A collection $\mathcal{F}$ of subsets of Ω is a field (i) if $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$ and $A \cap B \in \mathcal{F}$, (ii) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$, (iii) $\emptyset \in \mathcal{F}$.
σ-field: (i) $\emptyset \in \mathcal{F}$, (ii) if $A_1, A_2, \ldots \in \mathcal{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$, (iii) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$.

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad (A \cup B)^c = A^c \cap B^c \quad (A \cap B)^c = A^c \cup B^c \quad A \cup B = (A \cap B^c) \cup (A^c \cap B) \cup (A \cap B)$
$A \cap B^c = A - (A \cap B)$

$P(\emptyset) = 0 \quad P(\Omega) = 1 \quad P(A^c) = 1 - P(A) \quad$ If $A \subseteq B$ then $P(A) \leq P(B)$.

$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad P(A \cap B^c) = P(A) - P(A \cap B) = P(A - B)$

If A_i are disjoint then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$. In general $P\left(\bigcup_i A_i\right) \leq \sum_i P(A_i)$.

$P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} \quad P(A) = P(A \mid B)P(B) + P(A \mid B^c)P(B^c) \quad P(C \mid A \cap B)P(A \cap B) = P(C \mid A \cap B)P(B \mid A)P(A)$

$P(ABC) = P(A \cap B \cap C) = \dfrac{P(A \cap B \cap C)}{P(A \cap B)} P(A \cap B) = P(C \mid A \cap B)P(A \cap B)$

$P(C) = P(C \mid A \cap B)P(A \cap B) + P(C \mid A \cap B^c)P(A \cap B^c) + P(C \mid A^c \cap B)P(A^c \cap B) + P(C \mid A^c \cap B^c)P(A^c \cap B^c)$

Bayes' formula $P(A \mid X) = \dfrac{P(X \mid A)P(A)}{P(X \mid A)P(A) + P(X \mid B)P(B) + P(X \mid C)P(C)}$

Independence $\quad P(A \cap B) = P(A)P(B) \quad P(A \mid B) = P(A)$

Conditional independence $\quad P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$

"lim inf and lim sup are events." $\quad \liminf A_k = \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} A_j \quad \limsup A_k = \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} A_j$

$\cap_{k=1}^{\infty} A_k = \{\omega : \omega \in A_k \text{ for all } k\}$
$\liminf A_k = \{\omega : \omega \in A_k \text{ for all } k > n \text{ for some } n\}$
$\limsup A_k = \{\omega : \omega \in A_k \text{ for infinitely many } k\}$
$\cup_{k=1}^{\infty} = \{\omega : \omega \in A_k \text{ for at least one } k\}$

$$\bigcap_{k=1}^{\infty} A_k \subseteq \liminf A_k \subseteq \limsup A_k \subseteq \bigcup_{k=1}^{\infty} A_k$$

$\liminf(A_n \cap B_n) = (\liminf A_n) \cap (\liminf B_n) \quad (\limsup A_k)^c = \liminf(A_k^c)$
$\limsup(A_n \cup B_n) = (\limsup A_n) \cup (\limsup B_n) \quad \liminf A_n \subseteq \liminf(A_n \cup B_n)$

Borel-Cantelli lemma: Let $A_k \in \mathcal{F}$ such that $\sum_{k=1}^{\infty} P(A_k) < \infty$. Then $P(\limsup A_k) = 0$.

i.o. "infinitely often" is the sane as lim sup

Increasing sequence of events: $A_1 \subseteq A_2 \subseteq \ldots$ then $\lim_{i \to \infty} A_i = \bigcup_{i=1}^{\infty} A_i$

Decreasing sequence of events: $A_1 \supseteq A_2 \supseteq \ldots$ then $\lim_{i \to \infty} A_i = \bigcap_{i=1}^{\infty} A_i$

Continuity theorem: $\lim_{k \to \infty} P(A_k) = P(\lim_{k \to \infty} A_k)$

$I_A = 1 - I_{A^c} \quad I_{A \cap B} = I_A \cdot I_B \quad I_{A \cup B} = \max(I_A, I_B) = I_A + I_B - I_A I_B \quad I_A^2 = I_A$

If F is a distribution function then (i) $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$, (ii) if $x < y$ then $F(x) \leq F(y)$, (iii) $F(x)$ is right continuous, $\lim_{h \downarrow 0} F(x + h) = F(x)$. Converse is true, any function that satisfies (i), (ii), and (iii) is a distribution function of some random variable. If $F(x) \leq F(y)$ then $\{\omega : X(\omega) \leq x\} \subseteq \{\omega : Y(\omega) \leq y\}$.

X is discrete if it can have only countably many values.

If $f(x)$ is a probability mass function then (i) $0 \leq f(x) \leq 1$ for all x, (ii) $\sum_x f(x) = 1$. Converse is true, any function that satisfies (i) and (ii) is a p.m.f. of some random variable.

$f(x) = P(X = x) = F(x) - \lim_{h \downarrow 0} F(x - h) \quad F(x) = P(X \leq x) = \sum_{y \leq x} f(y)$

X is a continuous random variable if there is an $f(x)$ such that (i) $f(x) \geq 0$ for all x, (ii) $F(x) = \int_{-\infty}^{x} f(y)\, dy$. The function $f(x)$ is called the density of X. Density of a random variable has to satisfy two conditions (i) $f(x) \geq 0$, (ii) $\int_{-\infty}^{\infty} f(x) = 1$.
Remark $f(x) = dF/dx$. "A density is not a probability like mass is."

$P(X \le x) = F(x) \quad P(X > x) = 1 - F(X) \quad P(X < x) = \lim_{h \downarrow 0} F(x-h) \quad P(X \ge x) = 1 - \lim_{h \downarrow 0} F(x-h)$

$P(a < X < b) = \lim_{h \downarrow 0} F(b-h) - F(a) \quad P(a \le X \le b) = F(b) - \lim_{h \downarrow 0} F(a-h) \quad P(X = x) = F(x) - \lim_{h \downarrow 0} F(x-h)$

$$\sum_{k=1}^{\infty} \frac{1}{2^k} = 1 \quad \sum \frac{1}{k} = \infty \quad \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6} \quad \sum_{k=1}^{\infty} \frac{1}{k^4} = \frac{\pi^4}{90} \quad \sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_{k=0}^{n} p^k = \frac{1-p^{n+1}}{1-p} \quad \sum_{k=0}^{\infty} p^k = \frac{1}{1-p} \quad \sum_{k=1}^{\infty} p^k = \frac{p}{1-p} \quad \sum_{k=1}^{\infty} \frac{1}{k(k+1)} = 1$$

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x \quad \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n} \left(\frac{k}{n}\right)^m = \int_0^1 x^m \, dx = \frac{1}{m+1} \quad \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

$$\log ab = \log a + \log b \quad \log(a/b) = \log a - \log b \quad \log a^b = b \log a \quad \log_2 a = \frac{\log a}{\log 2} \quad \lim_{x \to -\infty} \tan^{-1} x = -\frac{\pi}{2}$$

$$\int e^{ax} = \frac{e^{ax}}{a} \quad \int x e^{ax} = \frac{e^{ax}(ax-1)}{a^2} \quad \int \frac{du}{a^2+u^2} = \frac{1}{a} \tan^{-1} \frac{u}{a} \quad \int \frac{du}{u} = \log |u|$$

Expected value $EX = \sum x f(x) \quad Eg(x) = \sum g(x) f(x) \quad E(aX+b) = aEX + b$

Theorem (i) $E1 = 1$, (ii) $E(aX + bY) = aEX + bEY$, (iii) if $X \ge 0$ then $EX \ge 0$.

$Var\,X = E((X - EX)^2) \quad Var\,X = E(X^2) - (EX)^2 \quad Var(aX + b) = a^2\,Var\,X$

If X and Y are independent then $E(XY) = E(X)E(Y)$. Converse is not true in general.

$Cov(X,Y) = E[(X-EX)(Y-EY)] = E(XY) - E(X)E(Y)$ If $Cov(X,Y) = 0$ then X and Y are uncorrelated.

If X and Y are uncorrelated then $Var(X+Y) = Var\,X + Var\,Y$

Correlation coefficient $\rho(X,Y) = \dfrac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}}$

$E(X(X-1)) = E(X^2) - EX \quad Var\,X = EX^2 - (EX)^2 = E(X(X+1)) - EX - (EX)^2$

Bernoulli(p) — $f(1)$ is the probability of heads, $f(0)$ is the probability of tails.
Binomial(n, p) — $f(k)$ is the probability of k heads in n tosses.
Poisson(λ)
Geometric(p) — $f(k)$ is the probability that it takes k tosses to get a head.
NegativeBinomial(n, p) — $f(k)$ is the probability of k tosses to get n heads.
Hypergeometric(N, b, n) — Urn with N balls, b are black and $N - b$ are white. Draw n balls without replacement. $f(k)$ is the probability of k black balls.

Joint distribution function $F(x,y) = P(X \le x \,\text{and}\, Y \le y)$ Joint mass function $f(x,y) = P(X = x \,\text{and}\, Y = y)$

Joined distribution function properties (i) $\lim_{x,y \to -\infty} F(x,y) = 0$, $\lim_{x,y \to \infty} F(x,y) = 1$, (ii) if $(x_1, y_1) \le (x_2, y_2)$ then $F(x_1, y_1) \le F(x_2, y_2)$, (iii) continuous from above, $\lim_{u,v \downarrow 0} F(x+u, y+v) = F(x,y)$. Property (ii) means $F(x,y)$ is nondecreasing. Another way of putting it is, for any $a_1 < a_2$, $b_1 < b_2$, it must be true that $F(a_2, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, b_1) \ge 0$. Any function that satisfies (i), (ii), and (iii) is a joined distribution function.

$\lim_{y \to \infty} F(x,y) = F_X(x) = P(X \le x) \quad \lim_{x \to \infty} F(x,y) = F_Y(y) = P(Y \le y)$

For random vectors, $x \le y$ means that $x_1 \le y_1$ and $x_2 \le y_2$.

$F(x,y)$ is jointly continuous if $F(x,y) = \displaystyle\int_{u=-\infty}^{x} \int_{v=-\infty}^{y} f(u,v)\,du\,dv$

Discrete random variables X and Y are independent iff $f(x,y)$ can be factored, that is, $f(x,y) = g(x)h(y)$.

For any intervals A and B, if $P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$ then A and B are independent.

Transformation theorem: If X and Y are independent, then $g(X)$ and $h(Y)$ are also independent.

$P(X \in A) = \displaystyle\sum_{x \in A} f(x)$

$P(a_1 < X \le a_2,\, b_1 < Y \le b_2) = F(a_2, b_2) - F(a_1, b_2) - F(a_2, b_1) + F(a_1, b_1)$

"We know that for a decreasing sequence of sets, the limit exists and it's the empty set."