1 Sets

$(A \cup B) \cup C = A \cup (B \cup C)$
$(A \cup B)^C = A^C \cap B^C$
Definition: σ-field
$\mathcal{F}$ subset of Ω is a $\sigma - field$, if
(a) $0 \in \mathcal{F}$
(b) if $A_1, A_2, ... \in \mathcal{F}$ then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$
(c) if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$

2 Probability

$\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$
If $B \supseteq A$ then $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \backslash A) \geq \mathbb{P}(A)$
$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
More generally:
$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_i \mathbb{P}(A_i) - \sum_{i<j} \mathbb{P}(A_i \cap A_j) - \sum_{i<j<k} \mathbb{P}(A_i \cap A_j \cap A_k) + \cdots + (-1)^{n+1} \mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n)$
Lemma 5, p. 7: Let $A_1 \subseteq A_2 \subseteq \ldots$, and write A for their limit:
$A = \bigcup_{i=1}^{\infty} A_i = \lim_{i\to\infty} A_i$ then $\mathbb{P}(A) = \lim_{i\to\infty} \mathbb{P}(A_i)$
Similarly, $B_1 \supseteq B_2 \supseteq B_3 \supseteq \ldots$, then $B = \bigcap_{i=1}^{\infty} B_i = \lim_{i\to\infty} B_i$
satisfies $\mathbb{P}(B) = \lim_{i\to\infty} \mathbb{P}(B_i)$
Multiplication rule
$P(A, B) = P(A)P(B \mid A)$
Conditional Probability
$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
$P(A \mid B, C, ...) = \frac{P(A,B,C,...)}{P(B,C,...)}$
Bayes formula
$\mathbb{P}(A|B) = \mathbb{P}(B|A)\mathbb{P}(A)\mathbb{P}(B)$
Total probability
$P(A) = P(A \mid B)P(B)$
$+ P(A \mid B^C)P(B^C)$
$P(A) = \sum_{i=1}^n P(A \mid B_i)P(B_i)$
Definition 1, p. 13:
A family $\{A_i : i \in I\}$is independent if:
$\mathbb{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathbb{P}(A_i)$ For all finite subset J of I

3 Random Variable

Lemma 11, p. 30:
Let F be a distribution function of X, then
(a) $\mathbb{P}(X > x) = 1 - F(x)$
(b) $\mathbb{P}(x < X \leq y) = F(y) - F(x)$
(c) $F(X = x) = F(x) - \lim_{y\to x} F(y)$
Marginal distribution:
$\lim_{y\to\infty} F_{X,Y}(x, y) = F_X(x)$
$\lim_{x\to\infty} F_{X,Y}(x, y) = F_Y(y)$
Lemma 5, p. 39:
The joint distribution function $F_{X,Y}$ of the random vector (X, Y) has the following properties:
$\lim_{x,y\to-\infty} F_{X,Y}(x, y) = 0$
$\lim_{x,y\to\infty} F_{X,Y}(x, y) = 1$
if $(x_1, y_1) \leq (x_2, y_2)$ then $F_{X,Y}(x_1, y_1) \leq F_{X,Y}(x_2, y_2)$
$F_{X,Y}$ is continuous from above, in that: $F_{X,Y}(x + u, y + v) \to F_{X,Y}(x, y)$ as $u, v \to 0$
Theorem:
If X and Y are independent and g, $h : \mathbb{R} \to \mathbb{R}$, then $g(X)$ and $h(Y)$ are independent too.
Definition:
The *expectation* of the random variable X is:
$\mathbb{E}(X) = \sum_{x:f(x)>0} xf(x)$
Lemma:
If X has mass function f and $g : \mathbb{R} \to \mathbb{R}$, then:
$\mathbb{E}(g(X)) = \sum_x g(x)f(x)$
Continous counterpart:
$E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x)dx$
Definition:
If k is a positive integer, the k:th moment m_k of X is defined $m_k = \mathbb{E}(X^k)$. The k:th central moment is $\sigma_k = \mathbb{E}((X - m_1)^k)$
Theorem:
The expectation operator $\mathbb{E}$:
(a) If $X \geq 0$ then $\mathbb{E}(X) \geq 0$
(b) If $a, b \in \mathbb{R}$ then $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$
Lemma:
If X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
Definition:
X and Y are *uncorrelated* if $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$
Theorem:
For random variables X and Y
(a) $\text{Var}(aX) = a^2\text{Var}(X)$ for $a \in \mathbb{R}$
(b) $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ if X and Y are uncorrelated.
Indicator function:
$\mathbb{E}I_A = \mathbb{P}(A)$

distribution function
$F : R- > [0, 1] : F(x) = P(X \leq x)$
mass function

3.1 Distribution functions

Constant variable
$X(\omega) = c$: $F(X) = \sigma(x - c)$
Bernoulli distribution $Bern(p)$
A coin is tossed one time and shows head with probability p with $X(H) = 1$ and $X(T) = 0$
$F(X) = 0 \quad x < 0$
$F(X) = 1 - p \quad 0 \leq x < 1$
$F(X) = 1 \quad x \geq 1$
$\mathbb{E}(X) = p, \quad \text{Var}(X) = p(1 - p)$
Binomial distribution $bin(n, k)$
A coin is tossed n times and a head turns up each time with probability p. The total number of heads is discribed by:
$f(k) = \binom{n}{k} p^k q^{n-k}$
$\mathbb{E}(X) = np, \quad \text{Var}(X) = np(1 - p)$
Poisson distribution
$f(k) = \frac{\lambda^k}{k!} \exp(-\lambda)$
$\mathbb{E}(X) = \text{Var}(X) = \lambda$
Geometric distribution
Independent Bernoulli trials are performed. Let W be the waiting time before the first succes occurs. Then
$f(k) = P(W = k) = p(1 - p)^{k-1}$
$\mathbb{E}(X) = 1/p, \quad \text{Var}(X) = (1 - p)/p^2$
negative binomial distribution
Let W_r be the waiting time before the r:th success. Then
$f(k) = P(W_r = k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r}$
$k = r, r + 1$
$\mathbb{E}(X) = \frac{pr}{1-p}, \quad \text{Var} = \frac{pr}{(1-p)^2}$
Exponential distribution:
$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$
$\mathbb{E}(X) = 1/\lambda, \quad \text{Var}(X) = 1/\lambda^2$
Normal distribution:
$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$
$\mathbb{E}(X) = \mu, \quad \text{Var}(X) = \sigma^2$
Cauchy distribution:
$f(x) = \frac{1}{\pi(1+x^2)}$ (no moments!)

3.2 Dependence

Joint distribution:
$F(x, y) = \mathbb{P}(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v)dvdu$
Lemma:
The random variables X and Y are independent if and only if
$f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$
$F_{X,Y}(x, y) = F_X(x)F_Y(y)$ for all $x, y \in \mathbb{R}$
Marginal distribution:
$F_X(x) = \mathbb{P}(X \leq x) = F(x, \infty) = \int_{-\infty}^{x} \left(\int_{-\infty}^{-\infty} f(u, y\right) dydx$
Marginal densities: $f_X(x) = \mathbb{P}\left(\bigcup_y(\{X = x\} \cap \{Y = y\})\right) = \sum_y \mathbb{P}(X = x, Y = y) = \sum_y f_{X,Y}(x, y)$
$f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$
Lemma:
$\mathbb{E}(g(X, Y)) = \sum_{x,y} g(x, y)f_{X,Y}(x, y)$
Definition:
$\text{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right]$

3.3 Conditional distributions

Definition:
The *conditional distribution* of Y given $X = x$ is:
$F_{Y|X}(y|x) = \mathbb{P}(Y \leq y|X \leq x)$
$= \int_{-\infty}^{x} \frac{f(v,y)}{f_Y(y)} dv, \quad \{y : f_Y(y) > 0\}$
Theorem: Conditional expectation

$\psi(X) = \mathbb{E}(Y \mid X), \quad \mathbb{E}(\psi(X)) = \mathbb{E}(Y)$
$\mathbb{E}(\psi(X)g(X)) = \mathbb{E}(Yg(X))$

3.4 Sums of random variables

Theorem:
$\mathbb{P}(X+Y=z) = \sum_x f(x, z-x)$
If X and Y are independent, then
$\mathbb{P}(X+Y=z) = f_{X+Y}(z) = \sum_x f_X(x)f_Y(z-x) = \sum_y f_X(z-y)f_Y(y)$

3.5 Multivariate normal distribution:

$f(\mathbf{x}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{V}^{-1}(\mathbf{x}-\mu)\right)}{\sqrt{(2\pi)^n \det(\mathbf{V})}}$
$\mathbf{X} = (X_1, X_2, \ldots, X_n)^T$
$\mu = (\mu_1, \ldots, \mu_n), \quad \mathbb{E}(X_i) = \mu_i$
$\mathbf{V} = (v_{ij}), \quad v_{ij} = \text{Cov}(X_i, X_j)$

4 Generating functions

Definition: Generating function
The generating function of the random variable X is defined by:
$G(s) = \mathbb{E}(s^X)$
Example: Generating functions
Constant: $G(s) = s^c$
Bernoulli: $G(s) = (1-p) + ps$
Geometric: $\frac{ps}{1-s(1-p)}$
Poisson: $G(s) = e^{\lambda(s-1)}$
Theorem: expectation $\leftrightarrow G(s)$
(a) $\mathbb{E}(s) = G'(1)$
(b) $\mathbb{E}(X(X-1)\ldots(X-k+1)) = G^{(k)}(1)$
Theorem: independance
X and Y are independent, iff
$G_{X+Y}(s) = G_X(s)G_Y(s)$

4.1 Characteristic functions

Definition: moment generating function
$M(t) = \mathbb{E}(e^{tX})$
Definition: characteristic function
$\Phi(t) = \mathbb{E}(e^{itX})$
Theorem: independance
X and Y are independent iff
$\Phi_{X+Y}(t) = \Phi_X(t)\Phi_Y(t)$
Theorem: $Y = aX + b$
$\Phi_Y(t) = e^{itb}\Phi_X(at)$
Definition: joint characteristic function
$\Phi_{X,Y}(s,t) = \mathbb{E}(e^{isX}e^{itY})$
Independent if:
$\Phi_{X,Y}(s,t) = \Phi_X(s)\Phi_Y(t)$ for all s and t
Theorem: moment gf $\leftrightarrow$ charact. fcn
Examples of characteristic functions:
Ber(p): $\Phi(t) = q + pe^{it}$
Bin distribution, bin(n,p): $\Phi_X(tt) = (q+pe^{it})^n$
Exponential distr. $\Phi(t) = \int_0^\infty e^{itx}\lambda e^{-\lambda x}dx$
Cauchy distr: $\Phi(t) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{e^{itx}}{(1+x^2)}dx$
Normal distr, $\mathcal{N}(0,1)$: $\Phi(t) = \mathbb{E}(e^{itX}) = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}\exp(itx - \frac{1}{2}x^2)dx$
Corollary:
Random variables X and Y have the same characteristic function if and only if they have the same distribution function. Theorem: Law of large Numbers
Let $X_1, X_2, X_3\ldots$ be a sequence of iid r.v's with finite mean μ.
Their partial sums $S_n = X_1 + X_2 + \cdots + X_n$ satisfy $\frac{1}{n}S_n \xrightarrow{D} \mu$ as $n \to \infty$
Central Limit Theorem:
Let $X_1, X_2, X_3\ldots$ be a sequence of iid r.v's with finite mean μ and finite non-zero σ^2, and let $S_n = X_1 + X_2 + \cdots + X_n$ then
$\frac{(S_n - n\mu)}{\sqrt{n\sigma^2}} \xrightarrow{D} \mathcal{N}(0,1)$ as $n \to \infty$

5 Markov chains

Definiton Markov Chain
$P(X_n = s \mid X_0 = x_0, \ldots X_n - 1 = x_n 1) = P(X_n \mid X_n - 1 = x_n - 1)$
Definition homogenous chain
$P(X_n + 1 = j \mid X_n = i) = P(X_1 = j \mid X_0 = i)$
Definition transition matrix
$\mathbf{P} = (p_{ij})$ with
$p_{ij} = P(X_{n+1} = j \mid X_n = i)$
Theorem: P stochastic matrix
(a) $\mathbf{P}$ has non-negative entries
(b) $\mathbf{P}$ has row sum equal 1
n-step transition
$p_{ij}(m, m+n) = P(X_{m+n} = j \mid X_m = i)$
Theorem Chapman Kolmogorov
$p_{ij}(m.m+n+r) =$
$\sum_k p_{ik}(m, m+n)p_{kj}(m+n, m+n+r)$
so
$\mathbf{P}(m, m+n) = \mathbf{P}^n$
Definiton: mass funtion
$\mu_i^{(n)} = P(X_n = i)$
$\mu^{(m+n)} = \mu(m)\mathbf{P}_n \Rightarrow \mu^{(n)} = \mu(0)\mathbf{P}^n$
Definition: persistent, transient
persistent:
$P(X_n = i$ for some $n \geq 1 \mid X_0 = i) = 1$
transient:
$P(X_n = i$ for some $n \geq 1 \mid X_0 = i) < 1$
Definition: first passage time
$f_{ij}(n) = P(X_1 \neq j, ., X_{n-1} \neq j, X_n = j \mid X_0 = i)$
$f_{ij} := \sum_{n=1}^{\infty} f_{ij}(n)$
Corollary: persistent, transient
State j is persistent if $\sum_n p_{jj}(n) = \infty$ and $\Rightarrow \sum_n p_{ij}(n) = \infty$ for all i
State j is transient if $\sum_n p_{jj}(n) < \infty$ and $\Rightarrow \sum_n p_{ij}(n) < \infty$ for all i
Theorem: Generating functions
$P_{ij}(s) = \sum_{n=0}^{\infty} s^n p_{ij}(n)$
$F_{ij}(s) = \sum_{n=0}^{\infty} s^n f_{ij}(n)$
then
(a) $P_{ii}(s) = 1 + F_{ii}(s)P_{ii}(s)$
(b) $P_{ij}(s) = F_{ij}(s)P_{ij}(s)$ if $i \neq j$
Definition: First visit time T_j
$T_j := min\{n \geq 1 : X_n = j\}$
Definition: mean recurrence time μ_i
$\mu_i := E(T_i \mid X_0 = i) = \sum_n n f_{ii}(n)$
Definition: null, non-null state
state i is null if $\mu_i = \infty$
state i is non-null if $\mu_i < \infty$
Theorem: nullness of a persistent state
A persistent state is null if and only if $p_{ii}(n) \to 0 (n \to \infty)$
Definition: period $d(i)$
The period $d(i)$ of a state i is defined by $d(i) = gcd\{n : p_{ii}(n) > 0\}$. We call i periodic if $d(i) > 1$ and aperiodic if $d(i) = 1$
Definition: Ergodic
A state is called ergodic if it is persistent, non-null, and aperiodic.
Definition: (Inter-)communication
$i \to j$ if $p_{ij}(m) < 0$ for some m
$i \leftrightarrow j$ if $i \to j$ and $j \to i$
Theorem: intercommunication
If $i \leftrightarrow j$ then:
(a) i and j have the same period
(b) i is transient iff j is transient
(c) i is null peristent iff j is null peristent
Definition: closed, irreducible
a set C of states is called:
(a) closed if $p_{ij} = 0$ for all $i \in C, j \notin C$
(b) irreducible if $i \leftrightarrow j$ for all $i, j \in C$
an absorbing state is a closed set with one state
Theorem: Decomposition
State space S can be partitioned as $T = T \cup C_1 \cup C_2 \cup \ldots$, T is the set of transient states, C_i irreducible, closed sets of persistent states
Lemma: finite S
If S is finite, then at least one state is persisten and all persistent states are non-null.

5.1 Stationary distributions

Definition: stationary distribution
π is called stationary distribution if
(a) $\pi_j \geq 0$ for all j, $\sum_j \pi_j = 1$
(b) $\pi = \pi P$, so $\pi_j = \sum_i \pi_i p_{ij}$ for all j
Theorem: existence of stat. distribution
An irreducible chain has a stationary distribution π iff all states are non-null persistent.
Then π is unique and given by $\pi_i = \mu_i^{-1}$
Lemma: $\rho_i(k)$
$\rho_i(k)$: mean number of visits of the chain to the state i between two successive visits to state k.
Lemma: For any state k of an irre-

ducible persistent chain, the vector $\rho(k)$ satisfies $\rho_i(k) < \infty$ for all i and $\rho(k) = \rho(k)P$

Theorem: irreducible, persistent

If the chain is irreducible and persistent, there exists a positive x with $x = xP$, which is unique to a multiplicative constant. The chain is non-null if $\sum_i x_i < \infty$ and null if $\sum_i x_i = \infty$

Theorem: transient chain if

s any state of an irreducible chain. The chain is transient iff there exists a non-zero solution $\{y_j : j \neq s\}$, with $|y_j| \leq 1$ for all j, to the equation:

$y_i = \sum_{j,j\neq s} p_{ij} y_j, \qquad i \neq s$

Theorem: persistent if

s any state of an irreducible chain on $S = \{0, 1, 2, ...\}$. The chain is persistent if there exists a solution $\{y_j : j \neq s\}$ to the inequalities

$y_i \geq \sum_{j,j\neq s} p_{ij} y_j, \qquad i \neq s$

Theorem: Limittheorem

For an irreducible aperiodic chain, we have that

$p_{ij}(n) \to \frac{1}{\mu_j}$ as $n \to \infty$ for all i and j

5.2 Reversibility

Theorem: Inverse Chain

Y with $Y_n = X_{N-n}$ is a Markov chain with $P(Y_{n+1} = j \mid Y_n = i) = (\frac{\pi_j}{\pi_i}) p_{ji}$

Definition: Reversible chain

A chain is called reversible if

$\pi_i p_{ij} = \pi_j p_{ji}$

Theorem: reversible → stationary

If there is a π with $\pi_i p_{ij} = \pi_j p_{ji}$ then π is the stationary distribution of the chain.

5.3 Poisson process

Definition: Poisson process

$N(t)$ gives the number of events in time t

Poisson process $N(t)$ in $S = \{0, 1, 2, ...\}$, if

(a) $N(0) = 0$; if $s < t$ then $N(s) \leq N(t)$

(b) $P(N(t + h) = n + m \mid N(t) = n) =$
$\lambda h + o(h)$ if $m = 1$
$o(h)$ if $m > 1$
$1 - \lambda h + o(h)$ if $m = 0$

(c) the emission per interval are independent of the intervals before.

Theorem: Poisson distribution

$N(t)$ has the Poisson distribution:

$P(N(t) = j) = \frac{(\lambda t)^j}{j!} e^{-\lambda t}$

Definition: arrivaltime, interarrivaltime

Arrival time: $T_n = inf\{t : N(t) = n\}$

Interarrivaltime: $X_n = T_n - T_{n-1}$

Theorem: Interarrivaltime

$X_1, X_2, ...$ are independent having exponential distribution Definition:

Birth process → Poisson process with intensities $\lambda_0, \lambda_1, \ldots$

Eq. Forward System of Equations:

$p'_{ij}(t) = \lambda_{j-1} p_{i,j-1}(t) - \lambda_j p_{ij}(t)$

$j \geq i, \quad \lambda_{-1} = 0, \quad p_{ij}(0) = \delta_{ij}$

Eq. Backward systems of equations:

$p'_{ij}(t) = \lambda_i p_{i+1,j}(t) - \lambda p_{ij}(t)$

$j \geq i \quad p_{ij}(0) = \delta_{ij}$

Theorem:

The forward system has a unique solution which satisfies the backward equation.

5.4 Continuous Markov chain

Definition: Continuous Markov chain

X is continuous Markov chain if:

$P(X(t_n) = j \mid X(t_1)) i_1, ..., X(t_{n-1}) = i_{n-1}) = P(X(t_n) = j \mid X(t_{n-1}))$

Definition: transistion probability

$p_{ij}(s, t) = P(X(t) = j \mid X(s) = i)$ for $s \leq t$

homogeneous if $p_{ij}(s, t) = p_{ij}(0, t - s)$

Def: Generator Matrix

$G = (g_{ij}), \quad p_{ij}(h) = g_{ij} h$ if $i \neq j$ and $p_{ii} = 1 + g_{ij} h$

Eq. Forward systems of equations:

$P'_t = P_t G$

Eq. Backward systems of equations:

$P'_t = G P_t$

Often solutions on the form $P_t = \exp(tG)$

Matrix Exponential:

$\exp(tG) = \sum_{n=1}^{\infty} \frac{t^n}{n!} G^n$

Definition:

Irreducible if for any pair i, j, $p_{ij}(t) > 0$ for some t

Definition: Stationary

$\pi, \pi_j \geq 0, \quad \sum_j \pi_j = 1$ and

$\pi = \pi P_t \ \forall t \geq 0$

Claim: $\pi = \pi P_t \Leftrightarrow \pi G = 0$

Theorem:

Stationary if $p_{ij}(t) \to \pi_j$ as $t \to \infty \ \forall i, j$

Not stationary if $p_{ij} \to 0$

6 Convergence of Random Variables

Norm

(a)$\|f\| \geq 0$

(b) $\|f\| = 0$ iff $f = 0$

(c) $\|af\| = |a| \cdot \|f\|$

(d) $\|f + g\| \leq \|f\| + \|g\|$

convergent almost surely

$X_n \xrightarrow{a.s.} X$,

if $\{\omega \in \Omega : X_n(\omega) \to X(\omega)$ as $n \to \infty\}$

convergent in rth mean

$X_n \xrightarrow{r} X$

if $\mathbb{E}|X_n^r| < \infty$ and $\mathbb{E}(|X_n - X|^r) \to 0$ as $n \to \infty$

convergent in probability

$X_n \xrightarrow{P} X$

if $P(|X_n - X| > \epsilon) \to 0$ as $n \to \infty$

convergent in distribution

$X_n \xrightarrow{d} X$

if $P(X_n < x) \to P(X < x)$ as $n \to \infty$

Theorem: implications

$(X_n \xrightarrow{a.s./r} X) \Rightarrow (X_n \xrightarrow{P} X) \Rightarrow (X_n \xrightarrow{D} X)$

For $r > s \geq 1$:

$(X_n \xrightarrow{r} X) \Rightarrow (X_n \xrightarrow{s} X)$

Theorem: additional implications

(a) If $X_n \xrightarrow{D} c$, where c is const, then $X_n \xrightarrow{P} c$

(b) If $X_n \xrightarrow{P} X$ and $P(|X_n| \leq k) = 1$ for all n and some k, then $X_n \xrightarrow{r} X$ for all $r \geq 1$

(c) If $P_n(\epsilon) = P(|X_n - X|) > \epsilon)$ satisfies $\sum_n P_n(\epsilon) < \infty$ for all $\epsilon > 0$, then $X_n \xrightarrow{a.s.} X$

Theorem: Skorokhod's representation t.

If $X_n \xrightarrow{D} X$ as $n \to \infty$

then there exists a probability space and random variable Y_n, Y with:

(a) Y_n and Y have distribution F_n, F

(b) $Y_n \xrightarrow{a.s.} Y$ as $n \to \infty$

Theorem: Convergence over function

If $X_n \xrightarrow{D} X$ and $g : \mathbb{R} \to \mathbb{R}$ is continuous then $g(X_n) \xrightarrow{D} g(X)$

Theorem: Equivalence

The following statements are equivalent:

(a) $X_n \xrightarrow{D} X$

(b) $\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$ for all bounded continuous functions g

(c) $\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$ for all functions g of the form $g(x) = f(x) I_{[a,b]}(x)$ where f is continuous.

Theorem: Borel-Cantelli

Let $A = \cap_n \cup_{m=n}^{\infty} A_m$ be the event that infinitely many of the A_n occur. Then:

(a) $P(A) = 0$ if $\sum_n P(A_n) < \infty$

(b) $P(A) = 1$ if $\sum_n P(A_n) = \infty$ and $A_1, A_2, ...$ are independent.

Theorem:

$X_n \to X$ and $Y_n \to Y$ implies $X_n + Y_n \to X + Y$ for convergence a.s., r:th mean and probability. Not generally true in distribution.

6.1 Laws of large numbers

Theorem:

$X_1, X_2, \ldots$ is iid and $\mathbb{E}(X_i^2) < \infty$ and $\mathbb{E}(X) = \mu$ then

$\frac{1}{n} \sum_{i=1}^{n} X_i \to \mu$ a.s. and in mean square

Theorem:

$\{X_n\}$ iid. Distribution function F. Then $\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{P} \mu$ iff one of the following holds:

1) $nP(|X_1| > n) \to 0$ and $\int_{[-n,n]} x dF$ as $n \to \infty$

2) Char. Fcn. $\Phi(t)$ of X_i is differen-

tiable at $t = 0$ and $\Phi'(0) = i\mu$

Theorem: Strong law of large numbers

$X_1, X_2, ...$ iid. Then

$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mu$ a.s. as $n \to \infty$.

for some μ, iff $\mathbb{E}|X_1| < \infty$. In this case $\mu = \mathbb{E}X_1$

6.2 Law of iterated logarithm

If $X_1, X_2, ...$ are iid with mean 0 and variance 1 then

$\mathbb{P}(\limsup_{n\to\infty} \frac{S_n}{\sqrt{2nloglogn}} = 1) = 1$

6.3 Martingales

Definition: Martingale

$S_n : n \geq 1$ is called a martingale with respect to the sequence $X_n : n \geq 1$, if

(a) $\mathbb{E}|S_n| < \infty$

(b) $\mathbb{E}(S_{n+1} \mid X_1, X_2, ..., X_n) = S_n$

Lemma:

$\mathbb{E}(X_1 + X_2|Y) = \mathbb{E}(X_1|Y) + E(X_2|Y)$

$\mathbb{E}(Xg(Y)|Y) = g(Y)\mathbb{E}(X|Y), \quad g : \mathbb{R}^n \to R$

$\mathbb{E}(X|h(Y)) = \mathbb{E}(X|Y)$ if $h : \mathbb{R}^n \to \mathbb{R}^n$ is one-one

Lemma: Tower Property

$\mathbb{E}[E(X|Y_1, Y_2)|Y_1] = \mathbb{E}(X|Y_1)$

Lemma:

If $\{B_i : 1 \leq i \leq n\}$ is a partition of A then $\mathbb{E}(X|A) = \sum_{i=1}^{n} \mathbb{E}(X|B_i)\mathbb{P}(B_i)$

Theorem: Martingale convergence

If $\{S_n\}$ is a Martingale with $\mathbb{E}(S_n^2 < M < \infty)$ for some M and all n then

$\exists S : S_n \xrightarrow{a.s, L^2} S$

6.4 Prediction and conditional expectation

Notation:

$\|U\|_2 = \sqrt{\mathbb{E}(U^2)} = \sqrt{\langle U, U\rangle}$

$\langle U, V\rangle = \mathbb{E}(UV), \|U_n - U\|_2 \to 0 \Leftrightarrow U_n \xrightarrow{L^2} U$

$\|U + V\|_2 \leq \|U\|_2 + \|V\|_2$

Def.

X, Y r.v. $\mathbb{E}(Y^2) < \infty$. The minimum mean square predictor of Y given X is $\hat{Y} = h(X) = \min \|Y - \hat{Y}\|_2$

Theorem:

If a (linear) space H is closed and $\hat{Y} \in H$ then $\min \left\|Y - \hat{Y}\right\|_2$ exists

Projection Theorem:

H is a closed linear space and $Y : \mathbb{E}(Y^2) < \infty$

For $M \in H$ then $\mathbb{E}((Y - M)Z) = 0 \Leftrightarrow \|Y - M\|_2 \leq \|Y - Z\|_2 \ \forall Z \in H$

Theorem:

Let X and Y be r.v., $E(Y^2) < \infty$. The best predictor of Y given X is $\mathbb{E}(X|Y)$

7 Stochastic processes

Definition: Renewal process

$N = \{N(t) : t \geq 0\}$ is a process for which $N(t) = \max\{n : T_n \leq t\}$ where $T_0 = 0, \quad T_n = X_1 + X_2 + \cdots + X_n$ for $n \geq 1$, and the X_n are iid non-negative r.v.'s

8 Stationary processes

Definition:

The Autocovariance function

$c(t, t + h) = \mathrm{Cov}(X(t), X(t + h))$

Definition: The process $X = \{X(t) : t \geq 0\}$ taking real values is called strongly stationary if the families $\{X(t_1), X(t_2), ..., X(t_n)\}$ and $\{X(t_1 + h), X(t_2 + h), ..., X(t_n + h)\}$ has the same joint distribution for all $t_1, t_2, ..., t_n$ and $h > 0$

Definition: The process $X = \{X(t) : t \geq 0\}$ taking real values is called weakly stationary if, for all t_1 and t_2 and $h > 0$:$\mathbb{E}(X(t_1)) = \mathbb{E}(X(t_2))$ and $\mathrm{Cov}(X(t_1), X(t_2)) = \mathrm{Cov}(X(t_1 + h), X(t_2 + h))$, thus if and only if it has constant means and its autocovariance function satisfies $c(t, t + h) = c(0, h)$

Definition:

The covariance of complex-valued $C_1$and C_2 is

$\mathrm{Cov}(C_1, C_2) = \mathbb{E}\left((C_1 - \mathbb{E}C_1)(\overline{C_2 - \mathbb{E}C_2})\right)$

Theorem:

$\{X\}$ real, stationary with zero mean and autocovariance $c(m)$.

The best predictor from the class of linear functions of the subsequence $\{X\}_{r-s}^{r}$ is

$\widehat{X}_{r+k} = \sum_{i=0}^{s} a_i X_{r-i}$

where $\sum_{i=0}^{s} a_i c(|i - j|) = c(k + j)$ for $0 \leq j \leq s$

Definition:

Autocorrelation function of a weakly stationary process with autocovariance function $c(t)$ is

$\rho(t) = \frac{\mathrm{Cov}(X(0), X(t))}{\sqrt{\mathrm{Var}(X(0))\mathrm{Var}(X(t))}} = \frac{c(t)}{c(0)}$

Theorem:

Spectral theorem: Weakly stationary process X with strictly pos. σ^2 is the char. function of some distribution F whenever $\rho(t)$ is continuous at $t = 0$

$\rho(t) = \int_{-\infty}^{\infty} e^{\lambda} dF(\lambda)$

F is called the spectral distribution function.

Ergodic theorem:

X is strongly stationary such that $\mathbb{E}|X_1| < \infty$ there exists a r.v Y with the same mean as X_n such that $\frac{1}{n}\sum_{j=1}^{n} X_j \to Y$ a.s and in mean

Weakly stationary processes:

If $X = \{X_n : n \geq 1\}$ is a weakly stationary process, there exists a Y such that $\mathbb{E}(Y) = \mathbb{E}(X_1)$ and $\frac{1}{n}\sum_{j=1}^{n} X_j \xrightarrow{m.s.} Y$.

8.1 Gaussian processes

Definition:

A real valued c.t. process is called Gaussian if each finite dimensional vector $(X(t_1), X(t_2), \ldots, X(t_n))$ has the multivariate normal distribution $\mathcal{N}(\mu(t), \mathbf{V}(t)), \quad t = (t_1, t_2, \ldots, t_n)$

Theorem:

The Gaussian process X is stationary iff. $\mathbb{E}(X(t))$ is constant for all t and $\mathbf{V}(t) = \mathbf{V}(t + h)$ for all t and $h > 0$

9 Inequalities

Cauchy-Schwarz:

$(\mathbb{E}(XY))^2 \leq \mathbb{E}(X^2)\mathbb{E}(Y^2)$

with equality if and only if $aX + bY = 1$.

Jensen's inequality:

Given a convex function $J(x)$ and a r.v. X with mean μ: $\mathbb{E}(J(X)) \geq J(\mu)$

Markov's inequality

$P(|X| \geq a) \leq \frac{\mathbb{E}|X|}{a}$ for any $a > 0$

$h : (R) \to [0, \infty]$ non-negative fcn, then $P(h(X) \geq a) \leq \frac{\mathbb{E}(h(X))}{a} \ \forall a > 0$

General Markov's inequality:

$h : (R) \to [0, M]$ non-negative function bounded by some M. Then

$P(h(X) \geq a) \leq \frac{\mathbb{E}(h(X))-a}{M-a} \quad 0 \leq a < M$

Chebyshev's Inequality:

$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}(X^2)}{a^2}$ if $a \geq 0$

Theorem: Holder's inequality

If $p, q > 1$ and $p^{-1} + q^{-1} = 1$ then

$\mathbb{E}|XY| \leq (\mathbb{E}|X^p|)^{\frac{1}{p}} (\mathbb{E}|Y^p|)^{\frac{1}{q}}$

Minkowski's inequality:

If $p \geq 1$ then

$[\mathbb{E}(|X + Y|^p)]^{\frac{1}{p}} \leq (\mathbb{E}|X^p|)^{\frac{1}{p}} + (\mathbb{E}|Y^p|)^{\frac{1}{q}}$

Minkowski 2:

$\mathbb{E}(|X + Y|^p) \leq C_P[\mathbb{E}|X^p| + \mathbb{E}|Y^p|]$

where $p > 0$ and

$$C_P \to \begin{cases} 1 & 0 < p \leq 1 \\ 2^{p-1} & p > 0 \end{cases}$$

Kolomogorov's inequality:

Let $\{X_n\}$ be iid with zero means and variances σ_n^2. Then for $\epsilon > 0$

$\mathbb{P}(\max_{1 \leq i \leq n} |X_1 + \cdots + X_i| > \epsilon) \leq \frac{\sigma_1^2 + \cdots + \sigma_n^2}{\epsilon^2}$

Doob-Kolomogorov's inequality:

If $\{S_n\}$ is a martingale, then for any $\epsilon > 0$

$\mathbb{P}(\max_{1 \leq i \leq n} |S_i| > \epsilon) \leq \frac{\mathbb{E}(S_n^2)}{\epsilon^2}$