Contents

Part I

Statistics

Chapter 1

Probability

1.1 Convergence

1. Convergence in probability.

$$T_n \xrightarrow{\mathbb{P}} T \iff \mathbb{P}(|T_n - T| \geq \epsilon) \xrightarrow[n\to\infty]{} 0 \quad \forall \epsilon > 0$$

2. Almost surely convergence.

$$T_n \xrightarrow{\text{a.s.}} T \iff \mathbb{P}(\{\omega \,|\, T_n(\omega) \to T(\omega)\}) = 1$$

3. Convergence in distribution.

$$T_n \xrightarrow{(d)} T \iff \mathbb{P}(T_n \leq x) \xrightarrow[n\to\infty]{} \mathbb{P}(T \leq x)$$

   for all x at which F_T is continuous.

4. $\mathcal{L}^p$ Convergence.

$$T_n \xrightarrow{:\surd} T \iff \mathbb{E}[|T_n - T|^p] \xrightarrow[n\to\infty]{} 0 \tag{1.1}$$

   For $p \geq 1$.

Theorem 1.1 (Weak/Strong law of large numbers). *Let $X_1, \ldots, X_n$ be i.i.d. with $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$ both finite. Then*

$$\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow[n\to\infty]{\mathbb{P}/\text{a.s.}} \mu$$

Theorem 1.2 (Central limit theorem). *Let $X_1, \ldots, X_n$ be i.i.d with $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathrm{Var}[X]$ both finite. Then*

$$\sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow[n\to\infty]{(d)} \mathrm{N}(0,1)$$

Theorem 1.3 (Hoeffding inequality). *Let $X_1, \ldots, X_n$ be i.i.d with $\mu = \mathbb{E}[X]$, $X \in [a, b]$ then for all $\epsilon > 0$*

$$\mathbb{P}\left(\left|\bar{X}_n - \mu\right| \geq \epsilon\right) \leq 2e^{-\frac{2n\epsilon^2}{(b-a)^2}}$$

Proposition 1.4. *The followings are equivalent*

1. $T_n \xrightarrow{(d)} T$.
2. $\mathbb{E}[f(T_n)] \xrightarrow[n\to\infty]{} \mathbb{E}[f(T)]$ *for all continuous and bounded* f.
3. $\mathbb{E}\left[e^{ixT_n}\right] \xrightarrow[n\to\infty]{} \mathbb{E}\left[e^{ixT}\right]$ *for all* x.

Proposition 1.5. *We have the following relationships*

$$T_n \xrightarrow{\text{a.s.}} T \implies T_n \xrightarrow{\mathbb{P}} T \implies T_n \xrightarrow{(d)} T$$

and

$$T_n \xrightarrow{L^1} T \implies T_n \xrightarrow{\mathbb{P}} T$$

and

$$T_n \xrightarrow{L^p} T \implies T_n \xrightarrow{L^q} T, \quad \forall q \leq p$$

Proposition 1.6. *Let* f *be a continuous function then*

$$T_n \xrightarrow{\text{a.s.}/\mathbb{P}/(d)} T \implies f(T_n) \xrightarrow{\text{a.s.}/\mathbb{P}/(d)} f(T)$$

Proposition 1.7. *If* $U_n \xrightarrow{\text{a.s.}/\mathbb{P}} U$ *and* $V_n \xrightarrow{\text{a.s.}/\mathbb{P}} V$ *then*

1. $$U_n + V_n \xrightarrow{\text{a.s.}/\mathbb{P}} U + V$$

2. $$U_n V_n \xrightarrow{\text{a.s.}/\mathbb{P}} UV$$

3. $$\frac{U_n}{V_n} \xrightarrow{\text{a.s.}/\mathbb{P}} \frac{U}{V}, \quad V \neq 0 \,\text{a.s.}$$

These propositions hold for convergence in distribution if the pair $(U_n, V_n) \xrightarrow{(d)} (U, V)$.

Theorem 1.8 (Slutsky's theorem). *Let* X_n, Y_n *be a sequence of random variable such that*

$$X_n \xrightarrow{(d)} X \qquad Y_n \xrightarrow{\mathbb{P}} c$$

where c *is a constant then* $(X_n, Y_n) \xrightarrow{(d)} (X, c)$. *In particular*

$$X_n + Y_n \xrightarrow{(d)} X + c, \quad X_n Y_n \xrightarrow{(d)} cX$$

Theorem 1.9 (Delta method). *Let* X_n *be a sequence of random variable such that*

$$\sqrt{n}(X_n - \mu) \xrightarrow{(d)} \mathrm{N}\big(0, \sigma^2\big)$$

where μ *and* σ *are finite valued constants. Then for any function* g *that* $g'(\mu)$ *exists and is non-zero*

$$\sqrt{n}(g(X_n) - g(\mu)) \xrightarrow{(d)} \mathrm{N}\big(0, \sigma^2\big(g'(\mu)^2\big)\big)$$

Proof. By Taylor's polynomial, there exists μ^* between X_n and μ such that

$$g(X_n) - g(\mu) = (X_n - \mu)g'(\mu^*)$$

then since $g'(\mu^*)$ convergers in probability – because of weak law of large number, continuity of g', and continuous mapping theorem – to $g'(\mu)$, by Slutsky's theorem

$$\sqrt{n}(X_n - \mu)g'(\mu^*) \xrightarrow{(d)} \mathrm{N}\Big(0, \sigma^2 (g'(\mu))^2\Big)$$

which proves the Delta method. ■

Chapter 2

Parametric Models

Chapter 3

Introduction

Definition: A statiscal model is parametric if it can be determined using a set of parameters. A statiscal model that can not be adquately determined by a set of parameters is called a non-parametric model. Models that have both components are called semi-parametric.

Definition: A parametric model is identifiable if

$$\theta_1 \neq \theta_2 \implies P_{\theta_1} \neq P_{\theta_2} \tag{3.1}$$

Definition: A statistic is a function from sample space $\mathcal{X}$ to some spave of values, $\mathcal{T}$.

Definition: Any parametric model that either

1. All of P_θ are continuous with denisities $p(x, \theta)$
2. All of P_θ are discrete with frequency functions $p(x, \theta)$, and the support set $\{x_1, x_2, \ldots\} \equiv \{x \mid p(x, \theta) > 0\}$

are called regular parametric models.

Chapter 4

Problems

4.0.1

Let U be any random variable and V be any other non-negative random variable. Show that $U + V$ is stochastically larger then U, that is

$$F_{U+V}(t) \leq F_U(t), \quad \forall t \tag{4.1}$$

Proof. Let

$$A_1 = \{\omega \,|\, U + V \leq t\}, \quad A_2 = \{\omega \,|\, U \leq t\}$$

since $U + V \geq U$ then $A_1 \subset A_2$ and hence

$$\mathbb{P}(A_1) = F_{U+V}(t) \leq F_U(t) = \mathbb{P}(A_2) \qquad \blacksquare$$

Chapter 5

Stochastic Processes

Chapter 6

Introduction

A stochastic process is set of indexed random variables. More formally, a stochastic process maps each sample to a sequence of random variables.

$$X : \Omega \to \mathbb{R}^{\mathbb{T}}, \quad X(\omega) = \{(X(t))(\omega) \,|\, t \in \mathbb{T}\}$$

Usually, $\mathbb{T}$ is interpreted as a set of time. Similarly, a stochastic process can be thought of a function time and sample to real numbers.

$$X : \Omega \times \mathbb{T} \to \mathbb{R}$$

Some definitions regarding stochastic processes are listed below

Discrete-time when $\mathbb{T}$ is discrete.

Continuous-time when $\mathbb{T}$ is continuous.

Discrete-state when $X(t)$ are discrete.

Continuous-state when $X(t)$ are continuous.

Equality Two stochastic processes are equal if

$$X(\omega, t) = Y(\omega, t), \quad \forall \omega \in \Omega, t \in \mathbb{T}$$

Equality in minimum squared sense is defined as

$$\mathbb{E}\left[(X(t) - Y(t))^2\right] = 0, \quad \forall t$$

6.1 Order distribution

The first order distribution of a stochastic process X is defined as

$$F(x; t) = \mathbb{P}(X(t) \le x)$$

Morever, the n_{th} distribution of X is

$$F(x_1, \ldots, x_n; t_1, \ldots, t_n) = \mathbb{P}(X(t_1) \le x_1, \ldots, X(t_n) \le x_n)$$

the n_{th} density of X is partials of distribution with respected x.

$$f(x_1, \ldots, x_n; t_1, \ldots, t_n) = \frac{\partial^n F(x_1, \ldots, x_n; t_1, \ldots, t_n)}{\partial x_1 \ldots \partial x_n}$$

To determine all statistical properties of a stochastic process we need every n_{th} distribution or density. However mostly, we need certain averages which are calculated using the first and second densities. These averages inclue

Mean

$$\eta(t) = \mathbb{E}[X(t)] = \int_{-\infty}^{\infty} x f(x;t)\,\mathrm{d}x$$

Autocorrelation

$$\begin{aligned} R(t_1, t_2) &= \mathbb{E}[X(t_1)X(t_2)] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 x_2 f(x_1, x_2; t_1, t_2)\,\mathrm{d}x_2\,\mathrm{d}x_1 \end{aligned}$$

The value of autocorrelation on $t_1 = t_2$ is the average power of $X(t)$.

Autocovariance

$$C(t_1, t_2) = R(t_1, t_2) - \eta(t_1)\eta(t_2)$$

The value of autocovariance on $t_1 = t_2$ is called the variance of $X(t)$.

Correlation coefficient

$$r(t_1, t_2) = \frac{C(t_1, t_2)}{\sqrt{C(t_1, t_1)C(t_2, t_2)}}$$

Example 6.1. Suppose X is a stochastic process and S is a random variable defined as

$$S = \int_a^b X(t)\,\mathrm{d}t$$

then

$$\begin{aligned} \mathbb{E}[S] &= \int_a^b \mathbb{E}[X(t)]\,\mathrm{d}t = \int_a^b \eta(t)\,\mathrm{d}t \\ \mathbb{E}\left[S^2\right] &= \mathbb{E}\left[\left(\int_a^b X(t_1)\,\mathrm{d}t_1\right)\left(\int_a^b X(t_2)\,\mathrm{d}t_2\right)\right] \\ &= \mathbb{E}\left[\int_a^b\int_a^b X(t_1)X(t_2)\,\mathrm{d}t_2\,\mathrm{d}t_1\right] \\ &= \int_a^b\int_a^b R(t_1, t_2)\,\mathrm{d}t_2\,\mathrm{d}t_1 \end{aligned}$$

6.2 Joint stochastic process

The joint distribution of two stochastic processes X and Y is determined in terms of the joint distribution of their random variables

$$F(x_1, \ldots, x_n, y_1, \ldots, y_m; t_1, \ldots, t_n, t'_1, \ldots t'_m)$$

We can expand the notion of stochastic processes to complex processes in which the process takes on complex value. Then, revising autocorrelation

$$R_{XX}(t_1, t_2) = \mathbb{E}\Big[X(t_1)\overline{X(t_2)}\Big]$$

Then we have

$$R_{XX}(t_2, t_1) = \overline{R_{XX}(t_1, t_2)}$$

and

$$R_{XX}(t, t) = \mathbb{E}\big[|X(t)|^2\big] \geq 0$$

Remark 1. R is a positive semi-definite function that is

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_i \bar{a}_j R(t_i, t_j) \geq 0$$

for all n, $t_1, \ldots, t_n$ and $a_1, \ldots, a_n \in \mathbb{C}$. Conversely, if R is a positive semi-definite function then there exists a stochastic X with autocorrelation R.

Similarly, autocovariance can be revised to

$$C(t_1, t_2) = R(t_1, t_2) - \eta(t_1)\overline{\eta(t_2)}$$

Expanding these notation for joint distribution, cross-correlation is

$$R_{XY}(t_1, t_2) = \mathbb{E}\Big[X(t_1)\overline{Y(t_2)}\Big]$$

and cross-covariance

$$C_{XY}(t_1, t_2) = R_{XY}(t_1, t_2) - \eta_X(t_1)\overline{\eta_Y(t_2)}$$

Definition: Two process X and Y are called (mutually) **orthogonal** if

$$R_{XY}(t_1, t_2) = 0, \quad \forall t_1, t_2$$

and they are called **uncorrelated** if

$$C_{XY}(t_1, t_2) = 0, \quad \forall t_1, t_2$$

Lastly, two processes are independent if for any n and any $t_1, \ldots, t_n$ and $t'_1, \ldots, t'_n$, the random variable $X(t_1), \ldots X(t_n)$ and $Y(t'_1), \ldots Y(t'_n)$ are mutually independent.

6.3 Stationary processes

6.3.1 Strict sense stationary process

X is an SSS process if its statistical properties are invariant to time shifts. That is, the processes $X(t)$ and $X(t+c)$ have the same statistics for any c. In other word, for any n and $x_1, \ldots, x_n$ and any $t_1, \ldots, t_n$ and c

$$F(x_1, \ldots, x_n; t_1, \ldots t_n) = F(x_1, \ldots, x_n; t_1 + c, \ldots, t_n + c)$$

Similarly, two process X and Y are jointly stationary if the joint statistics of $X(t)$ and $Y(t)$ are the same as the joint statistics of $X(t+c)$ and $Y(t+c)$.

Example 6.2. Suppose X is an SSS process. Then its first order distribution is independent of time t as

$$F(x; t) = F(x; t + c)$$

for all c. Implying that

$$F(x; t) = F(x) \implies \eta_X(t) = \eta$$

Its second order distribution is only dependent of the time difference.

$$F(x_1, x_2; t_1, t_2) = F(x_1, x_2; t_1 + c, t_2 + c) = F(x_1, x_2; \tau), \quad \tau = t_1 - t_2$$

therefore

$$R(t_1, t_2) = R(t_1 - t_2) = R(\tau)$$

6.3.2 Wide sense stationary process

A stochastic process X is called **wide sense stationary** if

$$\mathbb{E}[X(t)] = \eta \quad \text{(independent of time)}$$

and

$$R(t + \tau, t) = R(\tau) \quad \text{(depends on the difference)}$$

$R(\tau)$ can be written symmetrically

$$R(\tau) = \mathbb{E}\left[X\left(t + \frac{\tau}{2}\right)\overline{X\left(t - \frac{\tau}{2}\right)}\right]$$

In particular,

$$R(0) = \mathbb{E}\left[|X(t)|^2\right]$$
$$\mathrm{Var}[X(t)] = \sigma_X^2$$

are both independent of time.

Two processes are jointly WSS if each is WSS and their cross-correlation depends only on $\tau = t_1 - t_2$

6.3.3 Mean square peridicity

A process X is called mean square periodic if

$$\mathbb{E}\left[|X(t+\tau)-X(t)|^2\right]=0$$

for every t. Therefore by Markov's inequality

$$\mathbb{P}\left(|X(t+\tau)-X(t)|^2\geq\epsilon^2\right)\leq\frac{\mathbb{E}\left[|X(t+\tau)-X(t)|^2\right]}{\epsilon^2}=0$$

It follows that, for a specific t

$$x(t+\tau)=x(t)$$

with probability 1.

Proposition 6.1. *X is MS periodic if and only if its autocorrelation is doubly periodic.*

$$R(t_1+m\tau,t_2+n\tau)=R(t_1,t_2),\quad\forall m,n\in\mathbb{Z}$$

Proof. This follows from the fact that we can define a inner product on the space of random variables

$$\langle X,Y\rangle=\mathbb{E}[XY]$$

and thus by the Cauchy-Shwarz inequality we have

$$(\mathbb{E}[XY])^2\leq\mathbb{E}\left[X^2\right]\mathbb{E}\left[Y^2\right]$$

■

Definition: An independent and identically distributed process X has the following property

$$\forall n,\ \forall t_1,\dots,t_n,\ F(x_1,\dots,x_n;t_1,\dots,t_n)=F(x_1)\dots F(x_n)$$

hence X is SSS as well.

6.4 Ergodicity

We usually do not have access to multiple runs of a stochastic process to be able estimate its statistical properties. **Ergodic** processes is process that its statistical properties can be deduced from a single sufficiently run. Formally, A stationary process X is mean-ergodic if the time average estimates

$$\hat{\mu_T}=\frac{1}{2T}\int_{-T}^{T}X(t)\,\mathrm{d}t$$

converges in squared mean to ensemble average $\mu(t)=\mathbb{E}[X(t)]=\mu$ as $T\to\infty$. That is

$$\lim_{T\to\infty}\mathrm{Var}[\mu_T]=\mu$$

expanding the left hand side

$$\begin{aligned}\mathrm{Var}[\mu_T]&=\frac{1}{4T^2}\int_{-T}^{T}\int_{t-T}^{t+T}C(\tau)\,\mathrm{d}\tau\,\mathrm{d}t\\&=\frac{1}{4T^2}\int_{-2T}^{2T}(2T-|\tau|)C(\tau)\,\mathrm{d}\tau\\&=\frac{1}{2T}\int_{-2T}^{2T}\left(1-\frac{|\tau|}{2T}\right)C(\tau)\,\mathrm{d}\tau,\ \underbrace{C(\tau)=C(-\tau)}_{\text{for real processes}}\\&=\frac{1}{2T}\int_{0}^{2T}\left(1-\frac{|\tau|}{2T}\right)C(\tau)\,\mathrm{d}\tau\end{aligned}$$

Theorem 6.2 (Slutsky). *X is mean-ergodic if and only if*

$$\frac{1}{T}\int_0^T C(\tau)\,\mathrm{d}\tau \to 0 \text{ as } T \to \infty$$

Intuitively, X is loosely correlated to its past and thus it will eventually visits the whole space.

Remark 2. research Martingale process.

6.5 Systems with Stochastic Input

Suppose X is a stochastic process having $X(t, \omega_j)$. Let $Y(t)$ be a function of sample paths

$$Y(t) = T[X(t)]$$

Then, the system is described with T. The properties of a system

Deterministic

$$X(t, \omega_1) = X(t, \omega_2) \implies T[X(t, \omega_1)] = T[X(t, \omega_2)]$$

Stochastic not deterministic?!

Memoryless Output only depends on the input at the particular time.

Linear

$$T[aX_1(t) + bX_2(t)] = aT[X_1(t)] + bT[X_2(t)]$$

where a, b are random variable.

Time invariant

$$Y(t) = T[X(t)] \implies Y(t-c) = T[X(t-c)]$$

For linear time invariant system, the output can be written as

$$Y(t) = X(t) * h(t)$$

where h is the impulse response of the system. Moreover, if X is SSS then Y is SSS.