exclusive* ⇒ P(A∪B) = P(A)+P(B), i.e. P(A∩B) = 0, **so that** P(A) = 0 or P(B) = 0 as *required*.



(**Q14**) If two events (A are B) are independent, show that A and $B^c$ are **also** independent. A: We *know* that P(A∩B) = P(A)P(B) ⇒ P(A∪B) = P(A)+P(B)-P(A)P(B); P(A∪B)-P(B) = P(A)-P(A)P(B); P(A∪B)-P(B) = P(A)(1-P(B)); P(A∪B)-P(B) = P(A)P($B^c$). But *what is* P(A∪B)-P(B)? This is P(A∩$B^c$) by a Venn diagram. Therefore, P(A∩$B^c$) = P(A)P($B^c$) ⇒ A and $B^c$ are *independent* as **required**.

(**Q15**) If $A_1$, $A_2$ and $A_3$ are *independent*, such that P($A_3$) > 0, show that P[$A_1$∪$A_2$ | $A_3$] = P[$A_1$∪$A_2$]. A: If they are *independent,* then $A_1$∪$A_2$ will be **independent** of $A_3$ (we know that $A_1$ is independent of $A_3$, and that $A_2$ is independent of $A_3$). **Therefore,** P($A_1$∪$A_2$ | $A_3$) = P(($A_1$∪$A_2$)∩$A_3$)/P($A_3$) = P($A_1$∪$A_2$)P($A_3$)/P($A_3$) = P($A_1$∪$A_2$). **QED**.

(**Q16**) The number X of *aces in a hand* of 13 cards is a random variable with range {0, 1, 2, 3, 4}, and domain S having $^{53}C_{13}$ *sample* points. Find the probability distribution of X, if all hands of 13 cards are **equally** likely. A: X = 0: $^{49}C_{13}\,^{4}C_{0}$. X = 1: $^{49}C_{12}\,^{4}C_{1}$. X = 2: $^{49}C_{11}\,^{4}C_{2}$. X = 3: $^{49}C_{10}\,^{4}C_{3}$. X = 4: $^{49}C_{9}\,^{4}C_{4}$.

(**Q17**) A *random variable* has pf p(x) = cx for x = 1, 2, 3, 4. Find the **constant** c, and calculate P(X>1). A: p(1) = c, p(2) = 2c, p(3) = 3c, p(4) = 4c; so *summing,* 10c = 1; c = $^{1}/_{10}$.

(**Q18**) Look up the pf's of the following *common discrete probability distributions:* (i) Binomial, (ii) Poisson, (iii) Negative Binomial, (special case Geometric), and (iv) Hypergeometric. **Verify** in each case that $p(x) = P(X=x)$ is indeed a pf, i.e. that they are *non-negative and sum to unity.*

(i) **Binomial**: ${}^nC_x p^x(1-p)^{n-x}$. ${}^nC_x$ is *always* $\geq 0$. As $x = 0, 1, ..., n$, and $0 < p < 1$, $n-x \geq 0$, and **so** $p^x(1-p)^{n-x} \geq 0$. So *non-negative*. Now $\Sigma_{i=0}^{n}\ {}^nC_i p^i(1-p)^{n-i} = (p+(1-p))^n = 1^n = 1$ (*using* $(a+b)^n = \Sigma_{i=0}^{n}\ ({}^n{}_i)a^ib^{n-i}$). (ii) **Poisson**: $\exp(-\mu)\mu^x/x!$. As $\mu > 0$, *and as* $x = 0, 1, ...$, then this will *always* be $\geq 0$. Now $\Sigma_{i=0}^{\infty} \exp(-\mu)(\mu^i/i!) = \exp(-\mu)\Sigma_{i=0}^{\infty}(\mu^i/i!) = e^{-\mu}(e^{\mu}) = 1$. **QED**.

(iii) **Negative Binomial**: $({}^{x-1}{}_{k-1})p^k(1-p)^{x-k}$, *where* $x = k, k+1, ...$ $(0 < p < 1)$. The **book** gives an *alternative* formula: $({}^{r+x-1}{}_{x})p^rq^x = ({}^{-r}{}_{x})p^r(-q)^x$, for $x = 0, 1, 2,...$ Now *we know that* $(1+z)^{\alpha} = \Sigma_{k=0}^{\infty}({}^{\alpha}{}_{k})z^k$, so that $(1-q)^{-r} = \Sigma_{x=0}^{\infty}({}^{-r}{}_{x})(-q)^x$; $(1-q)^{-r}p^r = \Sigma_{x=0}^{\infty}\ ({}^{-r}{}_{x})(-q)^xp^r = p^{-r}p^r = 1$. **QED**. (iv) **Geometric**: $p(1-p)^{x-1}$, for $x = 1, 2, ....$ *This is always* $\geq 0$ as $0 < p < 1$. Now $\Sigma_{x=1}^{\infty}\ p(1-p)^{x-1} = p\Sigma_{x=1}^{\infty}\ (1-p)^{x-1} = p\Sigma_{x=0}^{\infty}\ (1-p)^x$. *Now* $\Sigma_{x=0}^{\infty}\ a^x \to {}^1/{}_{1-a}$ as $x \to \infty$, so we *have* $= p({}^1/{}_{1-(1-p)}) = {}^p/{}_p = 1$. **QED**.

Or, we use **Maclaurin Series:** $f(z) = \Sigma_{j=0}^{\infty}\ f^{(j)}(0)z^j/j!$ (---(1)). Let $f(z) = (1-z)^{-k}$ (---(2)). So $f^{(0)}(z) = f(z) = (1-z)^{-k}$; $f^{(0)}(0) = 1$. $f^{(1)}(z) = (-k)(1-z)^{-k-1}(-1)$; $f^{(1)}(0) = k = {}^{k!}/{}_{(k-1)!}$. $f^{(2)}(z) = k(-k-1)(1-z)^{-k-2}(-1)$; $f^{(2)}(0) = k(k+1) = {}^{(k+1)!}/{}_{(k-1)!}$. $f^{(3)}(z) = k(k+1)(-k-2)(1-z)^{-k-3}(-1)$; $f^{(3)}(0) = k(k+1)(k+2) = {}^{(k+2)!}/{}_{(k-1)!}$. It follows that $f^{(j)}(z) = k(k+1)...(k+j-2)(-k-j+1)(1-z)^{-k-j}(-1)$; $f^{(j)}(0) = k(k+1)...(k+j-1) = {}^{(k+j-1)!}/{}_{(k-1)!}$ (---(3)).

(1) to (3) $\Rightarrow (1-z)^{-k} = \Sigma_{j=0}^{\infty}\ {}^{(k+j-1)!}/{}_{(k-1)!}\ z^j/j! = \Sigma_{j=0}^{\infty}\ ({}^{k+j-1}{}_{k-1})z^j = (j = x-k) = \Sigma_{x=k}^{\infty}\ ({}^{x-1}{}_{k-1})z^{x-k}$ (---(4)). **Negative Binomial**: $\Sigma_{x=k}^{\infty}\ ({}^{x-1}{}_{k-1})p^k(1-p)^{x-k} = p^k\Sigma_{x=k}^{\infty}\ ({}^{x-1}{}_{k-1})(1-p)^{x-k} = ((4), 1-p = z) = p^k(1-(1-p))^{-k} = p^kp^{-k} = 1$ *as required*. **QED**.

(iv) **Hyper-Geometric**: $({}^k{}_x)({}^{N-k}{}_{n-k})/({}^N{}_n)$, for $x = 0, 1, ..., n$. This is *clearly* $\geq 0$ as all **combinations** are $\geq 0$. Now *as* $\Sigma_{i=0}^{m}\ ({}^a{}_i)({}^b{}_{m-i}) = ({}^{a+b}{}_m)$, then $\Sigma_{i=0}^{n}\ ({}^k{}_i)({}^{N-k}/{}_{n-i})/({}^N{}_n) = [1/({}^N{}_n)]\Sigma_{i=0}^{n}\ ({}^k{}_i)({}^{N-k}{}_{n-i}) = [1/({}^N{}_n)]({}^{k+(N-k)}{}_n) = ({}^N{}_n)/({}^N{}_n) = 1$. **QED**.

27th February 2001

## 2.4: The Riemann-Stieltjes Integral

**Definition** (*functions of bounded variation):* Let F be a function *defined on a generalised closed interval* [a,b] (i.e. either a or b could be infinite). Let $\Delta$ be a partition of [a,b], i.e. $\Delta$: $a = x_0 < x_1 < x_2 < ... < x_n = b$. F is said to be a function of bounded variation over [a,b] if $\Sigma_{k=1}^{n} |F(x_k)-F(x_{k-1})|$ is *bounded for all partitions* $\Delta$. Then ${}^{\sup}{}_{\forall\Delta \text{ on } [a,b]}\ \Sigma_{k=1}^{n} |F(x_k)-F(x_{k-1})|$ is called the **total** variation of F over [a,b].

**Example**: Any cdf $F(x) = P(X \leq x)$ *is of bounded variation*, with total variation 1. **Proof**: From Theorem 2.2, F(x) is a *monotonically increasing function*, which is bounded above by 1. **Hence**, for any [a,b], and for any partition $\Delta$, $\Sigma_{k=1}^{n} |F(x_k)-F(x_{k-1})| = \Sigma_{k=1}^{n} F(x_k)-F(x_{k-1}) = [F(x_1)-F(x_0)] + [F(x_2)-F(x_1)] + ... + [F(x_n)-F(x_{n-1})] = -F(x_0) + F(x_n) = F(b) - F(a)$. It follows that ${}^{\sup}{}_{\forall\Delta \text{ on } [a,b]}\ \Sigma_{k=1}^{n} |F(x_k)-F(x_{k-1})| = F(\infty)-F(-\infty) = 1-0 = 1$.

**Note**: A function of *bounded variation* has at most a countable number of discontinuities, and these are all **jumps** (e.g. the cdf of a discrete r.v.). **Definition**: Let g and F be any functions *defined on a finite closed interval* [a,b], and let F be of bounded variation over [a,b]. Let $\Delta$ be a *partition* of [a,b], $\Delta$: $a = x_0 < x_1 < x_2 < .... < x_n = b$. *Consider* $S(\Delta) = \Sigma_{k=1}^{n} g(x_{k-1}')[F(x_k)-F(x_{k-1})]$, *where* $x_{k-1}' \in (x_{k-1}, x_k)$. If, as $n \to \infty$, and as ${}^{\max}{}_{1\leq k\leq n} (x_k-x_{k-1}) \to 0$, the *sum* $S(\Delta)$ tends to a **finite** limit, *independent of the choice of the points* $x_{k-1}'$ and $\Delta$, then this limit is called the **Riemann-Stieltjes (R-S) Integral** of g w.r.t. F, and is *written* as $\int^{b}_{a} g(x)dF(x)$.

**Notes**: (1) *Sufficient* conditions for the existence of the R-S Integral are: (a) F is of bounded variation and g is continuous, or (b) F is cts and g is of bounded variation. (2) If F has *a cts derivative F'*, then $\int^{b}_{a} g(x)dF(x) = \int^{b}_{a} g(x)F'(x)dx$ (= the *Riemann* Integral).

(3) If F is the *cdf of a discrete r.v.*, with jumps at points $x_{k-1}'$ of size $p(x_{k-1}')$, then $\int^{b}_{a} g(x)dF(x) = \Sigma_k g(x_{k-1}')p(x_{k-1}')$. (4) If the *interval* [a, b] is **not** finite, then we have $\int_{-\infty}^{\infty} g(x)dF(x) = {}^{\lim}{}_{b\to\infty,\, a\to-\infty} \{\int^{b}_{a} g(x)dF(x)\}$. (5) *By (2) and (3) above*, the **expected** value of g(X) for a discrete **or** continuous r.v. can be written as $E(g(X)) = \int_{-\infty}^{\infty} g(x)dF(x)$, where this is *absolutely convergent*, i.e. $\int_{-\infty}^{\infty}|g(x)|dF(x) < \infty \Rightarrow \int_{-\infty}^{\infty} g(x)dF(x) < \infty$.

## 2.5: n-Dimensional Random Variables (Joint r.v.'s)

Let (S, $\mathcal{F}$, P) be our *basic probability space*. The **ordered** n-tuple $(X_1, X_2, ..., X_n)$ of *n real valued functions* defined on S is called an n-dim r.v. if for **arbitrary** reals $x_1, x_2, ..., x_n$, the *set* $\{e \in S \mid X_i(e) \subseteq x_i,\ i = 1, 2, ..., n\} \in \mathcal{F}$. So $(X_1, X_2, ..., X_n)$ *induces* from (S, $\mathcal{F}$, P) the space $(\mathbf{R}^n, \beta^n, P)$, where $\beta^n$ are *Borel sets* in $\mathbf{R}^n$.

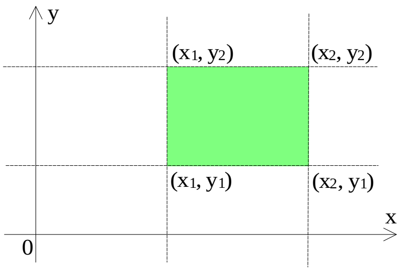
1st March 2001

**Example**: Let S be the set of *students* at UWB. Define on S a 4-dimensional r.v. $(X_1, X_2, X_3, X_4)$, where $X_1$ is the *height* of a student, $X_2$ the *weight*, $X_3$ the *income*, and $X_4$ the *age*. So to **each** student $e_n$, the r.v. assigns 4 real numbers $X_1(e_n)$, $X_2(e_n)$, $X_3(e_n)$, and $X_4(e_n)$. The *cdf of* $(X_1, X_2, ..., X_n)$ is defined for *every* point $(x_1, x_2, ..., x_n) \in \mathbf{R}^n$ as $F(x_1, x_2, ..., x_n) = P(X_1\leq x_1, X_2\leq x_2, ..., X_n\leq x_n)$.

The *cdf* of $(X_1, X_2, ..., X_k)$, where $k < n$, a subset of components of $(X_1, X_2, ..., X_n)$, is defined as $F_{(k)}(x_1, x_2, ..., x_k) = F(x_1, x_2, ..., x_k, \infty, \infty, ..., \infty)$, and is called the **marginal** cdf of $(X_1, X_2, ..., X_k)$. Similarly *for any other subset*. In particular, the marginal cdf of $X_i$, the $i^{th}$ component, is $F_i(x_i) = F(\infty, \infty, ..., \infty, x_i, \infty, ..., \infty)$ $(i = 1, 2, ..., n)$. The *random variables* $X_1, X_2, ..., X_n$ are termed *independent* iff $F(x_1, x_2, ..., x_n) = \Pi_{i=1}^{n} F_i(x_i)$.

### 2.5.1: The Bivariate or 2-Dimensional Case

There are good reasons for *considering these*. We might be **interested** in (i) the joint distribution of height X of a randomly chosen father and height Y of his adult son; (ii) the expected IQ Y of an *11 year old* given that he comes from a family with X number of children; and (iii) the joint distribution of *thrust* X and *mixture ratio* Y of rocket fuel, etc.



The cdf of a 2-dim r.v. is given by $F(x, y) = P(X\leq x, Y\leq y)$ with $(x, y) \in \mathbf{R}^2$. It is a *real non-negative function* of $(x, y)$. (i) $F(x, y)$ is **non-decreasing** in each variable. (ii) $F(x, -\infty) = F(-\infty, y) = 0$ (iii) $F(\infty, \infty) = 1$. (iv) $F(x+, y) = F(x, y+) = F(x,y)$ (*right continuity* in each variable). (v) $F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \geq 0$, where $(x_1, y_1)$ and $(x_2, y_2)$ are *any 2 points in* $\mathbf{R}^2$ s.t. $x_2 > x_1$ and $y_2 > y_1$. (v) *ensures* that $P(x_1<X\leq x_2, y_1<Y\leq y_2) \geq 0$. ($P = P(X\leq x_2, Y\leq y_2) - P(X\leq x_2, Y\leq y_1) - P(X\leq x_1, Y\leq y_2) + P(X\leq x_1, Y\leq y_1)$). <u>**Note**</u>: (i) to (iv) are *extensions* of the 1-dimensional case. The *marginal cdf's* of X and Y are given by $F_1(x) = F(x, \infty)$, and $F_2(y) = F(\infty, y)$. Interpret $F_1$ as the cdf of X with *no restriction of* Y, or when Y is *ignored*. Similarly for $F_2$.

## 2.5.1.1: The Discrete 2-Dimensional Case

(X, Y) is termed **discrete** if its range of values $(x_i, y_j)$ ($i, j = 1, 2, ...$) is **countable** such that $P(X=x_i, Y=y_j) = p(x_i, y_j) > 0$ for all $(x_i, y_j)$; and $\Sigma_i\Sigma_j\, p(x_i,y_j) = 1$. p is the *pmf* of (X,Y). The cdf is given by $F(x, y) = \Sigma_{x_i\leq x}\Sigma_{y_j\leq y}\, p(x_i,y_j)$. In *general*, $P((X,Y) \in B) = \Sigma\Sigma_{(x_i,y_j)\in B}\, p(x_i, y_j)$. The *marginal* pmf's are given by $p_X(x_i) = \Sigma_j p(x_i,y_j)$, and $p_Y(y_j) = \Sigma_i p(x_i,y_j)$. X and Y are *independent* iff $p(x_i,y_j) = p_X(x_i)p_Y(y_j)\ \forall\ (x_i, y_j)$.

## 2.5.1.2: The Continuous 2-Dimensional Case

(X, Y) is called a *cts r.v.* if $\exists$ a function $f(x,y) \geq 0$ such that for **all** $(x, y) \in \mathbf{R}^2$, we have $F(x, y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy$. (f is called the *pdf* of (X,Y)). Now $\partial^2F/\partial x\partial y = f(x,y)$ *almost* everywhere, i.e. at all points of **continuity** of f. <u>Further</u>, $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$, and $P(a<X\leq b, c<Y\leq d) = \int_a^b\int_c^d f(x,y)dxdy$. <u>**Marginals:**</u> $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$, and $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$. X and Y are *independent* iff $f(x,y) = f_X(x)f_Y(y)$ for *all* $(x, y) \in \mathbf{R}^2$.

(**Q21**) Let X be the number of **aces**, and Y the number of **kings**, in a hand of 13 cards drawn from a pack of 52 ordinary playing cards. What are the values of the *2-dimensional r.v.* (X, Y)? Find the <u>probability mass function</u> and the <u>marginal probability functions</u>. Are X and Y *independent*? Find the probability that a hand contains at **most** 2 aces and at **least** 3 kings. A: The **values** of (X, Y) are given by (x, y), where x, y = 0, 1, 2, 3, 4 (25 pairs). *The distribution* of (X, Y) is as follows: $P(X=x, Y=y) = \binom{4}{x}\binom{4}{y}\binom{44}{13-x-y}/\binom{52}{13}$ (x, y = 0, 1, 2, 3, 4). <u>*Marginal Distribution of X, Y*</u>: $p_X(x) = \Sigma_{y=0}^{4}\, p(x,y) = [\binom{4}{x}/\binom{52}{13}]\Sigma_{y=0}^{4}\binom{4}{y}\binom{44}{13-x-y}$ = (result *from the book*: $\Sigma_{t=0}^{k}\binom{k}{t}\binom{n}{r-t} = \binom{k+n}{r}$, where $n \geq r \geq k$) $= [\binom{4}{x}/\binom{52}{13}]\binom{48}{13-x}$, for x = 0, 1, 2, 3, 4. *Similarly*, $p_Y(y) = \Sigma_{x=0}^{4}\, p(x,y) = ... = \binom{4}{y}\binom{48}{13-y}/\binom{52}{13}$, for y = 0, 1, 2, 3, 4. X and Y are *independent* iff $p(x, y) = p_X(x)p_Y(y)$ for all x, y. **But** $p(0, 0) = \binom{4}{0}\binom{4}{0}\binom{44}{13}/\binom{52}{13} = \binom{44}{13}/\binom{52}{13}$, and $p_X(0) = \binom{4}{0}\binom{48}{13}/\binom{52}{13} = \binom{48}{13}/\binom{52}{13} = p_Y(0)$, so that $p(0,0) \neq p_X(0)p_Y(0)$. To finish, $P(X\leq 2, Y\geq 3) = \Sigma_{x=0}^{2}\Sigma_{y=3}^{4}\, p(x,y)$.

(**Q22**) The performance of a rocket engine depends on *thrust X* and mixture ratio of fuel Y. Assuming that (X, Y) is a continuous r.v., with pdf $f(x, y) = 2(x+y-2xy)$ for $0\leq x\leq 1$ and $0\leq y\leq 1$; and $f(x, y) = 0$ otherwise, show that the *pdf's of both X and Y* are **uniform** over the unit interval. A: $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{y=0}^{y=1} 2(x+y-2xy)dy = [2xy+y^2-2xy^2]_{y=0}^{y=1} = 2x+1-2x = 1$. *Similarly*, $f_Y(y) = ... = 1$ (<u>symmetrical</u>).

(**Q23**) In question 21, what is the probability that a hand contains *3 aces*, given that it contains *2 kings?* A: $P(X=3 \mid Y=2) = {}^{P(X=3,\ Y=2)}/_{P(Y=2)} = \{[\binom{4}{3}\binom{4}{2}\binom{44}{8}/\binom{52}{13}]/[\binom{4}{2}\binom{48}{11}/\binom{52}{13}]\} = \binom{4}{3}\binom{44}{8}/\binom{48}{11} = 0.03138$.

## The Bivariate Normal Distribution

$f(x,y) = \dfrac{\exp\{\frac{-1}{2(1-\rho_{XY}^2)}[(\frac{x-\mu_X}{\sigma_X})^2 - 2\rho_{XY}(\frac{x-\mu_X}{\sigma_X})(\frac{y-\mu_Y}{\sigma_Y}) + (\frac{y-\mu_Y}{\sigma_Y})^2]\}}{(2\pi)\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}$, for $-\infty < x, y < \infty$. Now $\mu_X$, $\mu_Y$, $\sigma_X$, $\sigma_Y$ and $\rho_{XY}$ are **parametres**, where $\sigma_X, \sigma_Y > 0$, and $|\rho_{XY}| < 1$. *Simple form of the pdf*: $\mu_X = \mu_Y = 0$, and $\sigma_X = \sigma_Y = 1$. *Example*: show that if (X, Y) has bivariate normal distribution, then X and Y each has **univariate** normal distribution. Also, if $\rho_{XY} = 0$, show that X and Y are *independent*.

A: Use the *simple* form. $f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy$. *Complete the square*, and make the substitution $t = (y-\rho_{XY}x)/\sqrt{(2(1-\rho^2_{XY}))}$ to get $f_X(x) = [e^{-\frac{1}{2}x^2}/\pi\sqrt{(2)}]\int_{\infty}^{-\infty} e^{-t^2}dt = ({}^1/_{\sqrt{2\pi}})e^{-\frac{1}{2}x^2}$, the *pdf* of N(0,1), as $\int_{-\infty}^{\infty} e^{-t^2}dt = \sqrt{\pi}$. *Similarly*, $f_Y(y) = {}^1/_{\sqrt{(2\pi)}}e^{-\frac{1}{2}y^2}$. If $\rho_{XY} = 0$, then $f(x,y) = {}^1/_{(2\pi)}e^{-\frac{1}{2}(x^2+y^2)} = ({}^1/_{\sqrt{(2\pi)}}e^{-\frac{1}{2}x^2})({}^1/_{\sqrt{(2\pi)}}e^{-\frac{1}{2}y^2}) = f_X(x)f_Y(y)$ for *all* $(x,y) \in \mathbf{R}^2 \Rightarrow$ X and Y are independent.

6th March 2001

## 2.5.1.3: Conditional Distributions

Let (X, Y) be *discrete*, with pmf p(x,y). The **conditional** pmf of X given Y = y is defined as $p(x|y) = p(x,y)/p_Y(y)$, where $p_Y(y) > 0$, and where $p_Y$ is the *marginal* pmf of Y. **Similarly,** if (X, Y) is *continuous* with pdf f(x, y), then the pdf of X given Y = y is given by $f(x|y) = f(x,y)/f_Y(y)$, where $f_Y(y) > 0$. Clearly, X and Y are *independent* if $f(x|y) = f_X(x)$, or if $f(y|x) = f_Y(y)$.

**Example**: Let (X, Y) have a *bivariate normal distribution*. Show that the conditional distribution of X given Y = $y_0$ is *univariate* normal (use the **simple** form of the pdf). A: The *conditional* pdf of X given Y = $y_0$ is given by $f(x|y_0) = f(x,y_0)/f_Y(y_0)$, where $f_Y(y_0) = {}^1/_{\sqrt{2\pi}}\, e^{-\frac{1}{2}y_0^2}$ (see *before*). Now $f(x|y_0) = [\frac{1}{(2\pi)\sqrt{1-\rho_{XY}^2}}\exp\{-\frac{1}{2(1-\rho_{XY}^2)}(x^2 - 2\rho_{XY}xy_0 + y_0^2)\}]/[\frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}y_0^2\}] = {}^1/_{\sqrt{2\pi}}{}^1/_{\sqrt{1-\rho_{XY}^2}}\ \exp\{-{}^1/_{2(1-\rho_{XY}^2)}(x^2 - 2\rho_{XY}xy_0 + y_0^2(1-(1-\rho^2_{XY})))\} = {}^1/_{\sqrt{2\pi}}{}^1/_{\sqrt{(1-\rho_{XY}^2)}}\exp\{-{}^1/_{2(1-\rho_{XY}^2)}(x-\rho_{XY}y_0)^2\}$, which is *univariate normal:* $N(\rho_{XY}y_0, (1-\rho_{XY}^2))$.

## Assignment 2

Q: A biased coin has *probability p of landing heads*. Anne and Betty toss the coin successively, Anne tossing first, until a **head** occurs. The person tossing the first head wins. Find the probability that *Anne* wins (and state the axiom(s) of a probability measure that you use). A: If Anne wins, she does so by throwing heads on the *1st, 3rd, 5th, ...* throw, with Betty **always** throwing tails. Because each throw is independent of the other (so that the event "*wins on the 3rd throw*" can be obtained by **multiplying** the three events q, q and p), and because no two ways of winning can be obtained *simultaneously*, i.e. $\{A_1, A_3, ...\}$ is a pairwise exclusive sequence, then P(Anne wins) = **(by Axiom 3)** = $P(A_1)+P(A_3)+... = p+qqp+qqqqp+... = \Sigma_{i=0}^{\infty} p(q^2)^i = p({}^1/_{1-q^2})$. **Check**: P(Betty Wins) = $qp+q^3p+... = qp(1+q^2+q^4+...) = qp({}^1/_{1-q^2})$; and so we can check that P(A)+P(B) = 1 holds.

(**Q10**) A population consists of k objects, of which any number of which may be of a *certain type A*. The process of selecting an object at random, noting its type, and **replacing** it in the population, is repeated n times. Given that at each of the n selections, the *type found is A*, find the probability that the population contains exactly r objects of the type A. Find the limiting values of this probability as $n \to \infty$ for all values of r.

A: P(population *contains r objects of type A* | n selections *found* A) = $P(A_r|B) = P(B|A_r)P(A_r)/\Sigma_{i=0}^{n} P(B|A_i)P(A_i)$. *Now* $P(B|A_i) = ({}^{i}/_{k})^{n}$, and $P(A_i) = {}^{1}/_{k+1}$ (*each sample point is* equally likely). Therefore, $P(A_r|B) = [({}^{r}/_{k})^{n}\,{}^{1}/_{k+1}]/[\Sigma_{i=0}^{k}({}^{i}/_{k})^{n}\,{}^{1}/_{k+1}] = (r^n/k^n)/[\Sigma_{i=0}^{k}(i^n/k^n)] = r^n/[\Sigma_{i=1}^{k} i^n]$. (1 *because* $0^n = 0$). Now, as $n \to \infty$, $P(A_r|B) \to r^n/k^n$, as $k^n$ is the **dominant** factor in the denominator. If $r = k$, then $P(A_r|B) \to 1$ as $n \to \infty$; and if $r \neq k$, then $P(A_r|B) \to 0$ *as* $n\to\infty$.

(**Q19**) Look up the **pdf's** of the following common continuous probability distributions: (i) Uniform over [a,b], (ii) Normal $N(\mu, \sigma^2)$, (iii) Cauchy, (iv) Gamma, (v) Exponential (Special case of (iv)), and (vi) Beta. Verify in each case that f(x) is *indeed* a pdf, i.e. that $f \geq 0$, and that $\int f\, dx = 1$. Also of interest in Physics: *Raleigh* and *Maxwell distributions*; also of interest in Statistics: $X^2$, F and t *distributions*.

A: (i) **Uniform** over [a,b]: $f(x) = {}^{1}/_{b-a}$, where $a \leq x \leq b$. *As* $b \geq a$, then $b-a \geq 0$, so that $f(x) \geq 0$. Now $\int_a^b {}^{1}/_{b-a}dx = {}^{1}/_{b-a}\int_a^b dx = {}^{1}/_{b-a}[x]_a^b = {}^{b-a}/_{b-a} = 1$. **<u>QED</u>**. (ii) **Normal** $N(\mu,\sigma^2)$: $f(x) = {}^{1}/_{\sigma\sqrt{2\pi}}\exp\{-{}^{1}/_{2\sigma^2}(x-\mu)^2\}$. As red > 0 (because $\sigma > 0$), and as blue $\geq 0$ (because $\exp(x) \geq 0\ \forall\ x$), then $f(x) \geq 0$. We must *now show that* $\int_{-\infty}^{\infty}f(x)dx = 1$. *Consider* $I = \int_{-\infty}^{\infty} {}^{1}/_{\sigma\sqrt{2\pi}}\exp\{{}^{-(x-\mu)^2}/_{2\sigma^2}\}dx$. Let $y = {}^{x-\mu}/_{\sigma\sqrt{2}}$, so that $y^2 = {}^{(x-\mu)^2}/_{2\sigma^2}$, and $dy = {}^{dx}/_{\sigma\sqrt{2}}$. It follows that $I = {}^{1}/_{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} \exp(-y^2)dy\sqrt{(2)}\sigma$; $I = {}^{\sqrt{2}}/_{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-y^2)dy$. As $\int_{-\infty}^{\infty}\exp(-z^2)dz = \sqrt{\pi}$, then $I = {}^{1}/_{\sqrt{\pi}}(\sqrt{\pi}) = 1$. **<u>QED</u>**.

(iii) **Cauchy**: $f(x) = {}^{1}/_{\pi}\,{}^{\alpha}/_{\alpha^2+(x-\theta)^2}$ ($\alpha > 0$, $-\infty < x < \infty$). *The numerator is* > 0 as $\alpha > 0$. *The denominator is* > 0 as $\pi > 0$, $\alpha^2 > 0$ and $(x-\theta)^2 \geq 0$. Conclusion: $f(x) \geq 0$. **Let** $I = \int_{-\infty}^{\infty} {}^{1}/_{\pi}\,{}^{\alpha}/_{\alpha^2+(x-\theta)^2}dx = {}^{\alpha}/_{\pi} \int_{-\infty}^{\infty} {}^{1}/_{\alpha^2+(x-\theta)^2}dx$. Let $y = x-\theta$, so that $y^2 = (x-\theta)^2$, and $dy = dx$. There is no *change* in the limits. So $I = {}^{\alpha}/_{\pi} \int_{-\infty}^{\infty} {}^{1}/_{\alpha^2+y^2}dy = {}^{\alpha}/_{\pi}[{}^{1}/_{\alpha}\tan^{-1}({}^{y}/_{\alpha})]_{-\infty}^{\infty} = {}^{1}/_{\pi}[\tan^{-1}({}^{y}/_{\alpha})]_{y\to-\infty}^{y\to\infty} = {}^{1}/_{\pi}[{}^{\pi}/_{2}-(-{}^{\pi}/_{2})] = {}^{\pi}/_{\pi} = 1$. **<u>QED</u>**.

(iv) **Gamma**: $f(x) = [\alpha^{\theta}/\Gamma(\theta)]x^{\theta-1}e^{-\alpha x}$, *where* x, $\alpha$, $\theta > 0$, and $\Gamma(\theta) = (\theta-1)!$ when $\theta \in$ **Z**. *Clearly*, $f(x) \geq 0$. Now **assume** that $\theta \in$ **Z,** and let us *find* $I = \int_0^{\infty} [\alpha^{\theta}/\Gamma(\theta)]x^{\theta-1}e^{-\alpha x}dx = [\alpha^{\theta}/(\theta-1)!]\int_0^{\infty} x^{\theta-1}e^{-\alpha x}dx$. Integrate by *parts*: $u = x^{\theta-1}$, ${}^{du}/_{dx} = (\theta-1)x^{\theta-2}$, ${}^{dv}/_{dx} = e^{-\alpha x}$, $v = -{}^{1}/_{\alpha}e^{-\alpha x}$. So $J = [(-x^{\theta-1}/\alpha)e^{-\alpha x}]_0^{\infty} + \int_0^{\infty}({}^{\theta-1}/_{\alpha})x^{\theta-2}e^{-\alpha x}dx = [0-0] + {}^{(\theta-1)}/_{\alpha} \int_0^{\infty}x^{\theta-2}e^{-\alpha x}dx$. It follows that $I = \alpha^{\theta}(\theta-1)/(\theta-1)!\alpha \int_0^{\infty} x^{\theta-2}e^{-\alpha x}dx$. Similarly, *integrate by parts* to obtain $I = \alpha^{\theta}(\theta-1)(\theta-2)/(\theta-1)!\alpha^2\int_0^{\infty} x^{\theta-3}e^{-\alpha x}dx$.

After we *integrate by parts* $(\theta-1)$ times, we will obtain $I = \alpha^{\theta}(\theta-1)!/(\theta-1)!\alpha^{\theta-1} \int_0^{\infty}x^0e^{-\alpha x}dx = \alpha \int_0^{\infty}e^{-\alpha x}dx = \alpha[-{}^{1}/_{\alpha}e^{-\alpha x}]_0^{\infty} = [-e^{-\alpha x}]_0^{x\to\infty} = [-e^{-\infty}-(-e^0)] = 0+1 = 1$. **<u>QED</u>**. (v) **Exponential:** $f(x) = \alpha e^{-\alpha x}$ ($x, \alpha > 0$). As $\alpha > 0$, and as $e^{-\alpha x} > 0$ for all $\alpha$, x, then $f(x) \geq 0$. *Now* $\int_0^{\infty} \alpha e^{-\alpha x}dx = \alpha \int_0^{\infty}e^{-\alpha x}dx = 1$ (*as just worked out above*). **<u>QED</u>**. (*Exponential* is a special case of *Gamma:* $\theta = 1$).

(vi) **Beta**: $f(t) = {}^{1}/_{B(a,b)}t^{a-1}(1-t)^{b-1}dt$ ($0 \leq t \leq 1$), where $B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx$. Now $0 \leq t \leq 1 \Rightarrow 0 \leq 1-t \leq 1$. Any *power* of t and (1-t) will be $\geq 0$, so that $t^{a-1}(1-t)^{b-1} \geq 0$, *implying that* $\int_0^1 x^{a-1}(1-x)^{b-1}dx \geq 0$, and so $f(t) \geq 0$. **Now** $\int_0^1 {}^{1}/_{B(a,b)}t^{a-1}(1-t)^{b-1}dt = {}^{1}/_{B(a,b)}\int_0^1 t^{a-1}(1-t)^{b-1}dt = \int_0^1 t^{a-1}(1-t)^{b-1}dt/\int_0^1 x^{a-1}(1-x)^{b-1}dt = 1$. **<u>QED</u>**.

# 2.6: Functions of a Random Variable

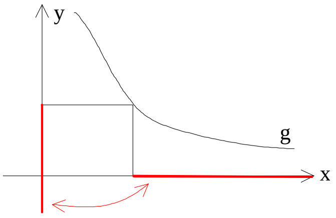
It can be shown that *functions of random variables* that are **monotonic** or **continuous,** and functions obtained from them by algebraic or limiting processes, are also *random variables.* **2.6.1: The one-dimensional case.** *Theorem 6.1*: Let X be a continuous r.v., with pdf f(x). Let $y = g(x)$ be a *strictly monotonic function* of x s.t. $g'(x)$ is *continuous and non-zero* in some open interval A.

Let $g^{-1}(y)$ be the **unique** inverse of g(x), and let B be the *image* of A in the range of y. Then the **random** variable Y = g(X) is cts, *whose pdf h(y) exists* for all $y \in B$, and is **given** by $h(y) = f(g^{-1}(y))|{}^{d}/_{dy}(g^{-1}(y))|$, and $\int_A f(x)dx = \int_B f(g^{-1}(y))|{}^{d}/_{dy}(g^{-1}(y))|dy$. **Proof:** Now g' cts ⇒ g' exists ⇒ g continuous ⇒ Y is cts, *since it is a cts function* of a cts r.v.

*Consider* g ↑ strictly (${}^{dy}/_{dx} > 0$). The **cdf** of Y is given by $H(y) = P(Y \le y) =$ (by Y = g(X)) $= P(g(X) \le y) = P(X \le g^{-1}(y)) = F(g^{-1}(y))$, where F is the cdf of X (red = x). **Hence the** pdf of Y is given by $h(y) = H'(y)$ ($= {}^{dH}/_{dy}$) = (*chain* rule) $= {}^{d}/_{dx}(H(y)){}^{dx}/_{dy} = {}^{d(F(x))}/_{dx}{}^{dx}/_{dy} = f(x){}^{dx}/_{dy}$ = (*by* $x = g^{-1}(y)$) $= f(g^{-1}(y)){}^{d}/_{dy}(g^{-1}(y)) = f(g^{-1}(y))|{}^{d}/_{dy}(g^{-1}(y))|$ as *required.*

8th March 2001

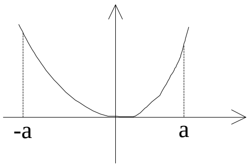
Now **consider** g ↓ *strictly* (${}^{dy}/_{dx} < 0$). Then $H(y) = P(Y \le y) = P(g(X) \le y) = P(X \ge g^{-1}(y))$ = (see the *diagram*) $= 1-P(X \le g^{-1}(y)) = 1-F(g^{-1}(y))$ (red = x). *Therefore,* $h(y) = H'(y)$ = (*chain* rule) $= {}^{d}/_{dx}(1-F(x)){}^{dx}/_{dy} = -{}^{dF}/_{dx}.{}^{dx}/_{dy} = {}^{dF}/_{dx}(-{}^{dx}/_{dy})$. *Note*: ${}^{dF}/_{dx} = f(x)$, the pdf of x. *Now* ${}^{dy}/_{dx} < 0$, so that ${}^{dx}/_{dy} < 0 \Rightarrow -{}^{dx}/_{dy} > 0$. Therefore, $h(y) = f(x)|{}^{dx}/_{dy}| = f(g^{-1}(y))|{}^{d}/_{dy}(g^{-1}(y))|$. The *second part* follows from the *first part*, and the fact that *B is the image of A* under g.



**Note 1:** The form of g and the *domain of X* determine whether or not g is monotonic, e.g. (i) $g(x) = x^2$ (monotonically **increasing** for x > 0, and **non-monotonic** for $-a < x < b$, where a, b > 0); and (ii) $g(x) = x^3$ (**monotonic** on **R**). **Note 2**: If g is *not* monotonic, then Theorem 6.1 is not applicable — use a direct approach: find the cdf of Y in *terms* of the cdf of X by finding the **event** in the domain of X equivalent to the event ($Y \le y$). Then *differentiate* w.r.t. y to obtain the pdf of Y in terms of a given pdf of X, and then substitute. The **domain** of Y is formed from that of X, and the *form* of g.

(**Q24**) If X is N(0, 1), find the pdf of $X^{1/3}$. A: Let $X \sim N(0, 1^2)$ i.e. $f(x) = {}^{1}/_{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}$ ($-\infty < x < \infty$). Find the *distribution* of $Y = X^{1/3}$. Now put $y = g(x) = x^{1/3}$ (monotonic *increasing*), so that $x = y^3 = g^{-1}(y)$. Now ${}^{dx}/_{dy} = 3y^2$ (${}^{d}/_{dy}(g^{-1}(y))$). Using Theorem 6.1 implies that the pdf of Y is given by $h(y) = f(g^{-1}(y))|{}^{d}/_{dy}(g^{-1}(y))| = {}^{1}/_{\sqrt{2\pi}}e^{-\frac{1}{2}(y^3)^2}|3y^2| = {}^{3y^2}/_{\sqrt{2\pi}}\exp(-\frac{1}{2}y^6)$ ($-\infty < y < \infty$).

Often, the conditions of *theorem 4* are not met, and we must use a **direct** method (start by writing down the *cumulative distribution function*). (**Q25**) Let X have a uniform distribution, with pdf ($F'(x) =$) $f(x) = {}^{1}/_{2a}$ if $|x| < a$ and a > 0; and f(x) = 0 *otherwise*. Find the **distribution** of $Y = X^2$. A: Let $y = g(x) = x^2$. This is *no monotonic*, since the range of x includes x = 0.



Now $H(y) = P(Y \le y) = P(X^2 \le y) = P(|X| \le \sqrt{y}) = P(-\sqrt{y} \le X \le \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y})$, since *Y is continuous*. The pdf of Y is given by $h(y)$ [= $H'(y)$] = $^{dH(y)}/_{dy} = {}^{d}/_{dy}(F(\sqrt{y}) - F(-\sqrt{y}))$ = (by the *chain* rule) = $F'(\sqrt{y})^{1}/_{2\sqrt{y}} - F'(-\sqrt{y})(-^{1}/_{2\sqrt{y}}) = {}^{1}/_{2\sqrt{y}}(F'(\sqrt{y}) + F'(-\sqrt{y})) = {}^{1}/_{2\sqrt{(y)}}[f(\sqrt{y}) + f(-\sqrt{y})] = {}^{1}/_{2\sqrt{y}}({}^{1}/_{2a} + {}^{1}/_{2a}) = {}^{1}/_{2a\sqrt{y}}$ if $0 < y < a^2$, and 0 *otherwise*.

## 2.6.2: The 2-Dimensional Case — Not in the Exam!

**Theorem 6.2:** Let (X, Y) be cts, with pdf $f(x, y)$; and let $u = g_1(x, y)$, and $v = g_2(x, y)$, be *single-valued, continuous* functions, with cts first order partial derivatives in the (x,y)-plane. Let the **Jacobian** of the transformation $({}^{\partial(u,v)}/_{\partial(x,y)})$ be non-zero in A. Let $x = g_1^{-1}(u, v)$ and $y = g_2^{-1}(u, v)$ be *unique* inverses for all (x, y) in A. Let B be the **image** of A in the uv-plane.

Then, the random variable (U, V), with $U = g_1(X, Y)$, and $V = g_2(X, Y)$, is cts, whose *pdf exists for all* (u, v) in B, and is given by $h(u,v) = f(g_1^{-1}(u,v), g_2^{-1}(u,v))|J|$, where $\iint_A f(x,y)dxdy = \iint_B f(g_1^{-1}(u,v), g_2^{-1}(u,v))|J|dudv$. *The proof is omitted* — refer to the G2M56 course, where $J = {}^{\partial(x,y)}/_{\partial(u,v)}$. **Note**: $J = ({}^{\partial(u,v)}/_{\partial(x,y)})^{-1} = \left(\begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix}\right)^{-1}$ (*substitute x*).

**Example**: Let the joint pdf of (X, Y) be given by $f(x, y) = e^{-(x+y)}$ for $x > 0$ and $y > 0$, and $f(x, y) = 0$ otherwise. Find the *pdf of (U, V)*, where $U = X+Y$, and $V = {}^{Y}/_{X}$. A: $u = x+y$, and $v = {}^{y}/_{x} \Rightarrow x = {}^{u}/_{1+v}$, and $y = {}^{uv}/_{1+v}$. Therefore, $J = {}^{\partial(x,y)}/_{\partial(u,v)} = \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} = \begin{vmatrix} 1/1+v & -u(1+v)^2 \\ v/1+v & u/(1+v)^2 \end{vmatrix} = {}^{u}/_{(1+v)^2}$. Hence the *joint pdf of* (U, V) is given by $h(u, v) = e^{-u}|{}^{u}/_{(1+v)^2}| = {}^{u\exp(-u)}/_{(1+v)^2}$ for $u, v > 0$, and $h(u, v) = 0$ *otherwise*.

## Tutorial

Q: Given the *bivariate normal distribution* (in its simple form) $f(x,y)$ = (**using** $\rho = \rho_{XY}$) = ${}^{1}/_{2\pi\sqrt{1-\rho^2}}\exp\{-^{1}/_{2(1-\rho^2)}(x^2-2\rho xy+y^2)\}$, show that the *marginal pdf*, $f_X(x)$, is **given** by $f_X(x)$ = (by definition) = $\int_{-\infty}^{\infty} f(x,y)dy = {}^{1}/_{\sqrt{2\pi}}\exp(-^{x^2}/_{2})$. A: Let $I = \int_{-\infty}^{\infty} {}^{1}/_{2\pi\sqrt{1-\rho^2}}\exp\{-^{1}/_{2(1-\rho^2)}(x^2-2\rho xy+y^2)\}dy$. Now let $t = {}^{y-\rho x}/_{\sqrt{[2(1-\rho^2)]}}$, so that $t^2 = {}^{(y-\rho x)^2}/_{2(1-\rho^2)}$, and $dt = {}^{dy}/_{\sqrt{[2(1-\rho^2)]}}$. Now *as* $x^2-2\rho xy+y^2 = (y-\rho x)^2 + (1-\rho^2)x^2$, then $I = \int_{-\infty}^{\infty} {}^{1}/_{2\pi\sqrt{(1-\rho^2)}}\exp\{-^{(y-\rho x)^2}/_{2(1-\rho^2)} - {}^{(1-\rho^2)x^2}/_{2(1-\rho^2)}\}dy = \int_{-\infty}^{\infty} {}^{1}/_{2\pi\sqrt{(1-\rho^2)}}\exp\{-t^2-^{x^2}/_{2}\}dt\sqrt{[2(1-\rho^2)]} = {}^{\sqrt{(2(1-\rho^2))}}/_{2\pi\sqrt{(1-\rho^2)}}\int_{-\infty}^{\infty} \exp\{-t^2-^{x^2}/_{2}\}dt = \exp\{-^{x^2}/_{2}\}^{\sqrt{2}\sqrt{(1-\rho^2)}}/_{2\pi\sqrt{(1-\rho^2)}}\int_{-\infty}^{\infty} \exp\{-t^2\}dt$. Now the red bit is $\sqrt{\pi}$, so we have $I = \exp\{-^{x^2}/_{2}\}^{1}/_{\sqrt{(2)}\pi}(\sqrt{\pi}) = \exp\{-^{x^2}/_{2})^{1}/_{\sqrt{2\pi}}$ as *required*. **QED**.

Q: If f is the *pdf of a cts r.v.* X, find the **pdf** of $Y = aX+b$ (a and b are *constants*, with $a \neq 0$). A: $y = ax+b$ ($a \neq 0$). Now $x = {}^{y-b}/_{a} \Rightarrow {}^{dx}/_{dy} = {}^{1}/_{a}$. Therefore, $h(y) = f({}^{y-b}/_{a})^{1}/_{|a|}$. Q: If X has *distribution* with pdf $f(x) = {}^{1}/_{a}$ for $0 < x < a$ ($a > 0$), and $f(x) = 0$ *otherwise*, find the **distribution** of $X^n$ ($n > 0$). A: $y = x^n \Rightarrow x = y^{1/n}$ (+**ve** root). Now ${}^{dx}/_{dy} = {}^{1}/_{n}y^{((1/n)-1)}$, and the pdf of Y is thus given by $h(y) = {}^{1}/_{a}{}^{1}/_{n}y^{(1/n)-1}$ for $0 < y < a^n$, and $h(y) = 0$ *otherwise*. Note: for these 2 questions, use Theorem 6.1 because of **monotonicity**.

13th March 2001

# Section 3: Expected Values and Moments

## 3.1: Expected Values of 1-Dimensional Random Variables

**Definition**: Let X be a 1-dim r.v, and let g(X), a *function* of X, be itself a (1-dim) r.v. (a Borel function on X) such that it is integrable w.r.t. F(x), the **cdf** of X, on $(-\infty, \infty)$. Then we define the expected value (the *expectation* or *mean*), denoted by E[g(X)], as $E[g(X)] = \int_{-\infty}^{\infty} g(x)dF(x)$ (---(1)), provided the integral is *absolutely convergent* (i.e. $\int_{-\infty}^{\infty} |g(x)|dF(x) < \infty$), and we say that the *expected* value does exist.

If $\int_{-\infty}^{\infty} |g(x)|dF(x) = \infty$ (*divergent*), we say that the expected value does not exist. In particular, if X is a cts r.v. with pdf f(x) (= F'(x)), then (1) **becomes** $E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$ (---(2)). If X is discrete, taking *values* $x_n$ with *probabilities* $p(x_n)$, then (1) becomes $E[g(X)] = \Sigma_n g(x_n)p(x_n)$ (---(3)).

**Theorem 1.1:** Let *functions* $g_1(X)$, $g_2(X)$, ... (i = 1, ..., n) be random variables, whose expected values exist. Then (i) for *constants* $a_i$ (i = 1, ..., n), $E[\Sigma_{i=1}^{n} a_ig_i(X)] = \Sigma_{i=1}^{n} a_iE[g_i(X)]$. (In *particular*, E[ag(X)] = aE[g(X)], and *also* E[a] = a). (ii) If $m \leq g(X) \leq M$, then $m \leq E[g(X)] \leq M$. And if $g_1(X) \leq g_2(X)$, then (iii) $E[g_1(X)] \leq E[g_2(X)]$, and (iv) $|E[g_1(X)]| \leq |E[g_2(X)]|$. The **proofs** follow from the *propositions* of an R-S Integral.

## 3.2: Moments of 1-Dimensional Random Variables

Moments about an *arbitrary* point: if, in the previous section, we take $g(X) = (X-a)^r$, where a is a **real** constant, and r is a **positive** integer, then $E[(X-a)^r]$ is called the $r^{th}$ moment about the point a of the r.v. X. *Moments* about the origin (or ordinary moments): if, in the above, we have a = 0, then $E[X^r]$ is called the $r^{th}$ moment about the origin, *denoted* by $\mu_r'(X)$, or simply by $\mu_r'$. In *particular*, if r = 1, then $\mu_1'$ is called the *mean* of X, and is denoted by μ.

**Putting** a = μ, we get $E[(X-\mu)^r]$, the $r^{th}$ *central moment* of a r.v. X, denoted by $\mu_r$. In *particular*, $\mu_2 = E[(X-\mu)^2]$ is called the **variance** of X, and is denoted by $\sigma^2$ or var X (σ is called the *standard deviation* of X). We have $\mu_2 = E[(X-\mu)^2]$ = (Theorem 1.1) = $E[X^2] - (E[X])^2 = \mu_2'-(\mu)^2$. Also, $E[|X|^r]$ is called the $r^{th}$ *absolute* moment about the origin; $E[|X-\mu|^r]$ is called the $r^{th}$ *absolute central moment*; and $E[X(X-1)(X-2)...(X-r+1)]$ is called the $r^{th}$ *factorial moment*, denoted by $\mu_{[r]}$.

**Example**: Let X be a r.v. with *values* $x_n = (-1)^n2^n/n$, where n = 1, 2, 3, ...; and *pmf* $p(x_n) = 1/2^n$. Show that the *mean* of X does **not** exist (even *though* $\Sigma_{n=1}^{\infty} x_np(x_n) = \Sigma_{r=1}^{\infty} (-1)^n (^1/_n) = -\ln(2)$). A: $\Sigma_{n=1}^{\infty} |x_n|p(x_n) = \Sigma_{n=1}^{\infty}|(-1)^n2^n/n|1/2^n = \Sigma_{n=1}^{\infty} {}^1/_n = \infty$ (*because* $\Sigma{}^1/_n = 1 + (½) + (^1/_3+^1/_4) + (^1/_5+^1/_6+^1/_7+^1/_8) + (^1/_9+...+^1/_{16}) > 1 + ½ + (^1/_4+^1/_4) + (^1/_8+^1/_8+^1/_8+^1/_8) + (^1/_{16}+...+^1/_{16}) + ... = 1+½+½+½+... = \infty$.

15th March 2001

**Example**: Let X be a r.v. which takes on *non-negative values* only, and such that its mean $\mu$ exists. *Show* that $\mu = \int^{\infty}_{0}(1-F(x))dx$ (F(x) is the *cdf* of X). **Proof:** $\mu = \int^{\infty}_{0} xdF(x) < \infty$. Now $\int_0^b xdF(x)$ = (by *parts*) = $[xF(x)]^b_0 - \int^b_0 F(x)dx = bF(b)-0-\int_0^bF(x)dx = -b(1-F(b))+\int^b_0(1-F(x))dx$. But $b(1-F(b)) = (F(\infty) = 1) = b\int_b^{\infty}dF(x) < \int_b^{\infty} xdF(x) \to 0$ as $b \to \infty$, as $\mu < \infty$. Therefore, $\mu = {}^{\lim}{}_{b\to\infty}\int^b_0 xdF(x) = \int_0^{\infty}(1-F(x))dx$.

**Theorem 2.1:** If $\mu_r'$ exists, then $\mu'_k$ exists for all k = 1, 2, 3, .., r-1. **Proof:** If $\mu_r'$ exists, then $\int_{-\infty}^{\infty} |x|^rdF(x) < \infty$. Now, *let n be any positive integer*, then $\int_{-\infty}^{\infty} |x|^ndF(x) = \int_{|x|\le1} |x|^ndF(x) + \int_{|x|>1} |x|^ndF(x) \le \int_{|x|\le1} 1dF(x) + \int_{|x|>1} |x|^ndF(x) \le 1 + \int_{|x|>1} |x|^ndF(x)$. But *for all* $1 \le k \le r-1$, $\int_{|x|>1} |x|^kdF(x) \le \int_{|x|>1} |x|^rdF(x) < \infty$. Therefore, $\int_{-\infty}^{\infty} |x|^kdF(x) \le 1+\int_{|x|>1} |x|^rdF(x) < \infty$, *completing the proof*. **Note**: If $\mu_r'$ does *not* exist, then $\mu'_n$ for n > r do **not** exist.

## 3.3: Standard Random Variables

A *random variable is called standard* if its **mean** is zero and its **variance** is 1. Clearly, if X is a r.v. with *mean* $\mu$ and *variance* $\sigma^2$, then $Y = {}^{X-\mu}/_{\sigma}$. **Proof:** $E(Y) = E({}^{X-\mu}/_{\sigma}) = {}^1/_{\sigma}E(X-\mu) = {}^1/_{\sigma}(E(X)-E(\mu)) = {}^1/_{\sigma}(\mu-\mu) = 0$. And $var(Y) = E(Y-E(Y))^2 = E(Y^2)-[E(Y)]^2 = E({}^{X-\mu}/_{\sigma})^2-0 = {}^1/_{\sigma^2}E(X-\mu))^2 = {}^1/_{\sigma^2}(\sigma^2) = 1$. **QED**.

**Example:** If X has the *binomial distribution*, with pmf $p(x) = ({}^n{}_x)p^x(1-p)^{n-x}$ for x = 0, 1, 2, ... ($\mu = np$, $\sigma^2 = np(1-p)$), then ${}^{X-np}/_{\sqrt{np(1-p)}}$ is *standard*. **Example**: *Poisson* Distribution: $p(x) = e^{-\mu}\mu^x/x!$ for x = 0, 1, 2, ... ($E(x) = var(x) = \mu$): ${}^{X-\mu}/_{\sqrt{\mu}}$ is *standard*. **Example**: If $X\sim N(\mu, \sigma^2)$, then $Y = {}^{X-\mu}/_{\sigma^2}$ is *standard*, $\sim N(0, 1^2)$.

**Evidence:** The pdf is $f(x) = {}^1/_{\sigma\sqrt{(2\pi)}}\exp\{-{}^1/_{2\sigma^2}(x-\mu)^2\}$. **Since** $E(X) = \mu$, and since $var(X) = \sigma^2$, then $Y = {}^{X-\mu}/_{\sigma}$ is *standard*, i.e. E(Y) = 0, and var(Y) = 1. The *distribution* of Y is $N(0, 1^2)$, i.e. with pdf ${}^1/_{\sqrt{(2\pi)}}\exp\{-½y^2\}$. *Recall* that $h(y) = f(g^{-1}(y))|{}^d/_{dy}(g^{-1}(y))|$. Now $X = \mu+\sigma Y \Rightarrow {}^{dx}/_{dy} = \sigma$. So $h(y) = {}^1/_{\sigma\sqrt{(2\pi)}}\exp\{-½y^2)\sigma = {}^1/_{\sqrt{(2\pi)}}\exp\{-½y^2\}$.

Effect of translation and change of scale on central moments: Let $\mu_r(X)$ be the r^th^ central moment of a r.v. X. If Y = aX+b, then $\mu_r(Y) = a^r\mu_r(X)$. **Proof:** $\mu_r(Y) = E(Y-E(Y))^r = E(aX+b - E(aX+b))^r = E(aX+b-(aE(X)+b))^r = a^rE(X-E(X))^r = a^rE(X-\mu)^r = a^r\mu_r(X)$.

**Tutorial** Problems. Let $X\sim N(½, 3^2)$. Find the *mean and central moments* of the r.v. Y = 2X-7. E(Y) = E(2X-7) = 2E(X)-7 = 2(½)-7 = -6. We now want to find $\mu_r(Y)$. Let r be a *positive integer*. Now $\mu_{2r+1}(X) = E((X-\mu)^{2r+1}) = \int_{-\infty}^{\infty}(x-\mu)^{2r+1}\ {}^1/_{\sigma\sqrt{2\pi}}\exp\{-½({}^{x-\mu}/_{\sigma})^2\}dx = 0$ (by *symmetry*), and $\mu_{2r}(X) = 1.3.5....(2r-3)(2r-1)\sigma^{2r}$, where $X\sim N(\mu, \sigma^2)$.

**Proof:** $\mu_{2r}(X) = E(X-\mu)^{2r} = \int_{-\infty}^{\infty} (x-\mu)^{2r}\ {}^1/_{\sigma\sqrt{(2\pi)}}\exp\{-½({}^{x-\mu}/_{\sigma})^2\}dx$. Let $y = {}^{x-\mu}/_{\sigma}$, so that $dy = {}^{dx}/_{\sigma}$, and we thus have $\int_{-\infty}^{\infty} (y\sigma)^{2r}\ {}^1/_{\sqrt{(2\pi)}}\exp\{-½y^2\}dy = \sigma^{2r}/\sqrt{(2\pi)} \int_{-\infty}^{\infty} y^{2r}\exp\{-½y^2\}dy$. *Look* at ${}^d/_{dy}e^{-½y^2} = -ye^{-½y^2}$. Let $J = \int_{-\infty}^{\infty} ye^{-½y^2}dy$ (set r = 1) $= \int_{-\infty}^{\infty} (-y)(-ye^{-½y^2}dy) = e^{-½y^2}.-y$. We now show that $I_{2r} = (2r-1)I_{2r-2}$. Let $K_{2r} = \int_{-\infty}^{\infty} y^{2r}\exp\{-½y^2\}dy = \int_{-\infty}^{\infty} y.y^{2r-1}\exp\{-½y^2\}dy$. Let $u = y^{2r-1}$, ${}^{du}/_{dy} = (2r-1)y^{2r-2}$; ${}^{dv}/_{dy} = y\exp\{-½y^2\}$, and so $v = -e^{-½y^2}$. Therefore, $K_{2r} = \int_{-\infty}^{\infty} e^{-½y^2}(2r-1)y^{2r-2} = (2r-1)K_{2r-2}$ ([uv] = 0, the -*ve's* cancel). It follows that $I_0 = \int_{-\infty}^{\infty} e^{-½y^2}dy = ... = \sqrt{2\pi}$. So we *obtain the result* using the recurrence relation.

**Example**: Let X be a r.v. with values $x_n = (-1)^n(2^n/n)^{½}$, where n = 1, 2, 3, ...; and *pmf* $p(x_n) = 1/2^n$. Show that the mean *exists*, but that the variance does **not** exist. A: *Consider* $\Sigma_{n=1}^{\infty} |x_n|p(x_n) = \Sigma_{n=1}^{\infty} |(-1)^n(2^n/n)^{½}|1/2^n = \Sigma_{n=1}^{\infty} (2^n/n)^{½}\ 1/2^n = \Sigma_{n=1}^{\infty} 1/\sqrt{n}2^n < \Sigma_{n=1}^{\infty} 1/2^{n/2} = \Sigma_{n=1}^{\infty} (1/2^{½})^n < \infty$. *Aside*: $S_n = r+r^2+..+r^n$; $rS_n = r^2+...+r^{n+1}$; $S_n(1-r) = r-r^{n+1}$; $S_n = r(1-r^n)/1-r$ (for $|r|<1$). Note that $S_n \rightarrow {}^{r}/_{1-r}$ as $n \rightarrow \infty$.

## Assignment 3

Q: A *continuous random variable* X has pdf given by $f(x) = kx^2e^{-x^2}$ for $x > 0$, and $f(x) = 0$ for $x \leq 0$ ($k = {}^{4}/_{\sqrt{\pi}}$). Find the pdf h(y) of $Y = X^2$. **Show** that $h(y) = \Gamma(\theta, \alpha; y)$, and *find* θ and α. A: Let $y = x^2$. For $x > 0$, $y = x^2$ is a *strictly monotonically increasing* function. Further, $g'(x) = 2x$ is **continuous and** non-zero for $x > 0$. Therefore, we can <u>apply Theorem 6.1.</u>

$y = x^2 \Rightarrow x = \sqrt{(y)}$ (take the *+ve root* — which is OK since we are **dealing** with $x > 0$). Therefore, $h(y) = f(\sqrt{y})|{}^{d}/_{dy}(\sqrt{y})| = k(\sqrt{y})^2\exp(-(\sqrt{y})^2)|{}^{1}/_{2\sqrt{y}}| = {}^{4}/_{\sqrt{\pi}}y\exp(-y){}^{1}/_{2\sqrt{y}} = {}^{2\sqrt{(y)}\exp(-y)}/_{\sqrt{(\pi)}}$ $(y > 0)$. <u>*Conclusion*</u>: $h(y) = {}^{2\sqrt{(y)}\exp(-y)}/_{\sqrt{(\pi)}}$ for $y > 0$, *and* $h(y) = 0$ for $y \leq 0$. **Now** $\Gamma(\theta, \alpha; y) = \alpha^{\theta}y^{\theta-1}e^{-\alpha y}/\Gamma(\theta)$ for $y > 0$, and $\Gamma(\theta, \alpha; y) = 0$ for $y \leq 0$, where $\Gamma(\theta) = \int_0^{\infty} e^{-t}t^{\theta-1}dt$.

h(y) and $\Gamma(\theta, \alpha; y)$ look *very alike*. Comparing **exponentials**, we see that we need $\alpha = 1$. Comparing **powers of y,** we need $\theta = {}^{3}/_{2}$ to obtain a *match*. So $\Gamma({}^{3}/_{2}, 1; y) = y^{½}e^{-y}/\Gamma({}^{3}/_{2})$ for $y > 0$, and $\Gamma({}^{3}/_{2}, 1; y) = 0$ for $y \leq 0$. <u>**Conclusion**</u>: $\theta = {}^{3}/_{2}$, and $\alpha = 1$, so *that* $\Gamma({}^{3}/_{2}) = {}^{\sqrt{(\pi)}}/_{2}$. <u>**Note**</u>: to complete the *verification*, we need to show that $\Gamma({}^{3}/_{2}) = {}^{\sqrt{\pi}}/_{2}$. Now $\Gamma({}^{3}/_{2}) = \int_0^{\infty} e^{-t}t^{½}dt$ = (by *parts*) $= [-e^{-t}t^{½}]_0^{\infty} + \int_0^{\infty}e^{-t}\,½t^{-½}dt = 0 + ½\Gamma(½)$. It *can be shown* that $\Gamma(½) = \sqrt{\pi}$, so that $\Gamma({}^{3}/_{2}) = ½(\sqrt{\pi}) = {}^{\sqrt{\pi}}/_{2}$ as *required*. (**Direct** Method: $H(y) = P(Y \leq y) = P(X^2 \leq y) = P(|X| \leq \sqrt{y}) = P(-\sqrt{(y)} \leq X \leq \sqrt{(y)}) = F(\sqrt{y}) - F(-\sqrt{y})$. But $\int_{-\infty}^{x} f(x)dx = 0$ for $x \leq 0$, so that $H(y) = F(\sqrt{y})$. To conclude, $h(y) = H'(y) = {}^{1}/_{2\sqrt{y}}F'(\sqrt{y}) = {}^{1}/_{2\sqrt{y}}f(\sqrt{y}) = {}^{1}/_{2\sqrt{y}}k.y.e^{-y} = {}^{2}/_{\sqrt{\pi}}y^{½}e^{-y}$ for $y > 0$ (as *above*), $h(y) = 0$ for $y \leq 0$.

Q: A discrete r.v. X **assumes** the values $x_n = (-1)^n((k^n)/n)^{1/2}$, for $k > 1$, and n = 1, 2, 3,...; with pmf $p(x_n) = (k-1)/k^n$, for n = 1, 2, 3, ... Verify that $p(x_n)$ is indeed a *probability mass function.* Show that E[X] **exists**, but that $E[X^2]$ does **not** exist. A: As $k > 1$, then the **numerator** is always +ve. Any *power of a +ve number* is +ve, so that the **denominator** is always +ve, and we conclude *that* $p(x_n) \underline{>} 0$ because $k > 1$. **Now** is $\Sigma_{n=1}^{\infty} (k-1)/k^n = 1$? From *Calculus and Sequences and Series*, for a **geometric** series, $\Sigma_{n=1}^{\infty} ar^{n-1} = {}^{a}/_{1-r}$ (*provided* that $|r| < 1$) $\Rightarrow \Sigma_{n=1}^{\infty} ar^n = {}^{ra}/_{1-r}$ (*provided* that $|r| < 1$).

In our **case**, $r = {}^{1}/_{k}$ (which is *always* < 1 as k > 1), and $a = k-1$. **Therefore,** $\Sigma_{n=1}^{\infty} (k-1)({}^{1}/_{k})^n = [({}^{1}/_{k})(k-1)]/[1-{}^{1}/_{k}] = (1-{}^{1}/_{k})/(1-{}^{1}/_{k}) = 1$. <u>**QED**</u>. **Consider** $\Sigma_{n=1}^{\infty} |x_n|p(x_n)$. If *this* is $< \infty$, then E(X) exists. Now $\Sigma_{n=1}^{\infty} |x_n|p(x_n) = \Sigma_{n=1}^{\infty}|(-1)^n(k^n/n)^{½}|(k-1)/k^n = \Sigma_{n=1}^{\infty} (k^n/n)^{½}[(k-1)/k^n] = \Sigma_{n=1}^{\infty} k-1/\sqrt{(n)}k^{n/2}$. But as $k > 1$, and as $n \underline{>} 1$, then $k-1/n^{½}k^{n/2} \leq k-1/k^{n/2}$ for n = 1, 2, 3, ...., and $\Sigma_{k=1}^{\infty} k-1/k^{n/2} = (k-1)\Sigma_{n=1}^{\infty} (1/k^{½})^n = (k-1)\Sigma_{n=1}^{\infty} r^n$, *where* $r = 1/k^{½}$. Therefore, we have ($|r| < 1$ as $k > 1$) $(k-1)({}^{r}/_{1-r}) = (k-1).\{[1/k^{½}]/[1-(1/k^{½})]\} = (k-1)/k^{½}-1 = k^{½}+1 < \infty$. So E(X) *must exist*. Now <u>consider</u> $\Sigma_{n=1}^{\infty} |(x_n)^2|p(x_n)$. If *this* is $< \infty$, then $E(X^2)$ *exists*. Now $\Sigma_{n=1}^{\infty}|(x_n)^2|p(x_n) = \Sigma_{n=1}^{\infty}|(-1)^n(k^n/n)^{½}]^2|(k-1)/k^n = \Sigma_{n=1}^{\infty}|(-1)^{2n}(k^n/n)|(k-1)/k^n = \Sigma_{n=1}^{\infty} (k^n/n)(k-1/k^n) = \Sigma_{n-1}^{\infty} {}^{k-1}/_{n} = (k-1)\Sigma_{n=1}^{\infty} {}^{1}/_{n}$ (---(3)). But $\Sigma_{n=1}^{\infty} {}^{1}/_{n} = \infty$, so *series (3) is divergent*, and we conclude that $E(X^2)$ does **not** exist. <u>**QED**</u>.

Q: Determine whether or *not the mean of the Cauchy random variable* X exists. Hence discuss the existence or **otherwise** of the moments $\mu'_r$ of X for all $r > 1$. A: If X is a Cauchy r.v., then $f(x) = {}^1/_\pi {}^1/_{1+x^2}$ (---(1)) for $-\infty < x < \infty$ (*simple form*). E(X) exists **provided that** $\int_{-\infty}^{\infty} |x|f(x)dx < \infty$ (---(2)). (1) $\Rightarrow \int_{-\infty}^{\infty} |x|f(x)dx$ = (*symmetry*) = $\lim_{\beta\to\infty} {}^2/_\pi \int_0^\beta {}^x/_{1+x^2}dx = ({}^1/_\pi)\lim_{\beta\to\infty} \int_0^\beta {}^{2xdx}/_{1+x^2} = {}^1/_\pi \lim_{\beta\to\infty}[\ln(1+x^2)]^\beta_0 = {}^1/_\pi . \lim_{\beta\to\infty}(\ln(1+\beta^2)) = \infty$ (---(3)). (2) and (3) $\Rightarrow$ E(X) *does not exist*, i.e. $\mu_1'$ of X does not exist. **Hence**, from Theorem 2.1 of Section 3.2, the $\mu'_r$ of X for $r > 1$ *do not exist*.

22nd March 2001

## 3.6: Moment Generating Functions (MGF's) (3.4 and 3.5 are Omitted)

Let X be a *1-dim. r.v.*, then the MGF of X is defined to be the *expected value* of $e^{tx}$, if it **exists**, for real t, where $|t| < h$ (an *open interval* containing t). It is denoted by $M_X(t)$, so that $M_X(t) = E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx}dF(x) = \Sigma_{\text{all } x_n} e^{tx_n}p(x_n)$ for X *discrete*, and $E(e^{tx}) = \int_{-\infty}^{\infty} e^{tx}f(x)dx$ for X *continuous*. **Example**: *The Binomial Distribution*, with parameters n and p, has pmf $p(x) = \binom{n}{x}p^x(1-p)^{n-x}$, for x = 0, 1, 2, ..., n. **MGF**: $M_X(t) = E[e^{tx}] = \Sigma_{x=0}^{x=n} e^{tx}\binom{n}{x}p^x(1-p)^{n-x} = \Sigma_{x=0}^{n} \binom{n}{x}(pe^t)^x(1-p)^{n-x} = (pe^t+(1-p))^n$ (using the *binomial* expansion $(a+b)^n = \Sigma_{x=0}^{n} \binom{n}{x}a^xb^{n-x}$).

**Theorem 1:** Let X be a r.v. where the *MGF* exists. Then $E(X^r) = M_X^{(r)}(0)$ (r = 1, 2, 3, ...), and $M_X(t) = 1 + \Sigma_{r=1}^{\infty} (t^r/r!)E(X^r)$. **Proof:** $(d^r/dt^r)(M_X(t)) = \int_{-\infty}^{\infty}(\partial^r/\partial t^r)(e^{tx})dF(x) = \int_{-\infty}^{\infty} x^re^{tx}dF(x)$ = (t = 0) = $\int_{-\infty}^{\infty} x^rdF(x) = E(X^r)$. So $M_X^{(r)}(0) = E(X^r)$ (r = 1, 2, ...). By *Maclaurin's* expansion of $M_X(t)$, $M_X(t) = 1 + \Sigma_{r=1}^{\infty} (t^r/r!)M_X^{(r)}(0)$ (*Using* $f(x) = 1+\Sigma_{r=1}^{\infty} (x^r/r!)f^{(r)}(0)$). Therefore, $M_X(t) = 1 + \Sigma_{r=1}^{\infty}(t^r/r!)E(X^r)$ as *required*.

**Example**: Calculate μ and σ² for the *binomial distribution* with parameters n and p, using the MGF. MGF: $M_X(t) = (pe^t+(1-p))^n$. $\mu = E(X) = M_X^{(1)}(t)|_{t=0} = {}^d/_{dt}[(pe^t+(1-p))^n]|_{t=0} = n(pe^t+(1-p))^{n-1}pe^t|_{t=0} = n(p+(1-p))^{n-1}p = np$. **QED**. Further, $E(X^2) = M_X^{(2)}(t)|_{t=0} = (n-1)n(pe^t+(1-p))^{n-2}(pe^t)^2 + n(pe^t+(1-p))^{n-1}pe^t|_{t=0} = (n-1)np^2+np = n^2p^2-np^2+np$. Therefore, $\sigma^2 = E(X^2)-(E(X))^2 = n^2p^2-np^2+np-(np)^2 = np(1-p) = npq$. **QED**.

## Tutorial

Q: Find the MGF of the *uniform (or rectangular)* distribution in [a, b]. A: $f(x) = {}^1/_{b-a}$ for $a \le x \le b$, and $f(x) = 0$ *otherwise*. Therefore, $M_X(t) = E[e^{tx}] = \int_a^b e^{tx}/b-a\ dx = [e^{tx}/t(b-a)]^b_a = {}^1/_{t(b-a)}(e^{bt}-e^{at})$. Q: Find the *MGF* of $X \sim N(0, 1^2)$. A: $f(x) = {}^1/_{\sqrt{(2\pi)}}e^{-½x^2}$ $(-\infty < x < \infty)$. Therefore, $M_X(t) = E[e^{tx}] = \int_{-\infty}^{\infty} {}^1/_{\sqrt{(2\pi)}}e^{tx}e^{-½x^2}dx = \int_{-\infty}^{\infty} {}^1/_{\sqrt{(2\pi)}} e^{tx-½x^2}dx = {}^1/_{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-½(x^2-2tx)}dx = {}^1/_{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-½[(x-t)^2-t^2]}dx = {}^1/_{\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{½t^2}e^{-½(x-t)^2}dx = {}^1/_{\sqrt{(2\pi)}}e^{½t^2}\int_{-\infty}^{\infty} e^{-½y^2}dy$ (*Letting* $y = x-t$) $= [e^{½t^2}/\sqrt{(2\pi)}][\sqrt{2\pi}] = e^{½t^2}$.

## 3.7: Determination of a Distribution by its Moments

A distribution may be *uniquely determined by its moments*. **Theorem 1:** The Moment Problem. If the *moments* $\mu'_n$ of a r.v. exist for $n = 1, 2, 3, \ldots$, and $\Sigma_{n=1}^{\infty} (\mu'_n/n!)r^n$ *is absolutely convergent* for some $r > 0$, then the **sequence** $\{\mu'_n\}$ of moments *uniquely determine* the cdf of X. **Proof**: omitted. **Corollary**: If a r.v. X is bounded from *both* sides (or if its range is *finite*), then **all** of its moments $\mu'_n$ ($n = 1, 2, 3, \ldots$) exist, and the *sequence* $\{\mu'_n\}$ determines uniquely the cdf of X. **Proof**: If X is *bounded*, then $\exists$ a, b ($a < b < \infty$) s.t. $f(a) = 0$, and $f(b) = 1$, where $f(x)$ is the *cdf* of X. Let $M = \max(|a|, |b|)$. Then $|\mu'_n| = |\int_a^b x^n dF(x)| \leq \int_a^b |x|^n dF(x) \leq M^n \int_a^b dF(x) = M^n < \infty$ $\forall$ n. It follows that $\mu'_n$ *exists* for every n. **Therefore,** $\Sigma_{n=1}^{\infty} |(\mu'_n/n!)r^n| \leq \Sigma_{n=1}^{\infty} |Mr|^n/r! = e^{|Mr|}-1 < \infty$ $\forall$ r, which *completes* the proof.

27th March 2001

## 3.8: Expected Values of n-Dimensional Random Variables

Let $g(X_1, X_2, \ldots, X_n)$ be a *real valued function* of an n-dimensional r.v. $\underline{X} = (X_1, X_2, \ldots, X_n)$ (g *itself* is a 1-dimensional r.v.). We define the expected value of g as $E(g(X_1, X_2, \ldots, X_n)) = \int\int\ldots_{-\infty}^{\infty}\ldots\int g(x_1,x_2,\ldots,x_n)dF(x_1,x_2,\ldots,x_n)$ (with *n integral signs*), provided that the integral is **absolutely convergent**. (**Note**: $E(g(X_1, X_2, \ldots, X_n)) = \Sigma_{i1}\Sigma_{i2}\ldots\Sigma_{in}\, g(x_{i1}, x_{i2}, \ldots, x_{in})p(x_{i1}, x_{i2}, \ldots, x_{in})$ if $\underline{X}$ is *discrete*, and $E(g(X_1, X_2, \ldots, X_n)) = \int\int\ldots_{-\infty}^{\infty}\ldots\int g(x_1, x_2, \ldots, x_n)f(x_1, x_2, \ldots, x_n)dx_1\ldots dx_n$ if $\underline{X}$ is *continuous*).

**Theorem 1**: For any *random* variables $X_1, \ldots, X_n$, we have $E(\Sigma_{i=1}^{n} a_i g_i(X_i)) = \Sigma_{i=1}^{n} a_i E(g_i(X_i))$, where each $g_i$ is a *real 1-dimensional function* of the $X_i$, each $a_i$ is a constant, and $i = 1, 2, \ldots, n$. **Proof:** $E(\Sigma_i a_i g_i) = \int\int\ldots_{-\infty}^{\infty}\ldots\int (a_1g_1(x_1) + \ldots + a_ng_n(x_n))dF(x_1, x_2, \ldots, x_n) = \int\int\ldots_{-\infty}^{\infty}\ldots\int a_1g_1(x_1)dF(x_1, x_2, \ldots, x_n) + \int\int\ldots_{-\infty}^{\infty}\ldots\int a_2g_2(x_2)dF(x_1, x_2, \ldots, x_n) + \ldots + \int\int\ldots_{-\infty}^{\infty}\ldots\int a_ng_n(x_n)dF(x_1, x_2, \ldots, x_n) = \int_{x1=-\infty}^{\infty} a_1g_1(x_1)\int_{x2=-\infty}^{\infty}\ldots \int_{xn=-\infty}^{\infty} dF(x_1, x_2, \ldots, x_n) + \ldots + \int_{xn=-\infty}^{\infty} a_ng_n(x_n)\int_{x1=-\infty}^{\infty}\ldots \int_{xn-1=-\infty}^{\infty} dF(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{\infty} a_1g_1(x_1)dF_1(x_1) + \ldots + \int_{-\infty}^{\infty} a_ng_n(x_n)dF_n(x_n)$ (*where* $F_i(x_i)$ is the **marginal** cdf of $X_i$) = (Section 2.5) = $\Sigma_{i=1}^{n} a_iE(g_i(X_i))$. (The red bit represents *n-1 integral* signs).

**Corollary**: $E(\Sigma_i a_iX_i) = \Sigma_i a_iE(X_i)$. **Theorem 2**: If r.v.'s $X_1, X_2, \ldots, X_n$ are *independent*, then $E[\Pi_{i=1}^{n}(g_i(X_i))] = \Pi_{i=1}^{n} E(g_i(X_i))$, where the $g_i$ are *real valued functions* of the $X_i$. **Proof:** $E(\Pi_{i=1}^{n} g_i) = \int\int\ldots_{-\infty}^{\infty}\ldots\int g_1(x_1)g_2(x_2)\ldots g_n(x_n)dF(x_1, x_2, \ldots, x_n)$ = (by *independence*) = $\int\int\ldots_{-\infty}^{\infty}\ldots\int g_1(x_1)\ldots g_n(x_n)dF_1(x_1)\ldots dF_n(x_n) = \int_{x1=-\infty}^{\infty} g_1(x_1)dF_1(x_1)\int_{x2=-\infty}^{\infty} g_2(x_2)dF_2(x_2)\ldots \int_{xn=-\infty}^{\infty} g_n(x_n)dF_n(x_n) = \Pi_{i=1}^{n} E(g_i(X_i))$. **Corollary**: $E(\Pi_{i=1}^{n} (X_i)) = \Pi_{i=1}^{n}(E(X_i))$ if the $X_i$ are *independent*. **Note**: the converse is *false*.

## 3.8.1: Moments of 2-Dimensional Random Variables

The $(r+s)^{th}$-order *moment* of (X, Y) about the **origin** is given by $\mu'_{r,s} = E(X^rY^s) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^ry^s dF(x,y)$. **Order 1**: $\mu'_{1,0} = \mu_X = E(X)$; and $\mu'_{0,1} = \mu_Y = E(Y)$. **Order 2**: $\mu'_{2,0} = E(X^2)$; $\mu'_{0,2} = E(Y^2)$; and $\mu'_{1,1} = E(XY)$. **Now** $E[(X-\mu_X)^r(Y-\mu_Y)^s]$ is the $(r+s)^{th}$-order *central moment* of (X,Y), **denoted** by $\mu_{r,s}$. Now $\mu_{1,0} = 0 = \mu_{0,1}$; $\mu_{2,0} = \sigma^2_X = var(X)$; $\mu_{0,2} = \sigma^2_Y = var(Y)$; and $\mu_{1,1} = E[(X-\mu_X)(Y-\mu_Y)] = cov(X, Y)$, the *covariance* of X, Y. **Clearly**, $cov(X, Y) = E(XY)-E(X)E(Y)$.

**Theorem 3**: If X and Y are *independent*, then cov(X, Y) = 0. **Proof:** X and Y *independent* $\Rightarrow$ (by a *corollary* of Theorem 2) E(XY) = E(X)E(Y) $\Rightarrow$ cov(X, Y) = 0. **QED**. **Note**: The *converse is* not true, i.e. if cov(X, Y) = 0, this does **not** imply that X and Y are *independent*. **Definition**: The *correlation coefficient* between X and Y is *defined* as $\rho = \text{cov}(X,Y)/\sigma_X\sigma_Y$.

23rd March 2001

**Theorem 4**: If X and Y are *independent,* then $\rho = 0$. **Proof:** X and Y independent $\Rightarrow$ (by *Theorem 3*) cov(X, Y) = 0 $\Rightarrow$ (by *definition*) $\rho = 0$. **Note**: The *converse* is false — see the following tutorial for an *example*. **Note**: if $\rho = 0$, then X and Y are termed *un-correlated*. **Theorem 5**: $|\rho| \le 1$. **Proof**: $0 \le E[((X-\mu_X)t+(Y-\mu_Y))^2] = \sigma^2_X t^2 + 2\text{cov}(X,Y)t + \sigma^2_Y \Rightarrow$ (*using* "$b^2-4ac \le 0$") $(2\text{cov}(X,Y))^2 - 4\sigma_X^2\sigma_Y^2 \le 0 \Rightarrow (\text{cov}(X,Y))^2/\sigma_X^2\sigma_Y^2 \le 1 \Rightarrow \rho^2 \le 1 \Rightarrow |\rho| \le 1$.

**Theorem 6**: If X and Y are *exactly linearly related,* i.e. Y = aX+b, with $a \ne 0$, then $\rho^2 = 1$. *Moreover*, if a > 0, then $\rho = 1$; and if a < 0, then $\rho = -1$. *Conversely*, if $\rho = \pm 1$, then X and Y are exactly linearly related, with *probability* 1. **Proof:** If Y = aX+b, then cov(X,Y) = (*by definition*) $= E[(X-\mu_X)(Y-\mu_Y)]$ = (Y = aX+b, $\mu_Y$ = E(Y)) $= E[(X-\mu_X)(aX+b-E(aX+b))] = E[(X-\mu_X)(aX+b-(a\mu_X+b))] = aE[(X-\mu_X)^2] = a\sigma_X^2$. And $\sigma_Y^2 = a^2\sigma_X^2 \Rightarrow \sigma_Y = |a|\sigma_X$. Therefore, $\rho = \text{cov}(X,Y)/\sigma_X\sigma_Y = a\sigma_X^2/\sigma_X|a|\sigma_X = \pm 1$ (or $\rho^2 = 1$).

If a < 0, then $\rho = -1$; if a > 0, then $\rho = 1$. *Converse:* Assume that $\rho^2 = 1$. Let $Z = (X-\mu_X)t + (Y-\mu_Y)$ ($\Rightarrow E(Z) = 0$). Now, according to the *proof* of Theorem 5, $\exists$ at least **one** value of t, say $t_0$, s.t. $E(Z^2) = 0$ (*because* if $E(Z^2) > 0\ \forall$ t, then $\rho^2 < 1$). It follows that $E(Z) = 0 \Rightarrow E(Z^2) = 0 = \text{var}(Z)$ for $t = t_0 \Rightarrow P(Z = E(Z) = 0) = 1$. *Therefore*, $(X-\mu_X)t_0 + (Y-\mu_Y) = 0 \Rightarrow Y = aX+b$.

**Note**: $\rho$ is a measure of the *degree of linear association* between two random variables. Values of $\rho$ **close** to $\pm 1$ indicate a *high degree of linear association*. If $\rho = 1$, then the linear association is **exact**, and the variables vary in the *same* direction (they increase and decrease together). If $\rho = -1$, then the linear association is **exact**, but the variables vary in the *opposite direction* (when one increases, the other decreases).

In these **extreme** cases, the whole *distribution* of (X, Y) is contained in the line Y = aX+b. (All **pairs** (x, y), with $p(x, y) \ne 0$, lie along a *straight* line). $\rho = 0$ implies the absence of **linear** association, but in such cases, a **non linear** association between the variables *may* exist — see the following tutorial for an example.

## 3.8.2: The Covariance Matrix

Let $\sigma_{ij} = E[(X_i-\mu_i)(X_j-\mu_j)]$ for i, j = 1, ..., n. Now $\sigma_{ij} = \text{var}(X_i)$ if i = j, and $\sigma_{ij} = \text{cov}(X_i, X_j)$ if $i \ne j$. The n×n *symmetric matrix of all variances* and **covariances** between $X_1$, $X_2$, ..., $X_n$ is called their *covariance matrix*, and is denoted by $\|\sigma_{ij}\|$ for i, j = 1, .., n. In the matrix, *variances* are along the diagonal, and the *covariances* are on the off-diagonal. We denote the **determinant** by $|\sigma_{ij}|$.

$$= \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1n} \\ \dots & & \dots \\ \sigma_{n1} & \dots & \sigma_{nn} \end{bmatrix}$$

The *inverse matrix* $\|\sigma_{ij}\|^{-1}$ exists if $\|\sigma_{ij}\|$ is *non-singular*, i.e. if $|\sigma_{ij}| \neq 0$. In this *case*, we have $|\,\|\sigma_{ij}\|^{-1}\,| = |\,\|\sigma_{ij}\|\,|^{-1}$. **Example:** (*restatement* of Theorem 6): **Two** r.v.'s $X_1$ and $X_2$ are *exactly linearly related* iff their covariance matrix is singular, i.e. iff $|\sigma_{ij}| = 0$. **Proof:** The condition is *equivalent* to $\rho^2 = 1$ — see the *tutorial* below.

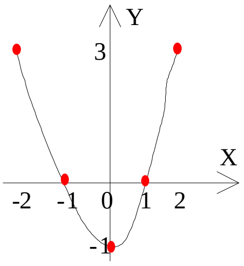
## Tutorial

**Example 1:** $\rho = 0$ does not imply that X and Y are independent. Consider the table shown, where $p_2(y)$ and $p_1(x)$ are the *row and column* totals respectively. X and Y independent means that $p(x, y) = p_1(x)p_2(y)\ \forall\ x$ and y. But *here,* $p(-1,0) = {}^1/_3 \neq p_1(-1)p_2(0) = {}^1/_3 \times {}^2/_3 = {}^2/_9$. Now $\rho = \mathrm{cov}(X,Y)/\sigma_X\sigma_Y$, *where* $\mathrm{cov}(X,Y) = E(XY)-E(X)E(Y)$. Here, $E(X) = \Sigma_{x=-1}^{x=1}\, x p_1(x) = 0$; $E(Y) = \Sigma_{y=0}^{1}\, y p_2(y) = {}^1/_3$; and $E(XY) = \Sigma_{x=-1}^{x=1}\, \Sigma_{y=0}^{1}\, xyp(x,y) = 0$. Therefore, $\mathrm{cov}(X,Y) = 0-({}^1/_3\times 0) = 0 \Rightarrow \rho = 0$ — but X and Y are **NOT** independent.

| y\x | -1 | 0 | 1 | $p_2(y)$ |
|---|---|---|---|---|
| 0 | 1/3 | 0 | 1/3 | 2/3 |
| 1 | 0 | 1/3 | 0 | 1/3 |
| $p_1(x)$ | 1/3 | 1/3 | 1/3 | 1 |

| y\x | -2 | -1 | 0 | 1 | 2 | $p_2(y)$ |
|---|---|---|---|---|---|---|
| 3 | 1/6 | 0 | 0 | 0 | 1/6 | 1/3 |
| 0 | 0 | 1/ 12 | 0 | 1/ 12 | 0 | 1/6 |
| -1 | 0 | 0 | 1/2 | 0 | 0 | 1/2 |
| $p_1(x)$ | 1/6 | 1/ 12 | 1/2 | 1/ 12 | 1/6 | 1 |

**Exercise 2:** Let X and Y have the shown *bi-variate* distribution. Find $\rho$ and the *functional relation* between X and Y. $E(X) = \Sigma_{x=-2}^{2}\, x p_1(x) = 0$; $E(Y) = \Sigma_{y=-1}^{3}\, y p_2(y) = ½$; and $E(XY) = \Sigma_{x=-2}^{2}\Sigma_{y=-1}^{3}\, xyp(x,y) = 0$. Therefore, $\mathrm{cov}(X,Y) = 0-(½\times 0) = 0 \Rightarrow \rho = 0$. But from the **graph**, even *though* that $\rho = 0$, there **is** a relationship ($Y = X^2-1$) *between* X and Y.



**Exercise 3:** Prove that $\rho^2 = 1 \Leftrightarrow |\sigma_{ij}| = 0$ (*for* $n = 2$). Now $\|\sigma_{ij}\| = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$, and $|\sigma_{ij}| = \sigma_{11}\sigma_{22}-\sigma_{21}\sigma_{12}$. *Now* $\mathrm{cov}(X_1,X_2) = \sigma_{12} = \sigma_{21}$. *Therefore,* $\sigma_{11}\sigma_{22}-\sigma_{21}\sigma_{12} = \sigma_{X1}^2\sigma_{X2}^2-(\mathrm{cov}(X_1,X_2))^2$. If this is *zero,* then $\sigma_{X_1}^2\sigma_{X_2}^2/(\mathrm{cov}(X_1,X_2))^2 = 1$, so that $\rho^2 = 1$. **QED**.

3rd April 2001

## 3.8.3: The n-Dimensional Normal Distribution

The n-dimensional normal distribution is **denoted** by $N(\underline{\mu}, \|\sigma_{ij}\|)$. Now $\underline{x}^T = (X_1, X_2, \ldots, X_n)$ has *pdf* $f(x_1, x_2, \ldots, x_n) = [1/(2\pi)^{n/2}\sqrt{(|\,\|\sigma_{ij}\|\,|)}]\times\exp\{-½(\underline{x}-\underline{\mu})^T(\|\sigma_{ij}\|)^{-1}(\underline{x}-\underline{\mu})]$, *where* $\underline{x} = (x_1\ x_2\ \ldots\ x_n)^T$, $\underline{\mu} = (\mu_1\ \mu_2 \ldots \mu_n)^T$, and $\|\sigma_{ij}\|$ is the *(positive definite)* **covariance** matrix ($\Rightarrow |\,\|\sigma_{ij}\|\,| \neq 0$). **Example**: Let $(X_1, X_2, \ldots, X_n)$ have a *multivariate normal distribution*, with **positive definite** covariance matrix. Show that the *random variables* $X_i$ are independent iff $\|\sigma_{ij}\|$ is a *diagonal* matrix.

**Proof**: (Only If): If the $X_i$ are *independent*, then $\mathrm{cov}(X_i, X_j) = 0\ \forall\ i, j\ (i \neq j)$ by Theorem 3 $\Rightarrow \sigma_{ij} = 0$ for $i \neq j \Rightarrow \|\sigma_{ij}\|$ is *diagonal*. (If): If $\|\sigma_{ij}\|$ is *diagonal*, then **so** is $\|\sigma_{ij}\|^{-1}$, so that the *pdf* $f(x_1, x_2, \ldots, x_n)$ factorises into a *product of n 1-dimensioal pdf's,* which implies that the cdf $F(x_1, \ldots, x_n)$ is given by $\Pi_{i=1}^{n} F_i(x_i)$, which implies that the $X_i$ are *independent.*

# Assignment 4

The *2-dimensional r.v.* $\underline{X}^T = (X_1, X_2)$ has the pdf of the **bivariate** normal distribution, $f(\underline{x}^T) = f(x_1, x_2) = [1/(2\pi)\sqrt{|\,\|\sigma_{ij}\|\,|}]\exp\{-\tfrac{1}{2}(\underline{x}-\underline{\mu})^T(\|\sigma_{ij}\|)^{-1}(\underline{x}-\underline{\mu})\}$ (---(1)), where $\underline{x}^T = (x_1\ x_2)$, $\underline{\mu}^T = (\mu_1\ \mu_2)$, and $\|\sigma_{ij}\| = \begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}$, with $|\,\|\sigma_{ij}\|\,| > 0$, and $\sigma_{21} = \sigma_{12}$ (*symmetric*). (a) State a **necessary** and **sufficient** condition on $\|\sigma_{ij}\|$ for the *independence* of $X_1$ and $X_2$. (b) If $f(\underline{x}^T) = C\exp[-\tfrac{1}{2}(3x_1^2+2x_1x_2+3x_2^2)]$ (---(2)), find C, $\underline{\mu}$, and $\|\sigma_{ij}\|$; and determine the *independence* of otherwise of the $X_1$ with the $X_2$. (c) If $\underline{y}^T = (Y_1, Y_2) = (X_1+X_2, -X_1+X_2)$, find the *pdf* $h(\underline{y}^T)$, and *show* that $Y_1$ and $Y_2$ are *independent*.

A: (a) $\|\sigma_{ij}\|$ *diagonal* $\Leftrightarrow$ $X_1$ and $X_2$ independent. (b) $3x_1^2+2x_1x_2+3x_2^2 = (x_1\ x_2)\begin{pmatrix}3 & 1\\ 1 & 3\end{pmatrix}(x_1\ x_2)^T$ (---(3)). (1), (2) and (3) $\Rightarrow$ $(\|\sigma_{ij}\|)^{-1} = \begin{pmatrix}3 & 1\\ 1 & 3\end{pmatrix}$ (---(4)). (4) $\Rightarrow$ $\|\sigma_{ij}\| = \tfrac{1}{8}\begin{pmatrix}3 & -1\\ -1 & 3\end{pmatrix}$ (---(5), the covariance matrix) (*using* $\begin{pmatrix}a & b\\ c & d\end{pmatrix}^{-1} = \tfrac{1}{ad-bc}\begin{pmatrix}d & -b\\ -c & a\end{pmatrix}$). (5) $\Rightarrow$ $|\,\|\sigma_{ij}\|\,| = \tfrac{1}{8}$ (---(6)). *Comparing* (2) with (1), and **using** (5) and (6), $C = 1/(2\pi)\sqrt{(\|\sigma_{ij}\|)} = \tfrac{1}{2\pi}\tfrac{1}{\sqrt{(1/8)}} = \tfrac{\sqrt{(2)}}{\pi}$ (---(7)). Now we can also say that $\underline{\mu} = \underline{0}$ (---(8)) (there are no *linear* terms in $x_1$ and $x_2$) — therefore, (5) $\Rightarrow$ $\|\sigma_{ij}\|$ is **not** diagonal $\Rightarrow$ $X_1$ and $X_2$ are not *independent*, i.e. they are dependent.

(c) $y_1, y_2 = (x_1+x_2, -x_1+x_2)$ (---(9)) $\Rightarrow$ $(x_1, x_2) = (\tfrac{1}{2}(y_1-y_2), \tfrac{1}{2}(y_1+y_2))$ (---(10)). Now the pdf of $\underline{Y}^T$ is given by $h(\underline{y}^T) = f(\underline{x})^T|J|$ (---(11)), *where* $J = \left|\tfrac{\partial(x_1,x_2)}{\partial(y_1,y_2)}\right| = \begin{vmatrix}\partial x_1/\partial y_1 & \partial x_1/\partial y_2\\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2\end{vmatrix}$ (---(12)). Now (10) and (12) $\Rightarrow$ $J = \begin{vmatrix}\tfrac{1}{2} & -\tfrac{1}{2}\\ \tfrac{1}{2} & \tfrac{1}{2}\end{vmatrix} = +\tfrac{1}{2}$ (---(13)). So, from *part* (b), if $f(\underline{x}^T) = \tfrac{\sqrt{(2)}}{\pi}\exp\{-\tfrac{1}{2}(3x_1^2+2x_1x_2+3x_2^2)\}$, then $h(\underline{y}^T) = \tfrac{\sqrt{(2)}}{\pi}\exp\{3x_1^2+2x_1x_2+3x^2\}(\tfrac{1}{2})$. *Substituting* for $x_1$ and $x_2$, $h(\underline{y}^T) = \tfrac{\sqrt{(2)}}{2\pi}\exp\{-\tfrac{1}{2}[3(\tfrac{1}{2}(y_1-y_2))^2 + 2(\tfrac{1}{2}(y_1-y_2)\tfrac{1}{2}(y_1+y_2)) + 3(\tfrac{1}{2}(y_1+y_2))^2]\} = \tfrac{\sqrt{(2)}}{2\pi}\exp\{-\tfrac{1}{2}[\tfrac{3}{4}(y_1^2-2y_1y_2+y_2^2) + \tfrac{1}{2}(y_1^2-y_2^2) + \tfrac{3}{4}(y_1^2+2y_1y_2+y_2^2)]\} = \tfrac{1}{2\pi\sqrt{\frac{1}{2}}}\exp\{-\tfrac{1}{2}[(\tfrac{3}{4}+\tfrac{1}{2}+\tfrac{3}{4})y_1^2 + (-\tfrac{3}{2}+\tfrac{3}{2})y_1y_2 + (\tfrac{3}{4}-\tfrac{1}{2}+\tfrac{3}{4})y_2^2]\} = \tfrac{1}{2\pi\sqrt{\frac{1}{2}}}\exp\{-\tfrac{1}{2}(2y_1^2+y_2^2)\} = h(\underline{y}^T)$.

Now the *above* implies that $h(\underline{y}^T) = \tfrac{1}{2\pi\sqrt{(\frac{1}{2})}}\exp[-\tfrac{1}{2}(y_1, y_2)\begin{pmatrix}2 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}y_1\\ y_2\end{pmatrix}]$ (the red bit is $|\,\|\sigma_{ij}\|\,|$, and the blue bit is $(\|\sigma_{ij}\|)^{-1}$, *comparing* to (1)). **So** $(\|\sigma_{ij}\|)^{-1} = \begin{pmatrix}2 & 0\\ 0 & 1\end{pmatrix}$ $\Rightarrow$ $\|\sigma_{ij}\| = \tfrac{1}{2}\begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}$, which **is** diagonal, *so that* $Y_1$ and $Y_2$ **are** *independent*. **QED**.

Now **check that** $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(\underline{y}^T)d\underline{y} = 1$. Let $I = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tfrac{1}{2\pi\sqrt{(\frac{1}{2})}}\exp\{-\tfrac{1}{2}(2y_1^2+y_2^2)\}dy_1dy_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \tfrac{\sqrt{(2)}}{2\pi}\exp(-y_1^2)\exp(-y_2^2/2)dy_1dy_2 = \tfrac{\sqrt{(2)}}{2\pi}\int_{-\infty}^{\infty}\exp(-y_2^2/2)[\int_{-\infty}^{\infty}\exp(-y_1^2)dy_1]dy_2$. **Now** as $\int_{-\infty}^{\infty}\exp(-x^2)dx = \sqrt{\pi}$, then $I = \tfrac{\sqrt{(2)}}{2\pi}\int_{-\infty}^{\infty}\exp(-y_2^2/2)[\sqrt{\pi}]dy_2 = \tfrac{\sqrt{(2\pi)}}{2\pi}\int_{-\infty}^{\infty}\exp(-y_2^2/2)dy_2 = \tfrac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty}\exp(-y_2^2/2)dy_2$. **Let** $z = y_2/\sqrt{(2)}$, so that $z^2 = y_2^2/2$, and $dz = dy_2/\sqrt{(2)}$. Therefore, $I = \tfrac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty}\exp(-z^2)dz(\sqrt{(2)}) = \tfrac{\sqrt{(2)}}{\sqrt{(2\pi)}}\int_{-\infty}^{\infty}\exp(-z^2)dz = \tfrac{\sqrt{(2)}}{\sqrt{(2\pi)}}[\sqrt{(\pi)}] = 1$. **QED**.

5th April 2001

## 3.10: The Central Limit Theorem (3.9 is Omitted)

If $X_1, X_2, \ldots, X_n$ are *independent r.v.'s* with the **same** distribution (of say X), with mean $\mu$ and variance $\sigma^2$; and if $\bar{X} = \tfrac{1}{n}\sum_{i=1}^{n} X_i$, then $Z = \bar{X} - \mu/(\sigma/\sqrt{n})$, the *standardised* form of $\bar{X}$, is *approximately* distributed as $N(0, 1^2)$ (a **standard** normal r.v.) for *large* values of n. **Notes**: (1) The MGF of $N(0, 1^2)$ is $M_X(\theta) = e^{\theta^2/2}$.

(2) We will *show* that, assuming the distributions have MGF's, then $M_Z(\theta) \to e^{\theta^2/2}$ as $n \to \infty$, i.e. that the MGF of Z *converges* to the MGF of $N(0, 1^2)$. This *assumes* that if the MGF's converge, then so do the **distributions** (which *can* be proved).

**<u>Proof:</u>** $M_Z(\theta) = E(e^{\theta z}) = E(\exp(\theta(\bar{X}-\mu/(\sigma/\sqrt{n}))) = E(\exp(\theta\bar{X}\sqrt{n}/\sigma)\exp(-{}^{\theta\mu\sqrt{n}}/_{\sigma})) = \exp(-{}^{\theta\mu\sqrt{n}}/_{\sigma})E(\exp({}^{\theta\sqrt{n}}/_{\sigma}{}^{1}/_{n}\Sigma_i X_i)) = \exp(-{}^{\theta\mu\sqrt{n}}/_{\sigma})E(\exp({}^{\theta}/_{\sigma\sqrt{n}}\Sigma_i X_i)) = \exp(-{}^{\theta\mu\sqrt{n}}/_{\sigma})E(\exp({}^{\theta}/_{\sigma\sqrt{n}} X_i)\times\ldots\times\exp({}^{\theta}/_{\sigma\sqrt{n}} X_n))$ = (by *Theorem 2 of Section 3.8* — the $X_i$'s are **independent**) $= \exp(-{}^{\theta\mu\sqrt{n}}/_{\sigma})E(\exp({}^{\theta}/_{\sqrt{(n)}\sigma} X_i))\ldots.E(\exp({}^{\theta}/_{\sigma\sqrt{n}}X_n)) = \exp(-{}^{\theta\mu\sqrt{n}}/_{\sigma})M_{X_1}({}^{\theta}/_{\sigma\sqrt{n}})\ldots M_{X_n}({}^{\theta}/_{\sigma\sqrt{n}}) = \exp(-{}^{\theta\mu\sqrt{n}}/_{\sigma})(M_X({}^{\theta}/_{\sigma\sqrt{n}}))^n$ (since the $X_i$'s follow the *same distribution* — which is, say, X).

Now taking **logs** on both sides, $\ln(M_Z(\theta)) = -{}^{\theta\mu\sqrt{(n)}}/_{\sigma} + n\ln(M_X({}^{\theta}/_{\sigma\sqrt{n}}))$ = (by *Theorem 1 of Section 3.6*) $= -{}^{\theta\sqrt{(n)}\mu}/_{\sigma} + n\ln[1 + {}^{\theta}/_{\sigma\sqrt{(n)}}E(X) + ({}^{\theta}/_{\sigma\sqrt{n}})^2\ \tfrac{1}{2}E(X^2)$ + terms in $(1/n^{3/2})]$. Now *using* $\ln(1+x) = x-\tfrac{1}{2}x^2+{}^{1}/_{3}x^3-\ldots$, we obtain $-{}^{\theta\mu\sqrt{(n)}}/_{\sigma} + n\{({}^{\theta}/_{\sigma\sqrt{(n)}}E(X) + ({}^{\theta}/_{\sigma\sqrt{(n)}})^2\ \tfrac{1}{2}E(X^2) + \ldots) - \tfrac{1}{2}(\ldots)^2 + {}^{1}/_{3}(\ldots)^3 - + - \ldots\} = -{}^{\theta\mu\sqrt{(n)}}/_{\sigma} + n\{{}^{\theta}/_{\sigma\sqrt{n}}E(X) + ({}^{\theta^2}/_{\sigma^2 n})(\tfrac{1}{2}E(X^2)) - \tfrac{1}{2}{}^{\theta^2(E(X))^2}/_{\sigma^2 n}$ + *terms* in $(1/n^{3/2})\}$.

*Now* $E(X) = \mu$, and $E(X^2)-(E(X))^2 = \sigma^2$, so that we can write $-{}^{\theta\mu\sqrt{n}}/_{\sigma} + {}^{\sqrt{(n)}\theta\mu}/_{\sigma} + {}^{\theta^2}/_{2\sigma^2}(E(X^2) - (E(X))^2)$ + n(terms in $(1/n^{3/2})$) $= -{}^{\theta\mu\sqrt{n}}/_{\sigma} + {}^{\sqrt{(n)}\theta\mu}/_{\sigma} + {}^{\theta^2\sigma^2}/_{2\sigma^2}$ + terms in $1/n^{1/2} = {}^{\theta^2}/_{2}$ + *terms* in $1/n^{1/2}$. Therefore, $\ln(M_Z(\theta)) \to {}^{\theta^2}/_{2}$ as $n \to \infty$, so that $M_Z(\theta) \to \exp(\theta^2/2)$ as $n \to \infty$. **<u>QED</u>**.

# Selected Questions and Answers from the Handout

(**Q27**) Two points A and B are *chosen independently* at random in a line of finite length $l$. If A and B are at **distances** $x_1$ and $x_2$ respectively from one given end, find $E[|x_1-x_2|^n]$, assuming *uniform distribution* of abscissae x on the line.

(**Q28**) A traffic light has constant probability $\lambda\delta t$ of changing to **green** after being **red**, or to red after being green in any *infinitesimal interval*, $\delta t$. Show that a car arriving at a random instant has *probability* 0.5 of passing **through** without waiting, and a probability of $\frac{1}{2}\lambda e^{-\lambda\tau}\delta\tau$ of waiting a *finite* time $\tau > 0$ before the light **changes** in the infinitesimal interval $\delta\tau$.

(**Q29**) Suppose that the distribution of a *random variable* X is $P(x_k) = {}^1/_{2n+1}$ for $x_k = {}^k/_{2n}$ ($k = 0, \pm1, \pm2, ..., \pm n$). Find the *uniform continuous distribution* corresponding to that of X. Hence show that, as $n \to \infty$, var X tends to the **variance** of the corresponding distribution, and, also, $\mu_{2r+1}$ is the same for **both** distributions.

For this question, you need to *know* that $S_2 = \Sigma_{k=1}^n k^2 = {}^1/_6 n(n+1)(2n+1)$. *Recall* that if $S_1 = 1+2+3+...+n = n+(n-1)+(n-2)+...+1$, **then** $2S_1 = (1+n)+(1+n)+(1+n)+....+(1+n)$. *Therefore*, $S_1 = \frac{1}{2}n(n+1)$. **Now** $(n+1)^3-n^3 = 3n^2+2n+1$; $n^3-(n+1)^3 = 3(n-1)^2+3(n-1)+1$; ...; $4^3-3^3 = 3.3^2+3.3+1$; $3^3-2^3 = 3.2^2+3.2+1$; $2^3-1^3 = 3.1^2+3.1+1$. Odd: $(n+1)^3-1^3 = 3.\Sigma_1^n k^2 + 3.\frac{1}{2}n(n+1) + n$. *Therefore*, $n^3+3n^2+3n-{}^3/_2n^2-{}^3/_2n-n = 3\Sigma k^2$; $n^3+{}^3/_2n^2+\frac{1}{2}n = 3\Sigma k^2$, ${}^1/_6n(2n^2+3n+1) = \Sigma k^2$; ${}^1/_6n(2n+1)(n+1) = \Sigma k^2$. We can use the *same* method to find **other** sums of powers, e.g. $S_3$ from $(n+1)^4-n^4$, etc. (*Answer*: $S_3 = {}^1/_4n^2(n+1)^2 = S_1^2$).

(**Q30**) A *necessary* and *sufficient* condition for $E[X^r]$ to exist is that $x^{r-1}P(|X| > x)$ is *absolutely integrable*. Show that $E[X^r] = \int_0^\infty rx^{r-1}(1-F(x))dx - \int_{-\infty}^0 rx^{r-1}F(x)dx$.

(**Q31**) Find the *means* and *variances* for the distributions in question 18. A: **Binomial Distribution:** $\mu = \int xdF = \Sigma_{x=0}^n x\binom{n}{x}p^xq^{n-x} = \Sigma_{x=1}^n x\ {}^{n!}/_{x!(n-x)!}p^xq^{n-x}$ (*since* ${}^0/_{0!} = 0$ as $0! = 1$) $= \Sigma_{y=0}^{n-1} (y+1)\ {}^{n!}/_{(y+1)!(n-y-1)!}p^{y+1}q^{n-y-1} = \Sigma_{y=0}^{n-1}\ {}^{n!}/_{y!(n-1-y)!}p^{y+1}q^{n-1-y} = np\Sigma_{y=0}^{n-1}\ {}^{(n-1)!}/_{y!(n-1-y)!}p^yq^{n-1-y} = np\Sigma_{y=0}^{n-1} \binom{n-1}{y}p^yq^{n-1-y} = np(p+q)^{n-1} = np$ *as* $(p+q)^{n-1} = 1$. To find var X, we use var $X = E(X^2)-\mu^2$, and compute $E(X^2)$ by the *above* method. **Answer** = npq.

**Note**: The Binomial Distribution can be thought of as a *series of Bernoulli Trials* (independent trials), each with parameter p. Let X be the number of **successes** in n trials. Then $X = \Sigma_i X_i$, and we *conclude* that $E(X) = E(\Sigma_i X_i) = \Sigma_i E(X_i) = n\times$the Bernoulli *mean*, and var $X = \text{var}\Sigma_i X_i = \Sigma_i \text{var}X_i + 2\Sigma_{i<j}\text{cov}(X_i,X_j) = \Sigma_i \text{var}X_i$ (since the $X_i$ *are independent*) $= n\times$the Bernoulli *Variance*. For Bernoulli ($p(x) = p^x(1-p)^{1-x}$, $0 < p < 1$, $x = 0, 1, ...$), $\mu = \Sigma_{x=0}^1 xp^xq^{1-x} = p$; var$X = E(X^2)-\mu^2 = E(X^2)-p^2$; and $E(X^2) = \Sigma_{x=0}^1 x^2p^xq^{1-x} = p$. *Therefore*, var$X = p-p^2 = p(1-p) = pq$.

**Poisson:** $E(X) = \mu$. To *find* $E(X^2)-\mu^2 = \text{var } X$, we want $E(X^2) = \Sigma_{x=0}^{\infty} x^2 (e^{-\mu}\mu^x/x!) = \Sigma_{x=1}^{\infty} xe^{-\mu}\mu^x/(x-1)!$ (*using* ${}^0/_{0!} = 0, 0! = 1$) $= \Sigma_{y=0}^{\infty} (y+1)e^{-\mu}\mu^{y+1}/y! = \Sigma_{y=0}^{\infty} ye^{-\mu}\mu^{y+1}/y! + \Sigma_{y=0}^{\infty} e^{-\mu}\mu^{y+1}/y! = \Sigma_{y=1}^{\infty} e^{-\mu}\mu^{y+1}/(y-1)! + e^{-\mu}\mu\Sigma_{y=0}^{\infty} \mu^y/y! = \Sigma_{z=0}^{\infty} e^{-\mu}\mu^{z+2}/z! + e^{-\mu}\mu e^{\mu} = \mu^2e^{-\mu}e^{\mu}+\mu = \mu^2+\mu$. *Therefore*, var $X = \mu^2+\mu-\mu^2 = \mu$.

(**Q32**) Find the *means* and *variances* for the distributions in question 19. A: **Normal Distribution:** [First *from* Q19: sum to unity: $\int dF = \int_{-\infty}^{\infty} f(x)dx = {}^1/_{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} \exp\{-{}^1/_{2\sigma^2}(x-\mu)^2\}dx$. Put $z = {}^{x-\mu}/_{\sqrt{(2\sigma)}}$, so that $dz = {}^{dx}/_{\sqrt{(2)}\sigma}$, and we obtain ${}^1/_{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\infty} e^{-z^2}\sqrt{(2)}\sigma dz = {}^1/_{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-z^2}dz = {}^2/_{\sqrt{(\pi)}}\int_0^{\infty} e^{-z^2}dz = {}^2/_{\sqrt{(\pi)}}I$. **Now** $I^2 = \int_0^{\infty} e^{-z^2}dz\int_0^{\infty}e^{-y^2}dy = \iint_{+\text{ve quad}.} e^{-(y^2+z^2)}dydz$. **Put** $y = r\cos\theta$, and put $z = r\sin\theta$, so that $dydz = rdrd\theta$ $({}^{\partial(y,z)}/_{\partial(r,\theta)}) = r$].

Now $\int_0^{\pi/2}d\theta\int_0^{\infty} e^{-r^2}rdrd\theta = {}^\pi/_4[e^{-r^2}]_0^{\infty} = {}^\pi/_4$. *Therefore*, $\int_{-\infty}^{\infty} {}^1/_{\sigma\sqrt{(2\pi)}} \exp\{-{}^1/_{2\sigma^2}(x-\mu)^2\} dx = 1$ (*). *Differentiate* * w.r.t. $\mu$, **giving** $\int_{-\infty}^{\infty} {}^1/_{\sigma\sqrt{(2\pi)}}\{{}^1/_{\sigma^2}(x-\mu)\}\exp\{-{}^1/_{2\sigma^2}(x-\mu)^2\}dx = 0$. *Therefore*, $\int_{-\infty}^{\infty} {}^1/_{\sigma\sqrt{(2\pi)}} x\exp\{-{}^1/_{2\sigma^2}(x-\mu)^2\}dx = \mu\int_{-\infty}^{\infty} {}^1/_{\sigma\sqrt{(2\pi)}} \exp\{-{}^1/_{2\sigma^2}(x-\mu)^2\} dx$ (**). *Therefore*, $E(X) = \mu.1$, by *. **So** $E(X) = \mu$. Now *differentiate* ** w.r.t. $\mu$, giving $\int_{-\infty}^{\infty} {}^1/_{\sigma\sqrt{(2\pi)}}\{x{}^1/_{\sigma^2}(x-\mu)\}\exp\{-{}^1/_{2\sigma^2}(x-\mu)^2\}dx = 1$. *Therefore*, $\int_{-\infty}^{\infty} {}^1/_{\sigma\sqrt{(2\pi)}} x^2\exp\{-{}^1/_{2\sigma^2}(x-\mu)^2\}dx - \mu\int_{-\infty}^{\infty} {}^1/_{\sigma\sqrt{(2\pi)}} x\exp\{-{}^1/_{2\sigma^2}(x-\mu)^2\}dx = \sigma^2$. So $E(X^2)-\mu E(X) = \sigma^2$; var$X = E(X^2)-\mu^2 = \sigma^2$. **QED**.

(iv) Take the *special case* $\theta = 1$. **Exponential:** $f(x) = \alpha e^{-\alpha x}$ when $x > 0$, and $f(x) = 0$ *when* $x \leq 0$. Now $\int_{-\infty}^{\infty} f(x)dx = \int_0^{\infty} \alpha e^{-\alpha x}dx = [-e^{-\alpha x}]_0^{\infty} = 1$. So $E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} \alpha xe^{-\alpha x}dx = [-xe^{-\alpha x}]_0^{\infty} + \int_0^{\infty} e^{-\alpha x}dx = 0+{}^1/_\alpha = {}^1/_\alpha$. And $E(X^2) = \int_{-\infty}^{\infty} x^2f(x)dx = \int_0^{\infty} \alpha x^2e^{-\alpha x}dx = [-x^2e^{-\alpha x}]_0^{\infty} + 2\int_0^{\infty} xe^{-\alpha x}dx = 0+{}^2/_\alpha E(X) = {}^2/_{\alpha^2}$. *Therefore*, var$X = {}^2/_{\alpha^2}-{}^1/_{\alpha^2} = {}^1/_{\alpha^2}$. **QED**.

(**Q33**) For the *normal distribution*, $N(\mu, \sigma^2)$, show that $\mu_{2r+1} = 0$, for $r \geq 0$, and that $\mu_{2r} = 1.3.5...(2r-1)\sigma^{2r}$, for $r \geq 1$. If X is $N({}^1/_2, 3)$, find the *mean* and *central moments* of the random variable $Y = 2X-7$. The **red** result should be remembered.

An example *similar* to the above question: for N(0,1), $\varphi(t) = e^{-\frac{1}{2}t^2}$ from **tables** $= \Sigma_{n=0}^{\infty} (-1)^nt^{2n}/2^nn! = \Sigma_{n=0}^{\infty} [(i)^{2n}/(2n)!].[(2n)!/2^nn!](t)^{2n} \Rightarrow \mu'_{2n+1} = \mu_{2n+1} = 0$. And $\mu'_{2n} = \mu_{2n} = (2n)!/2^nn! = 1.3.5...(2n-1)$. In *statistics*, this is a **useful** result.

(**Q34**) Let X be any r.v. such that $E(X^2)$ exists. Prove that for any $k > 0$. $P(|X| \geq k) \leq {}^{E(X^2)}/_{k^2}$. Show *that* if Y is a r.v. such that $E(e^{aY})$ exists, where a is a *positive constant*, then for all s, $P(Y \geq s) \leq E(e^{aY})/e^{as}$. A: *Let* $\bar{X}$ be the average of a *random sample* of size n drawn from a population having **uniform** distribution in the interval [0, 4]. Find n *such that* $P(|{}^1/_2\bar{X} -1| \geq 0.4) \leq 0.1$. (**Answer**: $n \approx 21$).

(**Q36**) A lucky dip is run at a garden fete. On payment of **50p** (not refunded in the case of a *win*), a participant draws 3 balls at random from a bag containing 5 red, 3 black, and 2 white balls. If he draws the **3** black balls, he receives a *prize* of £5; if he draws **both** white balls, he receives a *prize* of £1.50; and if he draws **3** red balls, he wins his *money back*. What is the expected **gain** to the organisers on each draw? What is the probability that *among the first 20 participants*, 3 will **win** a prize? In what sense is the number of participants relevant to your answer?

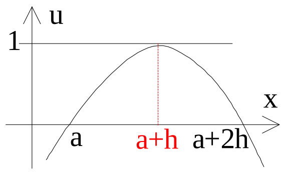
(**Q37**) In a game of dice, a player *continues to throw 2 fair die* until either the sum 2, the sum 4, or the sum 6 is obtained, in which case the game is **terminated**, with the player <u>winning</u> if 2 or 4 is obtained, and <u>losing</u> if 6 is obtained. Find the *probability* that the player wins in the first n throws. What is the probability of a win if the game is allowed to continue *indefinitely*? What odds should the player accept if the game is fair?

(**Q38**) Find the c.f.'s for the **distributions** in question 18.

(**Q39**) Find the c.f.'s for the **distributions** in question 19.

(**Q40**) If X is a r.v. such that $u(x) \geq 0$ *everywhere*, and $u(x) > a > 0$ for all $x \in I$, then show that $P(X \in I) \leq a^{-1}E(u(X))$. If, *on the other hand*, $u(x) \leq 0$ for all $x \notin I$, and $u(x) \leq 1$ for all $x \in I$, show that $P(X \in I) \geq E(u(X))$. Let X be *positive* (i.e. $F(0) = 0$), with $E(X) = 1$, and $E(X^2) = b$. The **polynomial** $u(x) = h^{-2}(x-a)(a+2h-x)$ is *positive* only for $a < x < a+2h$, and $u(x) \leq 1$ everywhere.

**When** $0 < a < 1$, show that $E(u(X)) \geq [2h(1-a)-b]h^{-2}$. *Choosing* $h = b(1-a)^{-1}$, deduce that $P(X > a) \geq (1-a)^2b^{-1}$. Hence **show** that for a r.v. X such that $E(X^2) = 1$, we have $E(X^4) = M$, and $P(|X| > t) \geq (1-t^2)^2M^{-1}$. (Various *non-trivial generalisations* of Chebychef's inequality can be derived by the sort of method **outlined** in this exercise).



A: (a) $E(u) = \int u\,dF$ $(\forall x) \geq \int_{x\in I} u\,dF$ (*since* $u \geq 0$) $\geq \int_{x\in I} a\,dF$ (*since* $u > a, x \in I$) $= aP(X \in I)$. (b) $E(u) = \int_{\forall x} u\,dF \leq \int_{x\in I} u\,dF$ (since $u \leq 0, x \notin I$) $\leq \int_{x\in}$ dF (since $u \leq 1, x \in I$) $= P(X \in I)$. <u>**Notes**</u>: from the *diagram*, $u(x) = 0$ at $x = a, a+2h$; $u'(x)$ at $x = a+h$, $u(a+h) = 1$. If $u = h^{-2}(x-a)(a+2h-x)$, then $E(u) = \int_0$ $\frac{1}{h^2}\{2x(a+h)-a(a+2h)-x^2\}dx = \frac{1}{h^2}\{2a+2h-a^2-2ah-b\} = \frac{1}{h^2}\{2h(1-a)+a(1-a)+a-h\} \geq \frac{1}{h^2}\{2h(1-a)-h\}$, *since* $a > 0$, and $1-a > 0$. By (b), $P(X \in I) = P(a < X < a+2h) \geq \frac{(1-a)^2}{b^2}\{2.\frac{b}{1-a}(1-a)-b\} = \frac{(1-a)^2}{b}$. But $(X > a) \supset (a+2h > X > a)$. *Therefore*, $P(X > a) \geq P(X \in I) \geq \frac{(1-a)^2}{b}$, and $P(|X| > t) = P(X^2 > t^2) \geq (1-t^2)^2\,\frac{1}{B}$, where $B = E((X^2)^2) = E(X^4) = M$.

(**Q41**) Given that $\phi_1(t) = \frac{2}{\pi}\int^{\infty} \frac{1-\cos x}{2}\cos(tx)\,dx$ and $\phi_2(t) = \frac{1}{2} + \frac{4}{\pi^2}\sum_0^{\infty} \frac{\cos(2n+1)\pi t}{(2n+1)^2}$ are two c.f.'s, find the forms of the *corresponding pdf and pmf.*

(**Q43**) An infinite population contains *4 different kinds of members in equal proportions.* A random sample is taken, stopping as soon as at least one of <u>each</u> kind has been obtained. (i) Find the p.f. for the *sample size N* when this stage is reached.

(ii) Find the **mean** and **variance** of N. (iii) Consider a process in which success occurs with probability p, where $0 < p < 1$, and *failure* occurs with probability 1-p, at a single trial. Let X be the number of trials until the first success is achieved in a sequence of *independent* trials. Show that $E(X) = \frac{1}{p}$, and that $\text{var } X = \frac{q}{p^2}$ $(q = 1-p)$, and hence devise a method for *obtaining the result in (ii).*

A: Do (iii) first. **Assume** that for $p(x) = q^{x-1}p$, we have $E(X) = p^{-1}$, and $\mathrm{var}X = {}^{q}/_{p^2}$, as *proved* in the notes. (iii) There are **four** different kinds of members in *equal* proportions. Let $X_1$ be the number of *observations* until any type is seen: $p_1 = 1$, $E(X_1) = 1$. Let $X_2$ be the number of observations until we see a **2nd** type: $p_2 = ¾$, $E(X_2) = {}^{4}/_{3}$. $X_3$: $p_3 = {}^{1}/_{2}$, $E(X_3) = 2$. $X_4$: $p_4 = ¼$, $E(X_4) = 4$.

So we *get* $E(N) = E(X_1+X_2+X_3+X_4) = 1+{}^{4}/_{3}+2+4 = 8{}^{1}/_{3}$, and $\mathrm{var}(N) = \Sigma_{i=1}^{4}\ \mathrm{var}X_i$ ($\mathrm{var}X = {}^{1-p}/_{p^2}$) $= 0 + ({}^{1/4}/_{9/16}) + ({}^{1/2}/_{1/4}) + ({}^{3/4}/_{1/16}) = {}^{16}/_{32}+{}^{4}/_{2}+{}^{48}/_{4} = {}^{4}/_{9}+2+12 = 14{}^{4}/_{9}$. (i) **We need**: (1) all 1's $\Rightarrow ({}^{1}/_{4})^{x-1}$; (2) all 1's and 2's, but not *only* 1's or 2's $\Rightarrow ({}^{1}/_{2})^{x-1}-2({}^{1}/_{4})^{x-1}$; and (3) all 1's, 2's and 3's, but not *only* 1's, 2's or 3's $\Rightarrow ({}^{3}/_{4})^{x-1} - 3\{({}^{1}/_{2})^{x-1}-2({}^{1}/_{4})^{x-1}\} - 3\{({}^{1}/_{4})^{x-1}\}$. Therefore, $P(N = n) = 4$[the red expression]${}^{1}/_{4}$ = the red expression.

(**Q44**) An **infinite** population contains 4 different kinds of members in the proportions 1:2:3:4. A random sample is taken, stopping as soon as *at least one member of each kind* is obtained. Denote the probability that the 1st n-1 sample elements are of type 1 or 2 or 3, with at least **one** of each by $p_4(n)$.

*Show* that $p_4(n) = ({}^{6}/_{10})^{n-1} - ({}^{5}/_{10})^{n-1} - ({}^{4}/_{10})^{n-1} + ({}^{2}/_{10})^{n-1} + ({}^{1}/_{10})^{n-1}$. Explain how to calculate this *probability mass function*, p(n), for the random sample size N, and state how it may be used to calculate the **mean** and **variance** of N. (Explicit formulae for p(n), E(N) and var N are not required).

(**Q45**) Define the *characteristic function* as $\phi(t)$ for the real random variable X. If X has the *Binomial* distribution, with probability mass function $p(x) = ({}^{n}{}_{x})p^x q^{n-x}$, for x = 0, 1, 2, ..., n, where $0 < p < 1$ and $q = 1-p$ are *constants*, obtain the characteristic function $\phi(t)$ of X. Write down the **moment generating function** M(t) corresponding to $\phi(t)$, and obtain ${}^{dM}/_{dt}$ and ${}^{d^2M}/_{dt^2}$.

Hence show that the *mean* and *variance* of X are respectively np and npq. Write down the characteristic function for a Bernoulli random variable (a *special case* of the Binomial, with n = 1). Hence verify that a Binomial random variable may be thought of as a sum on **n independent and identically distributed** Bernoulli random variables.

A: Method 1: **Direct Method**: The probability that the first (x-1) draws produce four different kinds of at least *one type* of each, and that the $x^{th}$ draw produces the $5^{th}$ type, is given by $({}^{4}/_{5})^{x-1} - 4({}^{5}/_{5})^{x-1} + 6({}^{2}/_{5})^{x-1} - 4({}^{1}/_{5})^{x-1} = p(x)$, which is *valid* for $x \geq 2$. [Proceed to **proof** that P((x-1) draws produce *type 1*) $= ({}^{1}/_{5})^{x-1}$; that P((x-1) draws produce *type 1 and type 2, with at least one of each)* $= ({}^{2}/_{5})^{x-1} - 2({}^{1}/_{5})^{x-1}$; that P((x-1) draws produce *types 1, 2 and 3, with at least one of each)* $= ({}^{3}/_{5})^{x-1} - 3\{({}^{2}/_{5})^{x-1}-2({}^{1}/_{5})^{x-1}\} - 3({}^{1}/_{5})^{x-1}$; that ....]

Now $E(X) = \Sigma_{x=5}^{\infty}\ xp(x) = \Sigma_{x=2}^{\infty}\ xp(x)$ (*since* $p(2) = p(3) = p(4) = 0$) $= \Sigma_{x=1}^{\infty}\ xp(x)+1 = [1/({}^{1}/_{5})^2] - 4[1/({}^{2}/_{5})^2] + 6[1/({}^{3}/_{5})^2] - 4[1/({}^{4}/_{5})^2] + 1 = 25-25+6.{}^{25}/_{8}-4.{}^{25}/_{11}+1 = 11{}^{5}/_{12}$. And $E(X^2) = \Sigma_{5}^{\infty}\ x^2p(x) = \Sigma_{x=1}^{\infty}\ x^2p(x)+1 = 2[[1/({}^{1}/_{5})^2] - 4[1/({}^{2}/_{5})^2] + 6[1/({}^{3}/_{5})^2] - 4[1/({}^{4}/_{5})^2]] - 10{}^{5}/_{12} + 1 = ... = 165{}^{74}/_{144}$. Therefore, $\mathrm{var}X = 165{}^{74}/_{144} - ({}^{137}/_{12})^2 = 25{}^{25}/_{144}$. **Note** that if $S = 1+r+r^2+...$, $rS = r+r^2+r^3+...$, thus $(1-r)S = 1$, and *therefore* $S = {}^{1}/_{1-r}$. **So** if $S = 1+2r+3r^2+...$, $(1-r)S = 1+r+r^2+... = {}^{1}/_{1-r}$, *and therefore* $S = {}^{1}/_{(1-r)^2}$. And *if* $S = 1+2^2r+3^2r^2+...$, $(1-r)S = 1+3r+5r^2+...$, $(1-r)^2S = 2(1+r+r^2+...)-1$, and *thus* $S = {}^{2}/_{(1-r)^3}-{}^{1}/_{(1-r)}$.

Now find the *types of object in the population*. Sample until **at least one** of each type is obtained, and then stop. Find the *mean* and *variance* of the r.v. X = the number of trials. Method 2: **Binomial method:** The *failure* probability is $q = 1-p$. Let $p(x) = P((x-1)$ failures, and a *success* on the $x^{th}$ trial) $= (1-p)^{x-1}p$. Now $E(X) = \Sigma x(1-p)^{x-1}p = p\Sigma x(1-p)^{x-1} = p\{1+2(1-p)+3(1-p)^2+...\} = {}^{p}/_{[1-(1-p)]^2} = {}^{1}/_{p}$.

*From* g.f., $P(x) = {}^{pt}/_{1-qt}$, and we get $E(X) = {}^{dP}/_{dt}|_{t=1} = {}^{1}/_{p}$. Further, $E(X^2) = \Sigma x^2(1-p)^{x-1}p = p\{1+2^2(1-p)+3^2(1-p)^2+...\} = p\{{}^{2}/_{p^3}-{}^{1}/_{p^2}\} = {}^{2}/_{p^2}-{}^{1}/_{p}$. Therefore, $\mathrm{var}(X) = E(X^2)-E(X)^2 = {}^{1-p}/_{p^2} = {}^{q}/_{p^2}$. *From* g.f., $E(X(X-1)) = {}^{d^2P}/_{dt^2}|_{t=1} = {}^{2ap}/_{p^3} = {}^{2a}/_{p^2}$. It follows that $\mathrm{var}X = E(X(X-1)) + E(X) - [E(X)]^2 = {}^{2a}/_{p^2}+{}^{1}/_{p}-{}^{1}/_{p^2} = {}^{q}/_{p^2}$.

In the *actual* problem, let $X_1$ be the number of draws needed to get **any** kind, let $X_2$ be the number of *extra* draws needed to get other kinds (*other* than the first drawn), let $X_3 = ...$ Then, we **want** $E(\Sigma X_i)$, and $\mathrm{var}(\Sigma X_i)$. Since the $X_i$'s are *independent*, we have $E(\Sigma X_i) = \Sigma E(X_i) = \Sigma 1/p_i = 1+{}^{5}/_{4}+{}^{5}/_{3}+{}^{5}/_{2}+{}^{5}/_{1} = 11\,{}^{5}/_{12}$, and $\mathrm{var}(\Sigma X_i) = 0 + [{}^{1/5}/_{(4/5)^2}] + [{}^{2/5}/_{(3/5)^2}] + [{}^{3/5}/_{(2/5)^2}] + [{}^{4/5}/_{(1/5)^2}] = ... = 25-{}^{25}/_{144}$.

## Bernoulli and Binomial Distributions

$p(x) = P(X = x) = p^x(1-p)^{1-x}$, for $0 < p < 0$, and $x = 0, 1$. Now $\mu = E(X) = \Sigma_{x=0}^{1}\, xp^xq^{1-x} = 0+p = p$, where $q = 1-p$. Further, $\sigma^2 = \mathrm{var}\ X = E(X^2)-\mu^2 = E(X^2)-p^2$. **Now** $E(X^2) = \Sigma_{x=0}^{1}\, x^2p^xq^{1-x} = 0+p = p$, *so that* $\sigma^2 = p-p^2 = p(1-p) = pq$. *Consider* $X = X_1+X_2+...+X_n$, where the $X_i$ are *independent*, and each has the **Bernoulli distribution**. Thus $E(X) = E(\Sigma_iX_i) = \Sigma_iE(X_i) = np$, and $\mathrm{var}\ X = \mathrm{var}\ \Sigma_i X_i = \Sigma_i\ \mathrm{var}\ X_i + 2\Sigma_{i<j}\ \mathrm{cov}\ (X_i,X_j) = \Sigma_i\ \mathrm{var}\ X_i = npq$, since the $X_i$ are **independent**.

The *Binomial Distribution* can be though of as series of n Bernoulli trials, all independent with *parameter* p (iid r.v.'s). Let X be the number of "**successes**" (the probability of *success* is p; and *failure,* q) in n trials. Then, by the above **reasoning**, the Binomial mean and variance are *respectively* np and npq.

Otherwise, consider a sequence of x **successes** and n-x **failures —** any particular sequence occurs with *probability* $p^x(1-p)^{n-x} = p^xq^{n-x}$. But there *are* ${}^nC_x$ possible ways in which x *successes* can occur, thus $p(x) = P(X = x) = {}^nC_xp^xq^{n-x}$. Then, *directly,* $\mu = E(X) = \Sigma_{x=0}^{n}\ x\ {}^nC_xp^xq^{n-x} = \Sigma_{x=1}^{n}\ x\ {}^nC_xp^xq^{n-x} = \Sigma_{y=0}^{n-1}\ (y+1){}^{n!}/_{(n-y-1)!(y+1)!}\ p^{y+1}q^{n-y-1} = np\Sigma_{y=0}^{n-1}\ {}^{(n-1)!}/_{(n-1-y)!y!}\ p^yq^{n-1-y} = np\Sigma_{y=0}^{n-1}\ {}^{n-1}C_y\ p^yq^{n-1-y} = np(p+q)^{n-1} = np$. Similarly, we may *calculate* $\sigma^2 = \mathrm{var}\ X = E(X^2)-\mu^2 = npq$.

(**Q46**) (a) The probability of success at each of a *series of independent trials* is p, where $0 < p < 1$. Show that the **probability generating function** of X, the number of the trial on which the *first* success occurs, is given by $P(t) = {}^{pt}/_{1-qt}$, where $q = 1-p$. *Evaluate* ${}^{dP}/_{dt}$ and ${}^{d^2P}/_{dt^2}$, and deduce that X has mean ${}^{1}/_{p}$, and variance ${}^{q}/_{p^2}$.

(b) The probability of *success* at each of a **series** of independent trials is p, where $0 < p < 1$. Show that the **moment generating function** of X, the number of the trial on which the *first* success occurs, is given by $M(t) = pe^t/1-qe^t$, where $q = 1-p$. *Evaluate* ${}^{dM}/_{dt}$ and ${}^{d^2M}/_{dt^2}$, and *deduce* that X has mean ${}^{1}/_{p}$, and variance ${}^{q}/_{p^2}$.

A *population* contains k different kinds of members in **equal** proportions. Members are selected at random, identified, and *replaced* into the population after noting their type. Sampling continues until every type has been observed at **least** once, and then stops. Show that the *mean number Y of samplings* is given by $E(Y) = k\Sigma_{i=1}^{i=k}\,{}^{1}/_{i}$. If k = 4, calculate the *mean* and the *variance* of Y.

(**Q47**) Find, using *characteristic functions*, the exact sampling distributions of (i) the **SAMPLE SUM,** S, and the **MEAN**, $\bar{X}$, of a random sample of size n drawn from a *population* with (a) <u>Normal</u>, (b) <u>Cauchy</u>, (c) <u>Exponential</u>, and (d) <u>Gamma</u> distributions; and (ii) the SUM S when the population has (e) <u>Binomial</u> distribution, and (f) <u>Poisson</u> distribution.

(**Q48**) Can you identity the *exact sampling distribution* of the sample mean for a sample drawn from a **Binomial** population?

## <u>The Gamma Function</u>

**<u>Definition</u>**: $\Gamma(z) = \int_0^\infty e^{-t}t^{z-1}dt = [-e^{-t}t^{z-1}]^\infty_0 + \int_0^\infty e^{-t}(z-1)t^{z-2}dt = 0+(z-1)\Gamma(z-1)$. When z is an *integer*, say z = n, we have $\Gamma(n) = (n-1)(n-2)....3.2.1.\Gamma(1)$, *where* $\Gamma(1) = \int^0_\infty e^{-t}dt = [-e^{-t}]^\infty_0 = 1$. *Hence*, for z = n, an **integer**, $\Gamma(n) = (n-1)!$. Also, we *need to know that* $\Gamma({}^1/_2) = \sqrt{(\pi)}$. By *definition*, $\Gamma({}^1/_2) = \int_0^\infty e^{-t}t^{-1/2}dt$. **Put** $\omega^2 = t$, so that $2\omega d\omega = dt$, and, therefore, $\Gamma(½) = \int_0^\infty e^{-\omega^2}\omega^{-1}2\omega d\omega = 2\int^\infty_0 e^{-\omega^2}d\omega$.

**Let** $I = \int_0^\infty e^{-\omega^2}d\omega$. Then $I^2 = \int_0^\infty e^{-\omega^2}d\omega\int_0^\infty e^{-x^2}dx$. **Put** $\omega = r\cos\theta$, and put $x = r\sin\theta$, so that $d\omega dx = |{}^{\partial(\omega,x)}/_{\partial(r,\theta)}|drd\theta = rdrd\theta$. *Therefore*, $I^2 = \int^{\pi/2}_0 d\theta \int_0^\infty re^{-r^2}dr = {}^{\pi}/_{2}{}^{1}/_{2}\int^\infty_0 e^{-r^2}d(r^2) = {}^{\pi}/_{4}[-e^{-r^2}]_0^\infty = {}^{\pi}/_{4}$. Thus $I = {}^{\sqrt{(\pi)}}/_{2}$, and, therefore, *as required*, $\Gamma(½) = \sqrt{\pi}$. **<u>Exercise 1</u>**: By *setting* $t = z+t\sqrt{z}$ in the *definition* of $\Gamma(n+1)$, obtain <u>Stirling's approximation</u> for n! for large n, i.e. $n! = \sqrt{(2\pi)}n^{n+½}e^{-n}$.

We sometimes *encounter* the so-called **Beta** function, $B(z_1, z_2) = 2\int^{\pi/2}_0 \sin^{2z_1-1}\theta.\cos^{2z_2-1}\theta d\theta$. **Or**, by *setting* $t = \sin^2\theta$, $B(z_1, z_2) = \int_0^1 t^{z_1-1}(1-t)^{z_2-1}dt$. **<u>Exercise 2</u>**: Consider the *product* $\Gamma(y)\Gamma(z) = \int_0^\infty e^{-t}t^{y-1}dt\int_0^\infty e^{-s}s^{z-1}ds$. Treat this product as a *double integral*. Set $t = w\cos^2\theta$, and set $s = w\sin^2\theta$, and thus *show that* ${}^{\partial(t,s)}/_{\partial(w,\theta)} = 2w\cos\theta.\sin\theta$. **Deduce** that $\Gamma(y)\Gamma(z) = \Gamma(y+z)B(y,z)$.

# Exam Paper: May 2001

**Answer 3 questions out of 5 (Questions Done: 1, 2, 4, 5)**

(1) (a) Define the upper limit $\lim_{n\to\infty}$ sup $A_n$ and the lower limit $\lim_{n\to\infty}$ inf $A_n$ of a sequence of sets. **[4 marks]**

(b) Define a convergent sequence $\{A_n\}$ in terms of its upper and lower limits. **[1 mark]**

(c) Prove that $\lim_{n\to\infty} \inf A_n \subset \lim_{n\to\infty} \sup A_n$. **[5 marks]**

(d) Find the upper and lower limits of the following sequences of events on the sample space ‘ and determine the convergence or otherwise of each sequence.

(i) $\{A_n\}$ where $A_n = (^1/_n - 1, ^1/_n)$ for n = 1, 2, 3, ....
(ii) $\{A_n\}$ where $A_n = [0, 1]$ for n even,
$A_n = \{0\}$ for n odd. **[10 marks]**

(2) (a) Explain carefully what is meant by a Probability Space ($S$, $F$, P) and state its axioms. **[5 marks]**

(b) Define a monotonic sequence of events $\{A_n\}$ in the sample space $S$ and state, without proof, the limit of a monotonic sequence distinguishing between the cases when the sequence is (a) increasing and (b) decreasing. **[5 marks]**

(c) If $\{A_n\}$ is a monotonically increasing sequence of events then prove that

$$\lim_{n\to\infty} P(A_n) = P(\bigcup_{n=1}^{\infty} A_n),$$

indicating which axioms of the probability measure P you are using. **[10 marks]**

(3) (a) Let X be a continuous random variable with probability density function f(x) and let y = g(x) be a strictly monotonic function of x such that $^{dg}/_{dx}$ is continuous and non-zero in some open interval A. If $g^{-1}(y)$ is the unique inverse of g(x) and B is the image of A in the range of Y, show that the random variable Y = g(X) is continuous and has probability density function h(y) for all $y \in$ B given by

$$h(y) = f(g^{-1}(y))\left|\frac{d(g^{-1}(y))}{dy}\right|.$$ **[10 marks]**

(b) A continuous random variable X has the probability density function

$$f(x) = \begin{cases} kx^2 e^{-x^2}, & x > 0 \\ 0, & x \le 0 \end{cases}.$$

Find the value of k. **[3 marks]**

(c) Find the probability density function h(y) of $Y = X^2$. **[5 marks]**

(d) Show that h(y) = $\Gamma(\theta, \alpha; y)$ and determine the values of the parameters $\theta$, $\alpha$ of the Gamma distribution. **[2 marks]**

(You may assume without proof that $\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}$ ).

(4) (a) A 2-dimensional continuous random variable (X, Y) has pdf f(x, y). Define the marginal pdf's $f_X(x)$, $f_Y(y)$. Give a necessary and sufficient condition in terms of f(x, y), $f_X(x)$, $f_Y(y)$ for X, Y to be independent. **[4 marks]**

(b) If (X, Y) has the bivariate normal distribution with pdf

$$f(x,y) = \frac{1}{2\pi\sqrt{1-\rho_{XY}^2}} \exp\left\{\frac{-1}{2(1-\rho_{XY}^2)}(x^2 - 2\rho_{XY}xy + y^2)\right\},$$

show that X, Y each has univariate normal distribution and that if $\rho_{XY} = 0$ then X, Y are independent. **[16 marks]**

(You may assume without proof that $\int_{-\infty}^{\infty} e^{-t^2} dt = \sqrt{\pi}$ ).

(5) (a) Define the moment generating function $M_X(t)$ of a one-dimensional random variable X. **[2 marks]**

(b) Show that $M_Z(t) = \exp(^{t^2}/_2)$ where Z denotes the standard normal random variable. **[4 marks]**

(c) Let $X_1, X_2, \ldots, X_n$ be independent random variables each having the same distribution with mean $\mu$ and variance $\sigma^2$. If $Z_n = \frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}$ , where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ , show that $M_{Z_n}(t) \to \exp(\frac{t^2}{2})$ as n → ∞. Deduce that (Central Limit Theorem) $Z_n \to Z$ as n → ∞. State without proof any theorems you use. **[14 marks]**