Stationarity vs Time Scale Dependence

or

Statistics vs Sample Path Variability

or

Tracking vs Estimation

MICRO-TUTORIAL

ISMA 2003 Bandwidth Estimation Workshop (BEst)

Darryl Veitch
“Estimating Available Bandwidth on Different Timescales”
Sounds good –
Sounds good – but what does it really mean?
“Estimating Available Bandwidth on Different Timescales”

Sounds good – but what does it really mean?

Must distinguish between:

- **Statistics** i.e. distributions, which are **not** random
  – could be **constant** over time: *stationary*, or
  – could be **varying** over time: *non-stationary*
  – try to measure them: **statistical estimation**

- **Sample paths** these are ‘random functions’
  – a single sample path is deterministic
  – variability across and within paths is natural, **regardless** of statistics
  – one path not enough for good estimation

**Note:** Using conceptual random process **model** – unavoidable, and natural
Consider a continuous r.v. \( X \), with

- Distribution function (CDF) \( F(x) = \Pr(X \leq x) \), probability density \( f(x) = F'(x) \),
- \textbf{Expectation:} \( \mu_X = \mathbb{E}[X] = \int x f(x) \, dx \), \quad \textbf{Variance:} \( \sigma_{X}^2 = \mathbb{E}[(X - \mu)^2] \),
A Random Variable: Distribution, Samples and Estimation

Consider a continuous r.v. $X$, with

- Distribution function (CDF) $F(x) = \Pr(X \leq x)$, probability density $f(x) = F'(x)$,
- **Expectation**: $\mu_X = \mathbb{E}[X] = \int x f(x) \, dx$,  **Variance**: $\sigma_X^2 = \mathbb{E}[(X - \mu)^2]$,

Now consider a sample value $x$ of $X$. How to estimate $\mu_X$?

Could set $\hat{\mu} = x$. Really this is a sample value of an **Estimator**

- **Estimator**: $\hat{\mu}_X = X$ is a r.v.
- With one sample $x$ of $X$, sample of $\hat{\mu}$ is also $x$.
- Not a great solution since although $\mathbb{E}[\hat{\mu}] = \mu$, have $\text{Var} [\hat{\mu}] = \sigma^2$.
- Can’t even estimate $\sigma_X^2$ with this, or much else!
A Random Variable: Distribution, Samples and Estimation

Consider a continuous r.v. $X$, with

- Distribution function (CDF) $F(x) = \text{Pr}(X \leq x)$, probability density $f(x) = F'(x)$,
- **Expectation**: $\mu_X = \mathbb{E}[X] = \int x f(x) \, dx$, **Variance**: $\sigma^2_X = \mathbb{E}[(X - \mu)^2]$,

Now consider a sample value $x$ of $X$. How to estimate $\mu_X$?

Could set $\hat{\mu} = x$. Really this is a sample value of an **Estimator**

- **Estimator**: $\hat{\mu}_X = X$ is a r.v.
- With one sample $x$ of $X$, sample of $\hat{\mu}$ is also $x$.
- Not a great solution since although $\mathbb{E}[\hat{\mu}] = \mu$, have $\text{Var} [\hat{\mu}] = \sigma^2$.
- Can’t even estimate $\sigma^2_X$ with this, or much else!

Would like to have more (independent) samples available:

Could set $\hat{\mu}_n = \bar{X} \equiv (\sum_{i}^{n} X_i)/n$, with sample values $x_i, i = 1, 2 \ldots$.

- $\mathbb{E}[\hat{\mu}] = \mu$, $\text{Var} [\hat{\mu}] = \sigma^2/n$.
- Gets rapidly better with increasing $n$!
- Now could estimate $\sigma^2$, e.g. using $S^2 = \sum_{i}^{n} (X_i - \bar{X})^2/(n - 1)$, the ‘sample variance’ (another r.v.).
Consider a time series $X(t)$, say $t \in \mathbb{Z}$.

- Distribution functions $F_{X(t)}(x)$ for all $t$, and 2-D joint $F_{X(t),X(t')} (x, x')$, all 3-D ....
- Expectations $\mu_{X(t)}$, variance, 3rd order moments, etc for each $t$.
- All statistics formed from joint distributions from any combination of the $X(t)$’s.
Consider a time series $X(t)$, say $t \in \mathbb{Z}$.

- Distribution functions $F_{X(t)}(x)$ for all $t$, and 2-D joint $F_{X(t),X(t')}(x,x')$, all 3-D ....
- Expectations $\mu_{X(t)}$, variance, 3rd order moments, etc for each $t$.
- All statistics formed from joint distributions from any combination of the $X(t)$’s.

Now consider a sample path $x(t)$. How to estimate $\mu_5 = E[X(5)]$?

As $X(5)$ is just one r.v., as before could set $\hat{\mu}_5 = X(5)$.

- $E[\hat{\mu}_5] = \mu_5$, $\text{Var} [\hat{\mu}_5] = \sigma^2$.
- As before, bad
Consider a time series $X(t)$, say $t \in \mathbb{Z}$.

- Distribution functions $F_{X(t)}(x)$ for all $t$, and 2-D joint $F_{X(t),X(t')}^{(x,x')}$, all 3-D ....
- Expectations $\mu_{X(t)}$, variance, 3rd order moments, etc for each $t$.
- All statistics formed from joint distributions from any combination of the $X(t)$’s.

Now consider a sample path $x(t)$. How to estimate $\mu_5 = E[X(5)]$?

As $X(5)$ is just one r.v., as before could set $\hat{\mu}_5 = X(5)$.

- $E[\hat{\mu}_5] = \mu_5$, $\text{Var}[\hat{\mu}_5] = \sigma^2$.
- As before, bad

Would like to have more (independent) samples available:

But there aren’t any! unless:

- We assume stationarity (and ergodicity), or
- We use simulation, or
- We assume real data can ‘repeat itself’, and can afford to wait.
Estimating With Non-Stationarity - how hard it can be

Estimates on ‘different time-scales’

- Not only time-scale, but time instant matters.
- Time varying statistics and sample variability mixed together.
- Only get ‘one sample’ of everything.
The Stationary Case

Consider a stationary time series $X(t)$.

- Stationary means that all statistics are time-origin invariant:
- eg.1: Marginal distributions all the same: $F_{X(t)} = F_{X(t')}$ (so $E[X(t)] = E[X(t')]$ etc.).
- eg.2: Covariance Function: $\gamma(t - t') = E[(X(t) - \mu)(X(t') - \mu)]$ depends only on $t - t'$
- eg.3: Any statistic formed from any combination of the $X(t)$'s.

Now consider a sample path $x(t)$. How to estimate $\mu = \mu_X(t)$?

Set $\hat{\mu} = \bar{X}$ as before:

- $E[\hat{\mu}] = \mu$, Var [$\hat{\mu}$] = $\sigma^2 / n$ (in white noise case)
- Now gets better with $n$.

More samples are available, and statistics are constant
Estimating With Stationarity - how hard can it be?

Gaussian White Noise sample paths

Estimates on ‘different time-scales’

- **Nothing** varies with time! – only have to worry about sample variability.
- Effectively get **many** samples of everything.
- Different time scales relates to **estimation variance**.
Time Scale Dependence of Estimates

Gaussian White Noise aggregated over different time intervals
AB will change with time – we want to track it.

However, can’t estimate an arbitrary non-stationarity, so:

- Must assume stationarity over some time interval $T$
- With $T$ bounded, so is estimation quality

Measurement over Different timescales refers to the tradeoff:

- Small $T$ makes stationarity assumption better
- But estimation variance (assuming stationarity) worse