LECTURE 1

1. DISCRETE PROBABILITY. Consider an experiment whose result is not known (i.e., not *deterministic*) but can take any one of a finite number of outcomes, say $\omega_1, \dots, \omega_n$. The set

$$\Omega = \{\omega_1, \dots, \omega_n\}$$

is called the *sample space* and any subset of Ω is called an *event*. To each event $A \subset \Omega$ is assigned a *probability*, $P(A)$ which can be interpreted as the relative frequency of that event if the experiment is performed over and over. Clearly

1. $0 \le P(A) \le 1$ and $P(\emptyset) = 0$, $P(\Omega) = 1$,
2. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Coin tossing. The experiment is to toss a coin once. The sample space is $\Omega = \{Head, Tail\}$ or $\{H, T\}$ or even $\{0, 1\}$. For purposes of idealization we ignore the possibility that the coin lands on its side, or that the universe comes to an end before completion of the experiment.

For n coin tosses the sample space could be written

$$\Omega = \{(x_1, \dots, x_n), x_i = 0, 1\} = \{0, 1\}^n.$$

Clearly the number of outcomes, $|\Omega| = 2^n$, and each one has probability $1/2^n$. A typical event one might be concerned with would be

$$A = \{\omega \in \Omega : y_1, \dots, y_n \le m, y_i = x_1 + \cdots + x_i - i/2\}.$$

Note that $P(A)$ is not particularly easy to compute.

Permutations. An urn contains n distinguishable balls. We select m of them one by one without replacement. The sample space is

$$\Omega = \{(x_1, \dots, x_m),\ x_i \ne x_j,\ i \ne j,\ x_i = 1, \dots, n\}$$

We count $|\Omega| = n(n-1)\cdots(n-m+1) =$ *the number of permutations of n objects taken m at a time.*

Combinations. Suppose we do the same experiment as in the previous example, but we ignore the order of the outcomes. The sample space is

$$\Omega = \{A \subset \{1, \dots, n\}, |A| = m\}$$

and

$$|\Omega| = \binom{n}{m} = \frac{n!}{(n-m)!m!}.$$

The Birthday Problem. Suppose there are n students in a class. What is the probability that two students have the same birthday?

Solution. The sample space is

$$\Omega = \{(x_1, \dots, x_n),\ x_i = 1, \dots, 365\}.$$

There are 365^n sample points and each has probability $1/365^n$. The event whose probability we are interested in is

$$A = \{(x_1, \ldots, x_n) : x_i = x_j, \text{ for some } i \neq j\}.$$

The probability $P(A)$ is unpleasant to calculate directly because there could be many coincidental birthdays. The complementary event is the one in which no two students have the same birthday,

$$A^C = \{(x_1, \ldots, x_n) : x_i \neq x_j, \ i \neq j\}.$$

By elementary rules of probability $P(A) = 1 - P(A^C)$. The probability $P(A^C)$ is easy to calculate since

$$|A^C| = 365 \cdot 364 \cdots (365 - n + 1).$$

So the probability that two students have the same birthday is

$$1 - \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n}.$$

Most people guess that you need quite a large class before the probability becomes larger than $1/2$, but you can check on a calculator that $n = 23$ is enough. One should note the idealization in the problem that all birthdays are equally likely, as well as the fact that we ignored Feb. 29.

Symmetric random walk. At times $1, \ldots, n$ you bet \$1 on a fair game which pays out \$2 if you win and nothing if you lose. The outcome of the experiment is the time evolution of your fortune $(x_1, \ldots, x_n)$ so the sample space is

$$\Omega = \{(x_1, \ldots, x_n), \ x_i - x_{i-1} = \pm 1\}$$

with the convention that $x_0 = 0$. We have $|\Omega| = 2^n$ and each possible outcome is equally likely, therefore has probability $1/2^n$.

Problem. Let $X =$ the number of wins. What is the *distribution* of X?

Solution. The possible values for X are $0, \ldots, n$. Let m be such a possible value. Sample points corresponding to trials with m wins are in one to one correspondence with subsets of size m of a set of size n. There are $\binom{n}{m}$ such subsets and therefore

$$P(X = m) = \frac{\binom{n}{m}}{2^n}.$$

Random variables. A real valued function X on the sample space such as in the previous example is called a *random variable*. It is a measurement of the outcome of the experiment. The range $\{x_1, \ldots, x_n\}$ of the random variable is simply the set of n real numbers it can obtain and the distribution of the random variable is the set of probabilities $P(X = x_i)$ where x_i is in the range. Note that $P(X = x_i)$ should really be written

$P(\{\omega : X(\omega) = x_i\})$ but no one ever does. The *expectation* of the random variable is the number

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i).$$

It is often called the mean and denoted by m or μ. The *variance* is

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

The square root of the variance is called the standard deviation and is always called σ.

Asymmetric random walk. Same as for the symmetric random walk except the game is not fair. The probability to win is p and to lose is $1 - p$ where $0 < p < 1$. The sample space is the same but each outcome is no longer equally likely. Instead $P(\omega) = p^{\text{number of wins}}(1-p)^{\text{number of losses}}$. Now

$$P(X = m) = \binom{n}{m} p^m (1-p)^{n-m}.$$

This is called the *Binomial distribution.* It depends on two parameters, p and n, and $\mu = E[X] = pn$ and $\sigma^2 = Var(X) = np(1-p)$.

DeMoivre-Laplace (central) limit theorem. A good way to compute numbers like these binomial probabilities when n and m are large is with *Stirling's formula*,

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}.$$

The symbol $\sim$ means that the left hand side over the right hand side approaches 1 as $n \to \infty$. You get

$$\binom{n}{m} p^m (1-p)^{n-m} \sim \frac{1}{\sqrt{2\pi n(\frac{m}{n})(1-\frac{m}{n})}} \frac{p^m(1-p)^{n-m}}{(\frac{m}{n})^m (1-\frac{m}{n})^{n-m}}.$$

If you introduce $\rho = \frac{m}{n}$ and $H(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$ then it becomes

$$\frac{1}{\sqrt{2\pi n\rho(1-\rho)}} e^{-nH(\rho)}.$$

One can check $H(p) = H'(p) = 0$ and $H''(p) = 1/p(1-p)$ so using Taylor's formula, $H(\rho) = \frac{(\rho-p)^2}{2p(1-p)} + \cdots$. Therefore

$$\binom{n}{m} p^m (1-p)^{n-m} \sim \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{z^2}{2}}$$

where $\sigma = \sqrt{np(1-p)}$ is the standard deviation and $z = \frac{m-pn}{\sqrt{np(1-p)}}$. One can check that the formula holds uniformly on bounded sets of z (actually it even holds for bounded sets of $n^{-1/6}|z|$.) This shows how the normal distribution approximates the binomial.

Problem. A coin is tossed 10,000 times and it comes up heads 5200 times. Should we conclude that the coin is biased?

Solution. In n Bernoulli trials the probability of between $np - 2\sqrt{np(1-p)}$ and $np+2\sqrt{np(1-p)}$ successes can be computed by the normal approximation $\int_{-2}^{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}\, dz$ which you can look up on a table and find to be about 0.9545.

Now suppose our coin is unbiased. Then $p = 1/2$ and $n = 10,000$ and the probability to have between 4950 and 5050 heads is about 95%. So with 95% certainty we reject the hypothesis that the coin is unbiased.

Poisson limit theorem. Choose some positive number λ and consider the behaviour of the binomial probabilities as $n \to \infty$ when $p = \lambda/n$. In other words we do a lot of trials with a very small probability to win. The expected number of wins is λ and the probability that there are m wins is

$$\binom{n}{m}\left(\frac{\lambda}{n}\right)^m\left(1-\frac{\lambda}{n}\right)^{n-m} = \left\{\frac{n(n-1)\cdots(n-m+1)}{n^m}\left(1-\frac{\lambda}{n}\right)^{-m}\right\}\frac{\lambda^m}{m!}\left(1-\frac{\lambda}{n}\right)^n.$$

As $n \to \infty$ the term in curly brackets goes to 1 and $(1-\frac{\lambda}{n})^n \to e^{-\lambda}$. So the limiting probability to have m wins is

$$\frac{\lambda^m e^{-\lambda}}{m!}.$$

This is called the *Poisson distribution.*

Problem. There were 500 guests at Sharon Stone's birthday party. What is the probability that 5 of them have the same birthday as her?

Solution. The answer $\binom{500}{5}(\frac{1}{365})^5(\frac{364}{365})^{495}$ is not particularly helpful. The Poisson approximation $\frac{(\frac{500}{365})^5 e^{-(\frac{500}{365})}}{5!}$ can be quickly computed to be about 1/100. The error in the approximation is very small, about 1/10000.

2. GENERAL SAMPLE SPACES. When the sample space is no longer countable there are a number of new issues that come up.

Bertrand's paradox. *Problem.* Choose a chord at random from a circle of radius 2. What is the probability p that it intersects a concentric circle of radius 1?

Solution 1. The chord is uniquely determined by its midpoint so $p = \frac{\text{area of inner circle}}{\text{area of outer circle}}$. This gives the solution $p = 1/4$.

Solution 2. By symmetry under rotation one can assume the chord is vertical. The diameter of the big circle is twice that of the smaller. Therefore $p = 1/2$.

Solution 3. By symmetry we can assume the end of the chord lies on a horizontal line through the centre. The angle the chord makes with the horzontal lies in $[-\pi/2, \pi/2]$ and the chord hits the inner circle if $[-\pi/6, \pi/6]$, so $p = 1/3$.

There are many other equally correct solutions, each corresponding to a different interpretation of the phrase "at random". Although similar misunderstandings arise in finite probability, they become harder to notice as the models become more complex.

Random walk. The sample space is

$$\Omega = \{(x_1, x_2, \ldots),\ x_i - x_{i-1} = \pm 1\}.$$

Obviously each sample point has probability 0, though events of the form

$$A = \{(x_1, x_2, \dots) \in \Omega : x_i = a_i, i = 1, \dots, n\}$$

where $a_1, \dots, a_n$ are specified are really just questions about finite random walks and the probabilities are the same as before.

Gambler's ruin. *Problem.* Suppose a gambler starts with a fortune $x > 0$ and repeatedly wagers \$1 on a game which pays \$2 with probability p and nothing with probability $1 - p$. What is the probability he is ruined?

Solution. This is the same as asking for the probability of the event

$$A = \{(x_1, x_2, \dots) \in \Omega : x_i > 0, i = 1, 2, \dots\}$$

if we let the random walk start at x. Note that it is not of the form previously discussed. It turns out to be a little easier to calculate the probability of

$$B = \{(x_1, x_2, \dots) \in \Omega : x_i > 0, i = 1, 2, \dots, \tau_b\}$$

where $\tau_b = \inf\{i : x_i \geq b\}$ is the first time that the walk goes above level b and then let $b \to \infty$ to recover the probability we want. B is really just the event that the gambler's fortune gets to b before 0. Let $f(x)$ denote $P(B)$. It depends on x through the initial fortune of the gambler and the big trick is to look for an equation that the whole function f satisfies. Since $x_1 - x$ is either $+1$ or -1,

$$P(B) = P(B \cap x_1 - x = +1) + P(B \cap x_1 - x = -1)$$

which we can rewrite in the language of conditional probability as

$$P(B) = P(B \mid x_1 - x = +1)P(x_1 - x = +1) + P(B \mid x_1 - x = -1)P(x_1 - x = -1)$$

Note however that once the random walk takes a jump up, the whole problem may as well have started at $x + 1$, and analogously for a jump down. So what we have is

$$f(x) = f(x+1)p + f(x-1)(1-p).$$

Clearly $f(0) = 0$ and $f(b) = 1$ as well. One can check the unique solution is $f(x) = x/b$ if $p = 1/2$ and $f(x) = \frac{(\frac{1-p}{p})^x - 1}{(\frac{1-p}{p})^b - 1}$ if $p \neq 1/2$. Letting $b \to \infty$ we get that the probability of ruin is 1 if $p \leq 1/2$ and $1 - (\frac{1-p}{p})^x$ if $p > 1/2$.

$\Omega = [0, 1)$ **as a model for symmetric random walk.** Each point ω in the real interval $[0, 1)$ has a binary expansion $\omega = 0.x_1 x_2 x_3 \cdots$ where $x_i \in \{0, 1\}$. It is unique if we disallow $111111\dots$ at the end. The random walk can be written

$$x_n(\omega) = \sum_{i=1}^{n} R_i(\omega)$$

where

$$R_i(\omega) = 2x_i - 1$$

is the ith Rademacher function. If the random walk is symmetric then the length $|b-a|$ is exactly the probability of the set of random walk paths corresponding to ω in the interval $[a,b)$ (or (a,b), $(a,b]$, $[a,b]$, endpoints having null probability). Clearly this will also be true for unions of such intervals. Lets call $P(A)$ the set function so obtained. Clearly $P(\emptyset)=0$, $P([0,1)) = 1$ and $P(A^C) = 1 - P(A)$ and $P(A \cup B) = P(A) + P(B)$ for $A \cap B = \emptyset$ whenever $P(A)$ and $P(B)$ make sense. Note that our P has also the special property that it is invariant under the group action $A \mapsto A + x = \{y + x : y \in A\}$ with addition mod 1. Lets call the class of subsets A for which $P(A)$ makes sense $\mathcal{B}$. By induction, if A is the finite union of disjoint $A_i \in \mathcal{B}$ then $A \in \mathcal{B}$ and $P(A) = \sum P(A_i)$. In fact one can check that if A is the countable union of disjoint sets in $\mathcal{B}$ and A itself is in $\mathcal{B}$ then $P(A) = \sum P(A_i)$.

Every subset of $[0,1)$ corresponds to some set of random walk paths and it is very tempting to believe that such a set should have some sensible probability. Unfortunately this is not the case as the following example shows.

A set with no probability. Consider the equivalence relation on $[0,1)$ given by $x \sim y$ if $x - y$ is rational. Each element of $[0,1)$ lies in exactly one equivalence class. Choose one element from each equivalence class to make a set A. Then $A+q$, q rational are all disjoint and their union is $[0,1)$. But $P(A+q) = P(A)$ for all q by translation invariance. But then $\sum_{q \text{ rational}} P(A+q) = \sum_{q \text{ rational}} P(A) = 1$ a contradiction.

Probability spaces. A *probability space* is a triple $(\Omega, \mathcal{F}, P)$ where Ω is a set, $\mathcal{F}$ is a *$\sigma-$field (=$\sigma-$algebra)* , i.e. a collection of subsets of Ω called *events* containing $\emptyset$ and closed under complements and countable unions, and P is a probability measure, i.e. a countably additive set function defined on $\mathcal{F}$ (a *measure*) with $P(\Omega) = 1$.

Countable additivity means that if $A_i \in \mathcal{F}$, $i = 1, 2, \ldots$ are disjoint then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$. That probabilities should be countably additive is obvious. On the other hand, if $A_1 \supset A_2 \supset \cdots$ is a decreasing sequence of sets in $\mathcal{F}$ going to zero in the sense that $\cap_{i=1}^{\infty} A_i = \emptyset$ then we certainly expect that $P(A_n) \to 0$. This together with finite additivity is equivalent to countable additivity.

The smallest $\sigma-$algebra on $[0,1)$ (or R) containing the open sets is called the *Borel sets* and the measure corresponding to length described above is called *Lebesgue measure.*

Random variables. A *random variable* is a real valued function X on Ω, i.e. for each $\omega \in \Omega$ there is exactly one real number $X(\omega)$, which is measurable with respect to $\mathcal{F}$, i.e. for any Borel set B the set $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ is in $\mathcal{F}$.

In particular the *distribution function* $F(x) = P(X \leq x)$ is meaningful and for any Borel set B the probability $P(X \in B)$ can, in principle, be computed from $F(x)$.

The collection $\{X^{-1}(B),\ B \textit{ a Borel set}\}$ of subsets of Ω is a σ-field called *the σ-field generated by X*. It is the smallest σ-field with respect to which X is measurable. If $X_1, \ldots, X_n$ are random variables then the σ-field generated by all of them is just the smallest σ-field with respect to which they are all measurable.

Filtrations. Consider the case of a simple random walk. The sample space is the set of paths

$$\Omega = \{(x_1, x_2, \ldots) : x_i - x_{i-1} = \pm 1\}.$$

The coordinate functions $x_1, x_2, \ldots$ are each random variables. Let $\mathcal{F}_n$ be the $\sigma-$field generated by $x_1, \ldots, x_n$. What it means is that if two paths differ only after the nth time step then there are not two disjoint subsets of $\mathcal{F}_n$ each containing one of the paths while if they differ at some stage on or before the nth time step then there are two such paths. In information theoretic language one says they are indistinguishable or distinguishable with the information contained in $\mathcal{F}_n$ in the respective cases. In terms of the model $[0, 1)$ of the random walk, $\mathcal{F}_1 = \{\emptyset, [0, 1/2), [1/2, 1), [0, 1)\}$, $\mathcal{F}_2$ consists of dyadic intervals of order 2 and all their unions, etc.

A *filtration* is a ordered collection of sub-σ-fields of a $\sigma-$field, e.g. $\mathcal{F}_n \subset \mathcal{F}$ all $\sigma-$fields $n = 1, 2, \ldots$ with $\mathcal{F}_i \subset \mathcal{F}_{i+1}$.

The example above makes clear the role of measurability. Any function $X(\omega)$ which involves more than the first n steps of the random walk is not going to be measurable with respect to $\mathcal{F}_n$.

Expectation. The *expectation* of a random variable X on $(\Omega, \mathcal{F}, P)$ is the integral

$$E[X] = \int_\Omega X(\omega) dP(\omega)$$

which is defined for positive random variables as the limit

$$\lim_{n\to\infty} \sum_{i=-\infty}^{\infty} \frac{i}{n} P(\frac{i}{n} < X \leq \frac{i+1}{n}).$$

For general random variables $E[X] = E[X_+] - E[X_-]$ as long as both are not infinite (here $X_+ = max(X, 0)$ and $X_- = max(-X, 0)$ are the positive and negative parts of X.) Note that the limit can be written in terms of the distribution function,

$$\lim_{n\to\infty} \sum_{i=0}^{\infty} \frac{i}{n} [F(\frac{i+1}{n}) - F(\frac{i+1}{n})]$$

which is called the Riemann-Stieltjes integral

$$E[X] = \int x dF(x).$$

If F is differentiable then $f = F'$ is called the *density* function and for any $g(x)$, $\int g(x) dF(x) = \int g(x) f(x) dx$. The *variance* is given by

$$Var(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2 = \int (x - \mu)^2 dF(x)$$

where $\mu = E[X]$.

Let $p \geq 1$. The collection of random variables on $(\Omega, \mathcal{F}, P)$ for which $E[|X|^p] < \infty$ is called $L^p(\Omega, \mathcal{F}, P)$ or just L^p when there is no confusion. For example, L^1 means random variables with finite expectation, and L^2 means random variables with finite variance. Since

P is a probability measure it is easy to check that the L^p spaces are decreasing in p. Each L^p is a Banach space (complete normed vector space) with the norm $\|X\|_p = (E[|X|^p])^{1/p}$. L^2 in particular is also a Hilbert space (complete inner product space) with the inner product $\langle X, Y\rangle = E[X \cdot Y]$. In a Hilbert space $\frac{\langle X,Y\rangle}{\|X\|\|Y\|}Y$ is the projection of X onto Y, but in a Banach space you don't have that.

The *characteristic function* of X is the Fourier transform

$$\phi(z) = E[e^{izX}] = \int e^{izx} dF(x).$$

Name	*Density*	*Characteristic function*
Normal	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$e^{i\mu z - \frac{z^2\sigma^2}{2}}$
Uniform	$\frac{1}{a}\chi_{[0,a]}$	$\frac{e^{iaz}-1}{iaz}$
Cauchy	$\frac{1}{\pi}\frac{a}{a^2+x^2},\ a > 0$	$e^{-a\|z\|}$
Poisson	$\sum_{n=0}^{\infty}\frac{\lambda^n e^{-\lambda}}{n!}\delta_n$	$e^{\lambda(e^{iz}-1)}$
Cantor	No	$\lim_{n\to\infty}\int e^{izx}d\mu_n(x)$

The Cauchy distribution is an example for which the expectation $\int x dF$ does not exist, i.e. Cauchy distributed variables are not in L^p for any $p \geq 1$.

The Poisson distribution does not really have a density but one can pretend it does to some extent. The δ_n refer to Dirac delta "functions" at the points n. They add and subtract just like functions, but cannot be multiplied. The Cantor distribution is even worse. It does not have a density in any sense. It is defined as the limit of the following sequence of distributions, $d\mu_1 = \chi_{[0,1]}dx$, $d\mu_2 = \frac{3}{2}\chi_{[0,1/3]\cup[2/3,1]}dx$, $\mu_3 = (\frac{3}{2})^2\chi_{[0,1/9]\cup[2/9,3/9]\cup[6/9,7/9]\cup[8/9,1]}dx, \ldots$

Radon-Nikodym. If μ and ν are two measures with respect to a $\sigma-$field $\mathcal{F}$, then we say μ is absolutely continuous with respect to ν if $\nu(A) = 0$ implies that $\mu(A) = 0$ for any $A \in \mathcal{F}$. Equivalently μ is absolutely continuous with respect to ν if there exists a measurable function f with $\int |f| d\nu < \infty$ and for any integrable function g,

$$\int g(x) d\mu(x) = \int g(x) f(x) d\nu(x).$$

Clearly the notion of a probability distribution having a density is the same as the corresponding probability measure on R being absolutely continuous with respect to Lebesgue measure.

Joint distributions. When you only have one random variable X the distribution function $F(x) = P(X \leq x)$ pretty much tells you the whole story. On the other hand if you have two random variables X_1 and X_2 with distribution functions $F_1(x) = P(X_1 \leq x)$ and $F_2(x) = P(X_2 \leq x)$, then it is not true that F_1 and F_2 together describe the random vector (X_1, X_2). The *joint distribution function* is

$$F(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2).$$

Clearly this can be extended to three or more random variables. The *covariance* of X_1 and X_2 is

$$Cov(X_1, X_2) = E[X_1; X_2] = E[(X_1 - E[X_1])(X_2 - E[X_2])],$$

and the *correlation* is the normalized covariance $\frac{Cov(X_1,X_2)}{\sqrt{Var(X_1)Var(X_2)}}$. If $X = (X_1, \ldots, X_n)$ is a random vector, the *covariance matrix* C is the $n \times n$ matrix with entries $C_{ij} = Cov(X_i, X_j)$.

Gaussian distributions. If X is a random vector with distribution function

$$F(x) = \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} \frac{1}{\sqrt{(2\pi)^n det(C)}} e^{-\frac{(x-m)\cdot C^{-1}(x-m)^T}{2}} dx_1 \cdots dx_n$$

where $m \in R^n$, $x = (x_1, \ldots, x_n)$ and C is a symmetric positive definite matrix then X is said to have a *Gaussian* or *multivariate normal* distribution. *Positive definite* means $x \cdot Cx^T > 0$ for any vector x. It does not suffice for X_1 and X_2 to both be Gaussian for $X = (X_1, X_2)$ to be Gaussian.

In a lot of applied work one computes the mean (vector) and covariance of a collection of random variables, and then pretends one has a Gaussian vector with those characteristics. Such L^2 theory is analogous to linearising everything in sight when one studies differential equations.

Independence. Two random variables are said to be independent if their joint distribution is the product of their respective distributions, $F(x_1, x_2) = F_1(x_1)F_2(x_2)$. Equivalently $P(X_1 \in B_1, X_2 \in B_2) = P(X_1 \in B_1)P(X_2 \in B_2)$ for any Borel sets B_1 and B_2. It is also the same thing to say that $E[f_1(X_1)f_2(X_2)] = E[f_1(X_1)]E[f_2(X_2)]$ for a large enough class of functions f_i. For example, e^{izx}, $z \in R$ is enough. It is certainly not enough to have $E[X_1X_2] = E[X_1]E[X_2]$.

If $X = (X_1, X_2)$ is Gaussian then X_1 and X_2 are independent if and only if they are uncorrelated i.e. $C_{12} = E[(X_1 - E[X_1])(X_2 - E[X_2])] = 0$.

Convolutions. *Problem.* If X and Y are independent random variables with distribution functions $F(x) = P(X \leq x)$ and $G(y) = P(Y \leq y)$ what is the distribution function of the random variable $X + Y$?

Solution. We can write $P(X + Y \leq z) = \sum_{i=-\infty}^{\infty} P(X + Y \leq z \text{ and } \frac{i}{n} < Y \leq \frac{i+1}{n})$. In terms of conditional probability the last term is

$$\sum_{i=-\infty}^{\infty} P(X + Y \leq z \mid \frac{i}{n} < Y \leq \frac{i+1}{n})[G(\frac{i+1}{n}) - G(\frac{i}{n})].$$

Letting $n \to \infty$ we get

$$\int P(X + Y \leq z \mid Y = y) dG(y).$$

Since X is independent of Y, $P(X + Y \leq z \mid Y = y) = P(X \leq z - y) = F(z - y)$, so the distribution function of $X + Y$ is

$$F * G(z) = \int F(z - y) dG(y).$$

The wonderful thing about characteristic functions is that if X and Y are independent random variables with characteristic functions $\phi_X(z) = E[e^{izX}]$ and $\phi_Y(z) = E[e^{izY}]$ then $X + Y$ has characteristic function

$$\phi_{X+Y}(z) = E[e^{iz(X+Y)}] = E[e^{izX}e^{izY}] = E[e^{izX}]E[e^{izY}] = \phi_X(z)\phi_Y(z).$$

PROBLEMS.

1. In the Russian *sportloto* each player on Monday pays 1 ruble and chooses six numbers between 1 and 49. The six numbers are all different, and there is no order. On Tuesday the master of ceremonies also chooses his six numbers. If your six numbers match those of the master of ceremonies, you win 100 million rubles payable Wednesday evening. Everybody knows Yeltsin will devalue the ruble by 50% sometime during the week, with each day equally likely. What is the expected return in the sportloto? Is it worth playing?
2. A needle one inch long is thrown onto a hardwood floor where the lines between the slats are exactly one inch apart. What is the probability that the needle does not cross one of the lines?
3. i. Let X be a random variable with density function f and let g be a nice function. Find the density of the random variable $g(X)$ when it makes sense, and discuss what properties g should have to make it exist.
   ii. Let X have a Gaussian distribution with mean 0 and variance σ^2. Find the density function of e^X.
   iii. Let $X_1, \ldots, X_n$ be independent standard Gaussian random variables (mean 0, variance 1). Find the density of $X_1^2 + \cdots + X_n^2$.
   iv. Let X be a random vector with a Gaussian distribution. Find a linear transformation of X which has a *standard* Gaussian distribution (mean $\mathbf{0}$, covariance $I =$ the identity matrix).
4. i. Let $a > 0$ and X be a random variable with an *exponential* density $f(x) = ae^{-ax}\chi_{[0,\infty)}$. Calculate the probability $P(x_1 \leq X \leq x_2)$ for any x_1, x_2. Find the mean and the *median* a defined by $P(X \leq a) = 1/2$. Finally calculate the conditional probability $P(X \geq a + b \mid X \geq b)$.
   ii. Let $X_1, X_2, \ldots$ be independent exponentially distributed random variables with mean μ. Fix $t > 0$ and let $N(t)$ be the unique number such that $X_1 + \cdots + X_{N(t)} \leq t$ and $X_1 + \cdots + X_{N(t)+1} > t$. Find the distribution of the random variable $N(t)$.