Given Notes

Sequence of R.V.

Random vector \( X = [X_1, X_2, \ldots, X_n] \)

\[
P\{X \in D\} = \int_D f_X(x) \, dx = \int_{x_1, x_2, \ldots, x_n} f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) \, dx_1 \, dx_2 \ldots \, dx_n
\]

where \( f_X(x) = \frac{\partial^n F(x)}{\partial x_1 \partial x_2 \ldots \partial x_n} \)

Joint or multivariate pdf of R.Vs \( X_i \):

\[
P_X(x) = P(X_1 = x_1, \ldots, X_n = x_n)
\]

Joint Distribution

Note:

\[
P(x_1, x_2, x_3) = F(x_1, x_2, x_3)_{n=4}
\]

\[
f(x_1, x_2) = \int f(x_1, x_2, x_3, x_4) \, dx_3 \, dx_4
\]

Transformations

\( X = [X_1, X_2, \ldots, X_n] \)

\( Y_1 = g_1(X), \ldots, Y_n = g_n(X) \)
Recall a RV \( X, Y \)
\[ Z = g(X, Y) \quad W = h(X, Y) \]

and \( f_{Z,W}(z,w) \)

\[ f_{Z,W}(z,w) = \frac{f_{X,Y}(x,y)}{|J(x,y)|} \quad + \quad \ldots \]

where \( J(x,y) = \det \begin{bmatrix} \frac{\partial x}{\partial x} & \frac{\partial x}{\partial y} \\ \frac{\partial y}{\partial x} & \frac{\partial y}{\partial y} \end{bmatrix} \)

\( \delta \) for a RV \( X, X_2, \ldots, X_n \)
\[ Y_1 = g_1(X), \quad \ldots \quad Y_n = g_n(X) \]

\[ f_{Y}(y) = \delta \]

Solution: \( \text{End with } (x_1, x_2, \ldots, x_n) \) of \( g_1(x_1, x_2, \ldots, x_n) = y_1, \quad g_2(x_1, x_2, \ldots, x_n) = y_2, \quad \ldots \quad g_n(x_1, x_2, \ldots, x_n) = y_n \)

\[ f_{Y}(y) = \frac{f_{X}(x_1, x_2, \ldots, x_n)}{|J(x_1, x_2, \ldots, x_n)|} \quad + \quad \ldots \]
Independent Experiments vs Related Trials

Recall: Cartesian product of 2 sets $S_1, S_2$

$$S = S_1 \times S_2$$

$S$ - sample space

All events of the form $A \times B$ where $A$ event in $S_1$

$B$ event in $S_2$.

Let us say that

$$P(A \times S_2) = P(A) \quad P(S \times B) = P(B)$$

where $P(A)$ prob. of event $A$ in event $S_1$

$P(B)$ prob. of event $B$ in event $S_2$.

$$P(A \times S_2) = P(A)$$

motivated by the fact that the event $A \times S_2$ if the event $S$ occurs if the event $A$ or event $S_2$ occurs no matter what the outcome of $S_2$.

What about events $A \times B$? One cannot always say

determined in terms of $P_1$ or $P_2$.

$$P(A \times B)$$

can be determined by counting in the usual way. If events $A, B$ indep.
Ex. 3.3

A ball drawn from each box.
P, ball from B₁ white, and ball from B₂ red; Y = ?

Easily seen that \( P(AB) = P(A)P(B) \)

Ex. 3.4

Coin tossed twice

\[ P(H_1H_2Y) = P(H_1)P(H_2) \]

How about \( P(Y) \) heads at the first toss

\[ = P(H_1H_2Y) + P(H_1H_2Y) \]

\[ = P(H_1)P(H_2) + P(H_1)P(H_2) \]
Then \( P(A \times B) = P(A) \times P(B) \)

Ex: \( P(y|y) = P(y) \times P(y) \)

**Independent Exps.**

Given RVs defined on product spaces \( S = S_1 \times S_2 \)

- Let \( X : S_1 \rightarrow \mathbb{R} \) and \( Y : S_2 \rightarrow \mathbb{R} \)
- \( S_1 \times S_2 \) is the set of all outcomes \( \{(s_1, s_2) \} \)

In the product space \( S_1 \times S_2 \), always \( X, Y \) are such that

\[ X(y, y_2) = X(y_2) \quad Y(y, y_2) = Y(y_2) \]

That is \( X \) depends only on outcomes in \( S_1 \)

\[ Y \]

**Thm 6.2** If \( X \) and \( Y \) independent then \( X, Y \) indeps.

**Proof:** Let \( A_x = \{ X = x \} \) \( \text{in} \ S_1 \) \( B_y = \{ Y = y \} \) \( \text{in} \ S_2 \)

In \( S_1 \times S_2 \), \( A_x \times B_y \) is \( \{(x, y) \} \)
If $A \times S_2 = S \times B$ by indp., then $X, Y$ indp.

On 7, suppose

$$S^* = S_1 \times S_2 \times \ldots \times S_n$$

combined exp. but $X_i$ defined only on $y_i$ of $S_i$.

That is $X_i(y_1, y_2, \ldots, y_n) = X_i(y_i)$ $i = 1, \ldots, n$

If exp. $S_i$ independent, then $X_i$ indp.

Now suppose RV $X$ defined on exp. $S$ in exp. $S$

performed $n$ times generating $S^* = S_1 \times \ldots \times S_n$.

In this exp., define RV $X_i$ by

$$X_i(y_1, y_2, \ldots, y_n) = X_i(y_i) \quad i = 1, \ldots, n \quad (7-12)$$

From this it follows that $F_i(x_i)$ of $F_i(x_i)$

of $X_i$ of $X$

Thus, if exp. performed $n$ times, RV $X_i$ defined in (7-12)

are independent or have the same distr. $F(x)$.

Thus RVs i.i.d.
Ex. 7.4 MEASUREMENT ERROR

Measure length \( L \) of object, say \( x \), with instrument of varying accuracy. The \( n \) measurements are \( n \) RVs

\[ X_i = \eta + V_i \quad E(V_i) = 0 \quad E(V_i^2) = \sigma_i^2 \]

Assume \( V_i \) measurement error, assume independent.

Want to estimate \( \eta \): \( \hat{\eta} \)

Want to find \( n \) constants \( a_i \): \( \hat{\eta} = a_1x_1 + \cdots + a_nx_n \)

Hence

\[ E(\hat{\eta}) = a_1E(x_1) + \cdots + a_nE(x_n) = \eta \]

and

\[ \text{Var}(\hat{\eta}) = a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2 \]

is minimum.

Solution: Apply constraint \( a_1 + \cdots + a_n = 1 \) (7-9)

From \( V = a_1\sigma_1^2 + \cdots + a_n\sigma_n^2 - \lambda(a_1 + \cdots + a_n - 1) \)

\[ V_{\text{min}} \text{ if } \left( \frac{dV}{da_i} \right)_{a_i = \lambda} = 2\lambda \sigma_i^2 - \lambda = 0 \Rightarrow a_i = \frac{\lambda}{2\sigma_i^2} \]

Inserting in \( (7-6) \) we solving for \( \lambda \), get \( \hat{\eta} \):

\[ \hat{\eta} \text{ from above} \]
Correlation or Variance Matrices

\[ \text{Cov}(X_i, X_j) = \text{C}_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)] \]
\[ = E(X_i X_j) - E(X_i)E(X_j) \]

\[ \text{Var}(X_i) = \text{C}_{ii} = \sigma_i^2 = E[(X_i - \mu_i)^2] = E(X_i)^2 - (E(X_i))^2 \]

RVs \( X_i \) called mutually uncorrelated if \( \text{C}_{ij} = 0 \) for \( i \neq j \)

In that case, for

\[ X = X_1 + \cdots + X_n \]
\[ \sigma_X^2 = \sigma_1^2 + \cdots + \sigma_n^2 \]

Ex: 7-5 Sample mean \& sample variance

\[ \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \]
\[ \bar{V} = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]

If \( X_i \) uncorrelated with some mean \( E(X_i) = \mu \)

and variance \( \sigma_i^2 \), then \( \sigma_i^2 = \sigma^2 \), then

\[ E(\bar{X}) = \mu \]
\[ \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} \]

and \[ E(\bar{V}) = \sigma^2 \]
Proof:  

(i) \[ E(\bar{X})^2 = E \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right)^2 = \frac{1}{n} \sum_{i=1}^{n} E(X_i^2) = \sigma^2 \]  

(ii) \( n = 2 \) \[ X = X_1 + X_2 \]  
\[ \sigma_x^2 = \sigma_1^2 + \sigma_2^2 = 2 \sigma^2 \]  
\[ \frac{X}{2} = \frac{X_1 + X_2}{2} \]  
\[ \sigma_x^2 = \frac{1}{4} 2 \sigma^2 = \frac{\sigma^2}{2} \]  
\[ \bar{X} = \frac{1}{2} (X_1 + X_2) \]  
\[ \sigma_x^2 = \frac{\sigma^2}{n} \]  

(iii) For the sample variance \[ \bar{V} = \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X})^2 \]  
\[ E(\bar{V}) = \sigma^2 \]  
\( \text{Cov}(i \neq j) \)  

Proof: text.
Correlation matrix

\[ X = [X_1 \; X_2 \cdots X_n] \]

\[ R_n = E \left[ \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}^T \right] \]

Covariance matrix

\[ C_n = E \left[ \begin{bmatrix} X_1 - \mu_1 \\ \vdots \\ X_n - \mu_n \end{bmatrix} \begin{bmatrix} X_1 - \mu_1 \\ \vdots \\ X_n - \mu_n \end{bmatrix}^T \right] \]

\( \mu \): variance of "centered" data \( X_i - \mu_i \)

Properties of \( R_n, C_n \)

\[ C = \alpha^T R_n \alpha \geq 0 \quad \forall \; \alpha \neq 0 \]

Proof: \( n = 2 \)

\[ E \left[ (\alpha X_1 + \alpha_2 X_2)^2 \right] \geq 0 \]

\[ = E \left[ \alpha^2 (X_1, X_2) (X_1, X_2)^T \right] \]

\[ = \sum_{i,j=1}^{2} \alpha_i \alpha_j E(X_i X_j) = \alpha^T R_2 \alpha \]
RV Independence

RV $X_i$ linearly independent if

$$E \left[ \left( \sum a_i X_i \right)^2 \right] > 0 \quad \text{all } a_i$$

($E_n \text{ is PD}$)

RV $X_i$ linearly dependent if

$$a_1 X_1 + \cdots + a_n X_n = 0 \quad \text{for some } a_i$$

Show corresponding $Q = 0$

($R_n \text{ is PD}$)

Correlation matrix $R_n$ determinant.

$$R_n \text{ PSD } \Rightarrow |R_n| \geq 0$$

$$R_n \text{ PD } \Rightarrow |R_n| > 0$$
**Normal Vectors**

Recall, \( x, y \) jointly normal if

\[
fx(y) = A \exp \left[ -\frac{1}{2(1-r^2)} \left( \frac{(x-\mu_1)^2}{\sigma_1^2} - 2r \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1 \sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right) \right]
\]

\[
A = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-r^2}} \quad |r| < 1
\]

Recall \( r = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y} \)

Recall also equivalent statements:

2 RV unrelated if:

\[
\text{Cov}(x,y) = 0 \quad \text{or} \quad \rho_{xy} = 0 \quad \text{or} \quad E(xy) = E(x)E(y)
\]

Also, \( x \perp y \) if:

\[
E(xy) = 0
\]

Ex: 6-30 show that for \( f(x,y) \) given above (using \( \sigma_1 = \sigma_2 = 0 \))

\[
E(xy) = \rho \sigma_1 \sigma_2
\]
$N$ random variables $X_1, X_2, \ldots, X_n$

jointly normal if joint pdf $f(x)$

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \Delta}} \exp \left\{ -\frac{1}{2} X C^{-1} X^T \right\}$$

where $X = [X_1, X_2, \ldots, X_n]^T$

$C$ covariance matrix

$\Delta = \text{det}(C)$
7.3 MEAN SQUARE ESTIMATION

Ex:  S - sample of all children in a community.

Y - RV: \( S \rightarrow \mathbb{R} \) height of a child.

i.e. \( y(y) = \text{a number for child } y \).

Problem: Estimate height by a constant \( c \).

i.e. \( \text{Min } E \{ (Y-c)^2 \} = \int_{-a}^{a} (y-c)^2 f(y) \, dy \) (7.63)

\[ \frac{\partial E}{\partial c} = \int_{-a}^{a} 2(y-c) f(y) \, dy = 0 \]

\[ c = \int_{-a}^{a} y f(y) \, dy = E(Y) \]

Ex: Subtract now that each child is weighted.

On the basis of this, the height can be improved.

Problem: Estimate height \( Y \) by a function \( c(x) \) where \( x \) is weight of child.
Minimise MS estimation

So: \[ \min e = \mathbb{E} \left[ (Y - c(x))^2 \right] \]

\[ e = \iint (y - c(x))^2 f(y|x) dxdy \]

\[ = \iint (y - c(x))^2 f(y|x) dxdy \]

\[ e = \int f(x) \int (y - c(x))^2 f(y|x) dy \]  \[ dx \]  

smallest integrand possible.

Hence \( e = \min \) if inner integral minimum for every \( x \).

But \( \int (y - c(x))^2 f(y|x) dy \) same as \( \int (y - c(x))^2 f(y) dy \) (7.11)

if \( e \to c(x) \) and \( f(y) \to f(y|x) \).

\[ \therefore \text{Solution} \quad c(x) = \int y f(y|x) dy = E[Y|X = x] \]

If \( Y = g(x) \), then \( c(x) = g(x) \).

Also, \( \int y f(y|x) dy = \int g(x) \frac{f(y|x)}{f(x)} dy = g(x) \)

And, if \( X, Y \) independent, then \( E[Y|X] = E[Y] = \text{constant} \).

Knowledge of \( X \) has no effect on \( Y \).
**Linear MS Estimation**

\[ y = ax + b \]

\[ \text{I.e. } \min_{a,b} E \left( (y - (ax + b))^2 \right) = \min_{a,b} e \]

**Solution:**

\[ a = \frac{\mu_1}{\mu_2} = \frac{E(x-\bar{x})(y-\bar{y})}{E((x-\bar{x})^2)} = \frac{\text{cov}(x,y)}{\text{var}(x)} \]

\[ b = \bar{y} - a \bar{x} \]

**Proof:** Ths p. 263

**Note:** \( x, y \) normal, nonlinear, linear estimate the same.

**Orthogonality Principle**

From (7-72) can show \( e \) min. if

\[ \frac{\partial E}{\partial a} = \frac{\partial E}{\partial b} = 0 \]

\[ \frac{\partial E}{\partial a} = E \left[ 2 (y - (ax + b)) (-x) \right] = 0 \]

\[ \Rightarrow E \left[ (y - (ax + b)) \cdot x \right] = 0 \]

\( \text{i.e. Estimation error } (y - (ax + b)) \perp x. \)

(Orthogonality Principle)
Homogeneous case:

\[ y \sim ax \]

\[ \min \ e = E[(y-ax)^2] \]

Ans: \( (y-ax) \perp x \)

\[ \text{i.e. } E[(y-ax)x] = 0 \]

And \( E[y|x] \) - linear MS estimate of \( y \) given \( x \)

\[ E[y|x] = ax \quad \text{where} \quad a = \frac{E(xy)}{E(x^2)} \]
**Linear Estimation (Normal One)**

\[ S = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n \]

is the linear estimate of \( S \) in terms of \( n \) RVs \( X_1, X_2, \ldots, X_n \).

Determine constants \( a_i \) to minimize estimation error

\[ E = E \left[ (S - S')^2 \right] = E \left[ E \left[ (S - (a_1 X_1 + \ldots + a_n X_n))^2 \right] \right] \]

\[ \text{minimized.} \]

**Orthogonality Principle**

\[ \text{Err} (S - S') = \text{truth} \quad X_i \]

i.e.,

\[ E \left[ E \left[ S - (a_1 X_1 + \ldots + a_n X_n) \right] \cdot X_i \right] = 0 \quad i = 1, 2, \ldots, n \]

**Proof:** First set \( \frac{\partial E}{\partial a_i} = 0 \) to get above result.

\[ \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix} \begin{bmatrix} n_1 & n_2 & \ldots & n_n \\ n_1 & n_2 & \ldots & n_n \\ \vdots & \vdots & \ddots & \vdots \\ n_1 & n_2 & \ldots & n_n \end{bmatrix} = \begin{bmatrix} \sum \left( X_i \right) \\ \sum \left( X_i X_j \right) \\ \vdots \\ \sum \left( X_i^n \right) \end{bmatrix} \]

\[ R_{ij} = E \left[ X_i X_j \right] \quad R_{ij} = E \left[ S X_j \right] \]

\[ R = \begin{bmatrix} n_1 & \sum \left( X_i X_j \right) \\ \sum \left( X_i X_j \right) & \sum \left( X_i^n \right) \end{bmatrix} \]

\[ R_{ij} = \frac{1}{n} \quad \text{and} \quad a = R^{-1} n \]
7.4 **Stochastic Convergence or Limit Theorems.**

Suppose we want to measure length $\lambda$ on object $X = a + N$.

Assume $N$ is an integer.

Intuitively, if $\sigma$ is of order small compared to $\lambda$, then observed value $X(a)$ of $X$ in a single measurement is a satisfactory estimate.

Or, in the context of probability,

$$P(|X - a| < \epsilon) > 1 - \frac{\sigma^2}{\epsilon^2}$$

If $\sigma < \epsilon$,

then prob. $|X - a| < \epsilon$ close to 1.

: almost certainly $X$ lies in $(a - \epsilon, a + \epsilon)$

If $\sigma$ not small compared to $\epsilon$, then take a number of measurements in average.
So underlying probability model in part gives

\[ S^2 = S_1^2 \times S_2^2 \cdots \times S_k^2 \]

i.e. experiment is repeated \( n \) times.

If measurements independent, then

\[ X_i = a + N_i \]

\( N_i \) zero mean \( \mu_i \) variance \( \sigma_i^2 \)

Then the sample mean

\[ \bar{X} = \frac{X_1 + \cdots + X_n}{n} \]

is a RV with mean \( a \) and variance \( \frac{\sigma^2}{n} \).

So intuitively if \( \frac{\sigma^2}{n} \ll \sigma^2 \), set good estimate in a single performance of \( \exp S^2 \) (i.e. \( n \) measurements).

To bound the error, so learn to Chebyshev:
Let $n$ be so large that \[ \frac{\sigma^2}{n\sigma^2} = 10^{-4}. \]

Qs: What is prob. that $X$ lies between $\cdot9a$ and $1.1a$?

Tchebychev: \[ P \left( a-e < X < a+e \right) > 1 - \frac{\sigma^2}{e^2}. \]

But $e = 0.1a$

So, \[ P \left( 0.9a < X < 1.1a \right) > 1 - \frac{\sigma^2}{n \cdot (0.1a)^2}. \]

\[ > 1 - 10^{-4} \times 10^{-2} = 0.99 \]

So, if NPT performed $n = \frac{10^4 \cdot \sigma^2}{a^2}$. Hence,

almost certainly estimate $X$ if $n$ will be between $\cdot9a$ and $1.1a$. 

Law of Large Numbers \( (p275\text{Text}) \).

Recall Bernoulli Th. \( (p38\text{Text}) \).

Let \( A \) denote event with prob. \( P \) of occurrence in a single trial \( = p \).

If \( n \) denotes \# of occurrences of \( A \) in \( n \) independent trials, then

\[
P\left( \left| \frac{k}{n} - p \right| > \varepsilon \right) < \frac{pq}{n\varepsilon^2} \tag{3.27}
\]

(axiomatic \( \frac{pq}{n\varepsilon^2} \) def. of probability)

Equation (3.27) states that the above 2 can be made compatible with almost any def desired before \( n \) accuracy, (provided we take a large \# of trials) (Proof \( p38\text{Text} \))
The same result can be established as a limit of i.i.d. random variables.

Let \( X_i = 1 \) if \( A \) occurs and \( 0 \) otherwise.

Can show that the sample mean

\[
\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \to p
\]

as \( n \to \infty \).

Proof. p.276 Text.
Central Limit Theorem (CLT)

Given $n$ independent $N(0, \sigma^2)$ $X_i$, the sum
\[ X = X_1 + X_2 + \cdots + X_n \]

Then
\[ \bar{X} = \frac{X_1 + \cdots + X_n}{n} \]
\[ \sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2 \]

CLT states that (under certain regular conditions)
\[ F_X(x) \approx \Phi \left( \frac{x - \mu}{\sigma} \right) \]

as $n \to \infty$.

That is, CLT can be expressed as a property of convolutions of positive functions.

Ex: $X_1$ i.i.d. in $[0, T]$.

\[
\begin{align*}
\frac{1}{T} f_t(x) &\quad \frac{1}{T} \begin{cases} 1 \quad &0 \leq x < T \\ 0 \quad &\text{otherwise} \end{cases} \\
X = X_1 + X_2 &\quad X = X_1 + X_2 + X_3
\end{align*}
\]