Continuous Probability Distributions and Stochastic Dependencies
(Bertsimas & Freund 2004, Chapter 3)

Howard Hao-Chun Chuang (莊皓鈞)

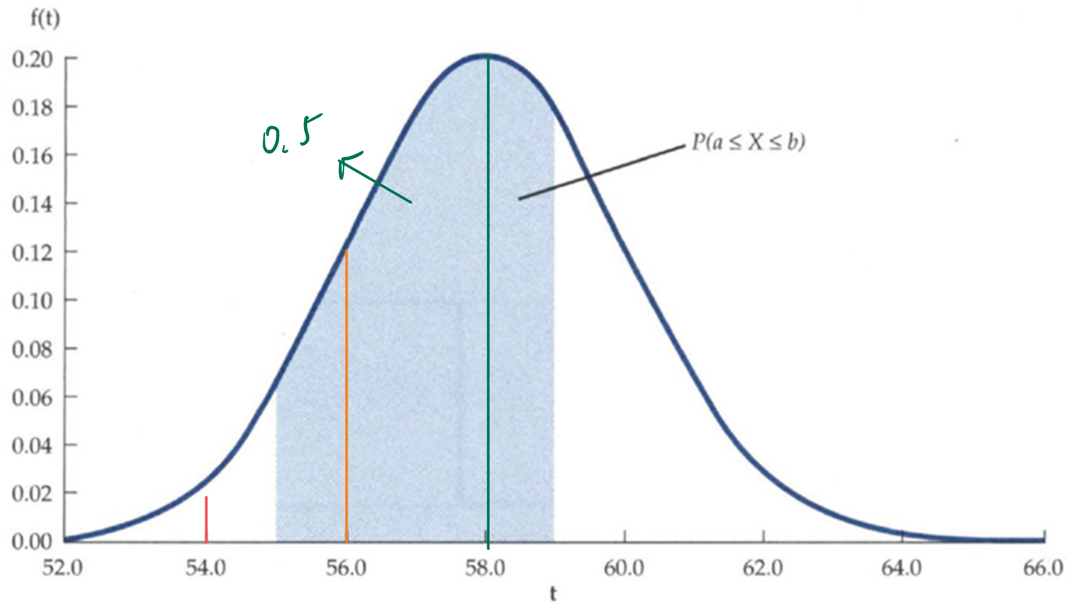
Professor
College of Commerce
National Chengchi University

October, 2024
Taipei, Taiwan

© 2024 H. Chuang All rights reserved

Random Variables

- For a **continuous random variable (RV)** X, its probability density function (pdf) $f(t)$ has the following characteristics
  - The area underlying the pdf curve is equal to one
  - $P(a <= X <= b)$ is equal to the area under the curve between a and b

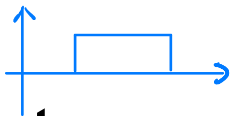


For any t, $P(X=t)=0$ https://en.wikipedia.org/wiki/Probability_density_function

© 2024 H. Chuang All rights reserved

Random Variables

- For a **continuous random variable (RV)** ***X***, its probability density function (pdf) $f(t)$ has the following characteristics
  - The area underlying the pdf curve is equal to one
  - P(a<= X <=b) is equal to the area under the curve between a and b

- **X ~ Uniform(a, b)** 連續型均勻分布 
  - uniform[loc, loc+scale] in scipy.stats

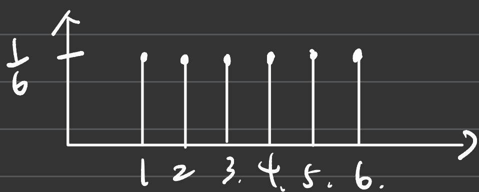
$$f(t) = \left\{ \begin{matrix} \dfrac{1}{b-a} & if\ a \leqq t \leqq b, \\ 0 & \text{otherwise} \end{matrix} \right\}$$

  - e.g., Automobile Life (L) ~ Uniform[2, 7.5]
    - What would the pdf of L look like?
  - Uniform(0, 1) is extremely important for simulation!

© 2024 H. Chuang All rights reserved

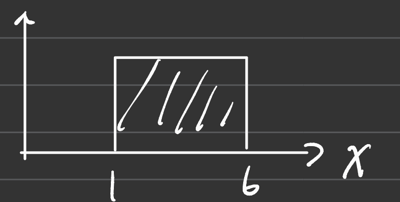
$X \sim$ Discrete Uniform $(1, 6)$

$f(x) = P(x)$



$X \sim$ Continuous Uniform

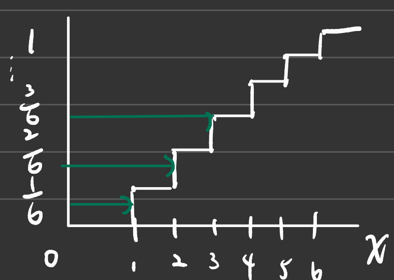
$f(x)$ prob density



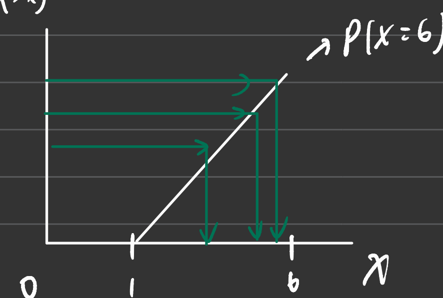
CDF: $P(X \le x^{max}) = 1$, $P(X < x^{min}) = 0$

$F(x) = P(X = x)$

*: CDF 只能从 0. ~ 1.



$F(x) = P(X = x)$



电脑去模拟.

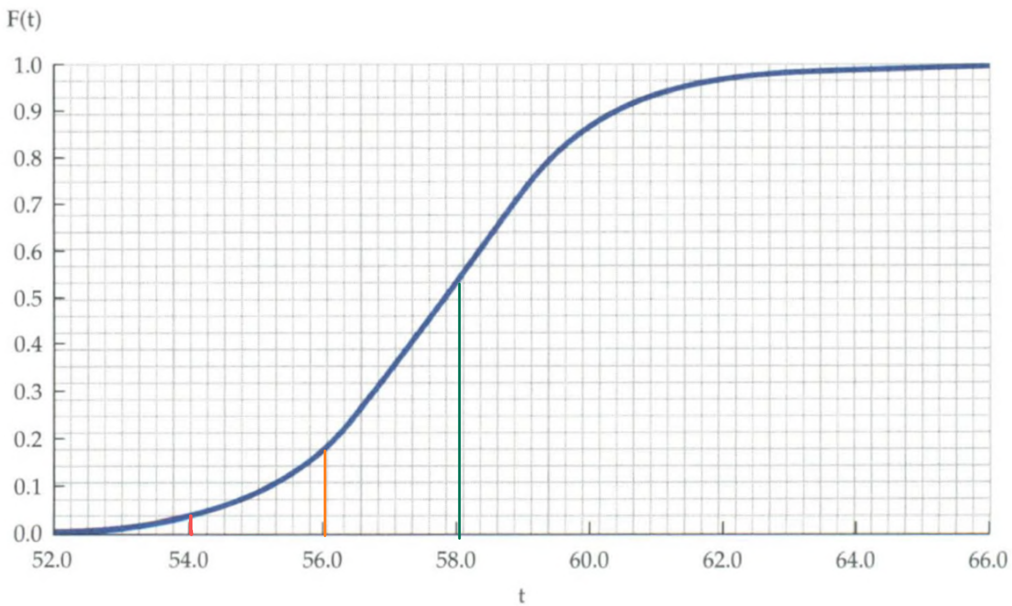
Random Variables

- For a **continuous RV *X***, its cumulative distirubition function (cdf) of *X* is *F*(*t*)=*P*(*X*<=*t*)

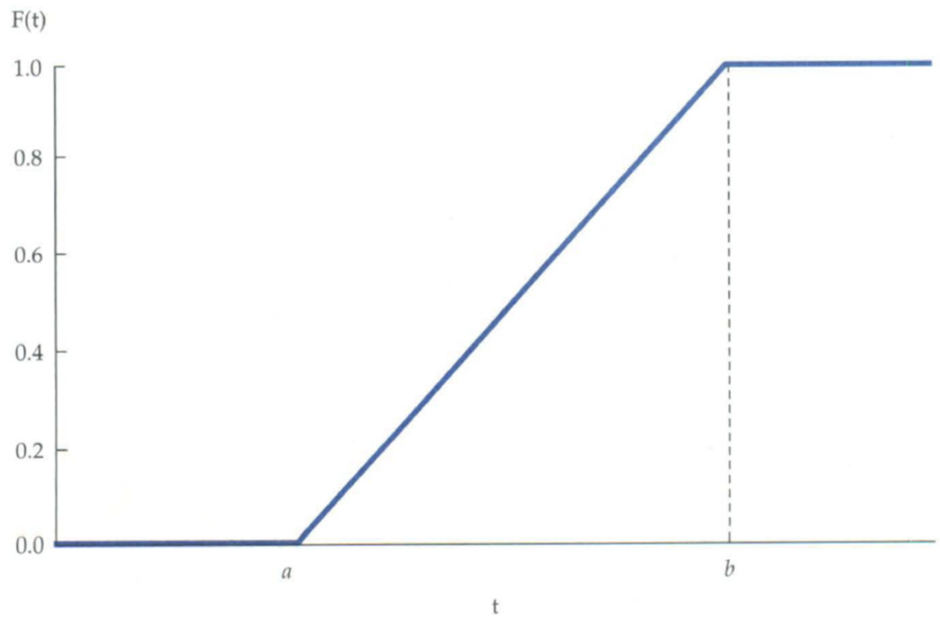
  P(*X* > *t*) = 1 – *F*(*t*); $\boxed{P(c\le X\le d) = F(d) - F(c)}$ why

  ? We **can interchange <= & <** in any probability expressions involving *X*

  X: Width of a steel plate

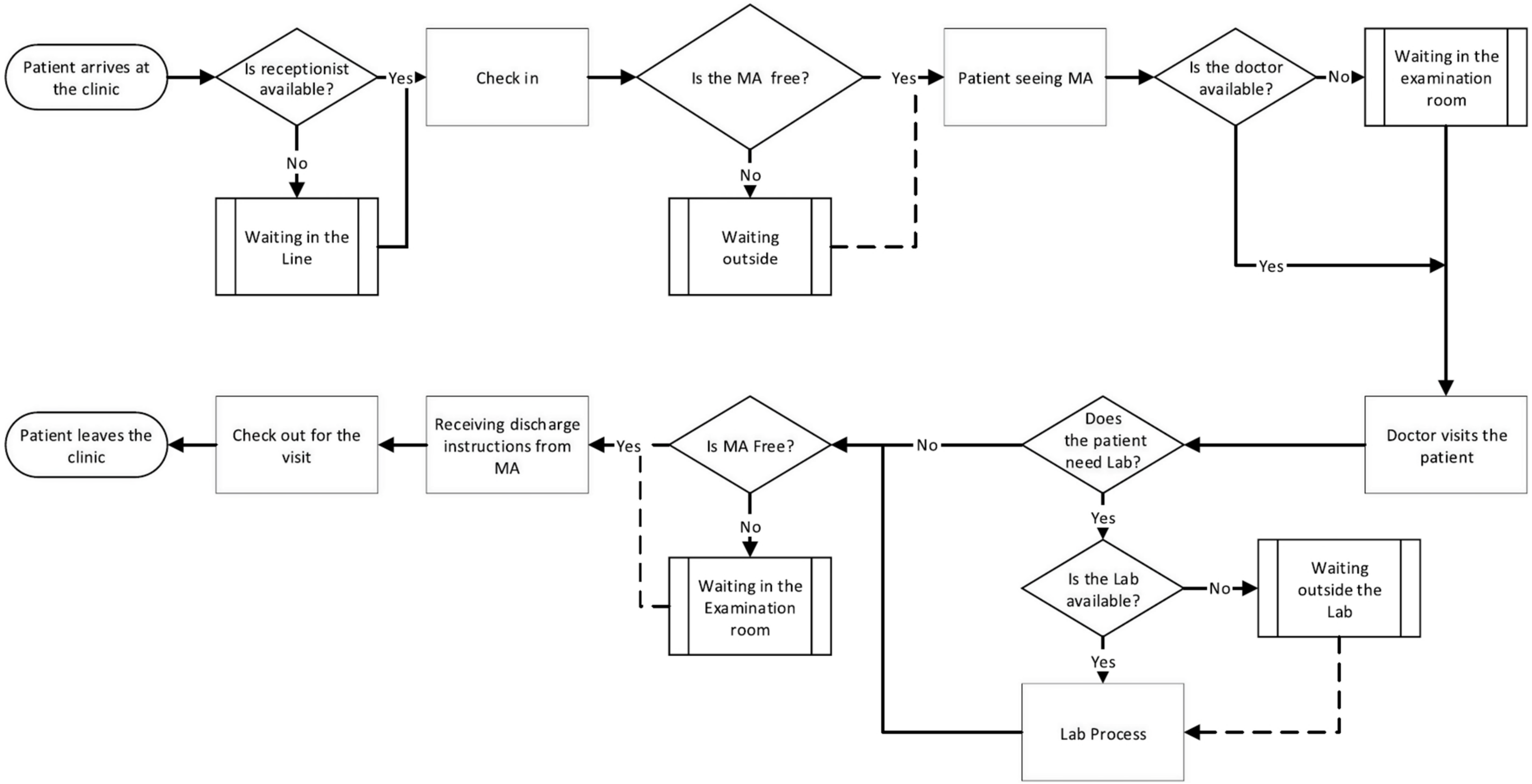


X: A Uniform(a, b) RV



© 2024 H. Chuang All rights reserved

Continuous Uniform (a, b) in Practice



© 2024 H. Chuang All rights reserved

The process starts with the arrival of a patient with a scheduled appointment. If either receptionist is available, the patient will start the primary check-in process; otherwise, the patient will wait in the registration area. For an established patient, the primary check-in time is set to be $C1 \sim U(2, 3)$ minutes, that is, a uniformly distributed random variable between two and three minutes. For a new patient, they experience an additional paperwork time $P \sim U(15, 20)$ minutes. As per COVID-19 protocols, all patients will leave the clinic after the primary check-in and wait outside until an MA calls them in. Usually, a primary care doctor has a designated MA during a shift. The MA is responsible for settling the patient in an available examination room. The MA's service time is set as $M1 \sim U(8, 12)$ minutes, including basic questionnaires and vital measurements. After this, the patient will wait for the doctor to become available. TI's historical data show that service times for primary care doctors fit well with the uniform distributions in Table 1.

After the doctor sees the patient, if the patient needs any laboratory testing done, they will go directly to the laboratory; otherwise, the patient will wait for the MA to give discharge instructions, which is set to be $M2 \sim U(2, 6)$ minutes. The laboratory service time is set as $L \sim U(8, 12)$ minutes. Finally, the patient will check out, taking $C2 \sim U(2, 6)$ minutes, and leave the clinic.

	Primary care 1	Primary care 2
Established	$[U(15,20), 24]$	$[U(8,12), 48]$
New	$U(35,45)$	$U(30,40)$
Monday	p.m.	a.m. and p.m.
Wednesday	p.m.	a.m. and p.m.

© 2024 H. Chuang All rights reserved

The Normal Distribution (section 3.4)

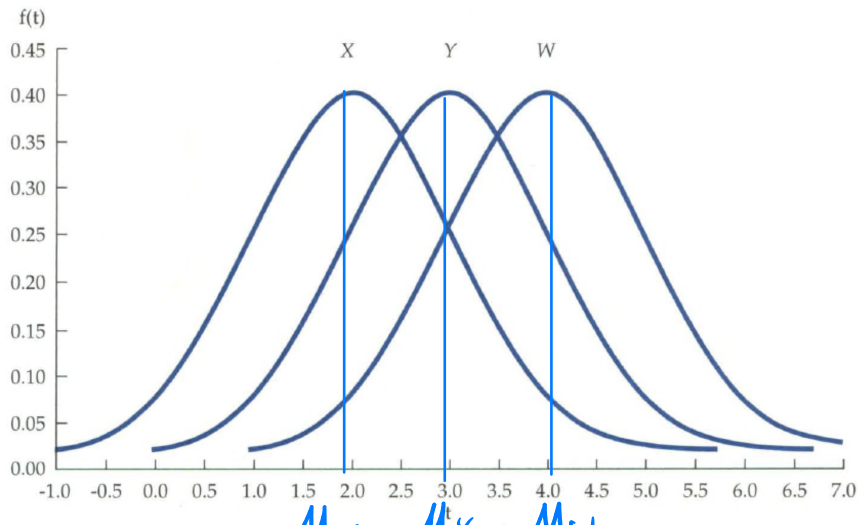
- ***X* ~ Normal(*mu*, *sigma*)** $\mu \quad \sigma$ $E(x) = \mu$ 中心在哪 $Var(x) = \sigma^2$ 有多分散

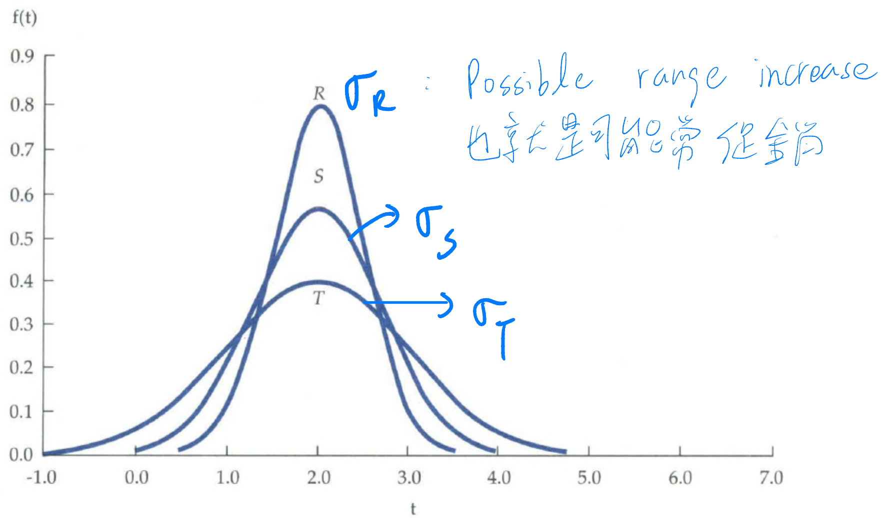
  ***f*(x): norm.pdf (scipy)**

  ***P*(*X* <= *x*)/*F*(x): norm.cdf**

  Find the smallest *k* such that *P*(*X* <= *k*)>=*q*: norm.ppf Quantile

  Simulate *S* random realizations: norm.rvs





Z = (*X*-*mu*)/*sigma* ~ Standard Normal(0, 1) → Normalization

$z = \frac{(x - \bar{x})}{\sigma_x}$ 標準化

© 2024 H. Chuang All rights reserved

The Normal Distribution (section 3.4)

- ***X* ~ Normal(*mu*, *sigma*)**

  ***f*(x): norm.pdf (scipy)**

  ***P*(*X* <= *x*)/*F*(x): norm.cdf**

  Find the smallest *k* such that *P*(*X* <= *k*)>=*q*: norm.ppf

  Simulate *S* random realizations: norm.rvs

開始時間

	Underwriting Step 1	Rating Step 2
(100, 80) Case 1	0.	100.
(170, 30) Case 2	100.	270 (>180)
(110, 90) Case 3	270.	

Example 3.8 in page 130

Issuing an insurance policy requires two steps – *underwriting* the policy with process time (a normal RV X_1 with μ_1=150 mins & σ_1=30 mins) and *rating* the policy with process time (a normal RV X_2 with μ_2=75 mins & σ_2=25 mins).

Assuming X_1 and X_2 are independent of each other.

What is the probability that it takes 180 minutes or less to issue a policy?

What is the 95th percentile of the time for issuing a policy?

What is the probability of issuing three policies in an 8 hour day?

What if X_1 and X_2 are NOT independent of each other?

© 2024 H. Chuang All rights reserved

$\boxed{\text{Under}} \longrightarrow \boxed{\text{Rating}} \longrightarrow$

$\sigma = 30$ independent $\sigma = 25$ $X = X_1 + X_2$

150 X_1 75 X_2

who ≤ 180 → $P(X \le 180)$ ← cdf

$N(\mu = 150+75, \sigma = \sqrt{30^2+25^2})$

S_1 S_2 S_3 130 160 78 89 = 206 = 249

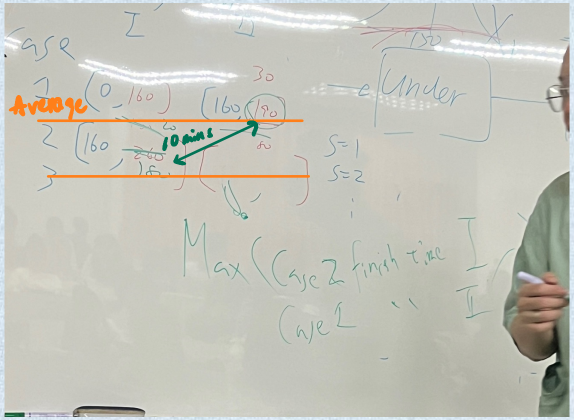
2.9 Covariance & Correlation (Table 2.23)

ppf

$P(X \le ?) \ge 0.95$

找到第95分位數

Probability	Number of Sunglasses Sold	Number of Umbrellas Sold
p_i	x_i	y_i
0.10	35	41
0.15	78	10
0.05	81	0
0.10	30	13
0.20	16	42
0.05	29	22
0.10	35	1
0.10	14	26
0.10	52	11
0.05	46	23



$P(X = x, Y = y)$

CoRR > 0

CoRR < 0

相關係數

© 2024 H. Chuang All rights reserved

2.9 Covariance & Correlation

- **Covariance for two random variables *X* and *Y***

$$\mathrm{COV}(X, Y) = \sum_{i=1}^{n} p_i(x_i - \mu_X)(y_i - \mu_Y)$$

where $p_i = P(X=x_i, Y=y_i)$ is the joint probability distribution function

e.g., outcome of rolling two fair dies

X & Y are independent if $P(X=x, Y=y)=P(X=x)*P(Y=y)$

e.g., study hour (X) & test score (Y)

獨立：各自相乘!

X \ Y	Y_1 71-80	Y_2 81-90	Y_3 91-100	Total
x_1 1 Hour	7	5	2	14
x_2 2 Hours	8	3	3	14
x_3 3 Hours	3	5	8	16
x_4 4 Hours	4	9	7	20
Total	22	22	20	64

$P(X=x_1) = \frac{14}{64} = \frac{7}{32}$

$P(Y=y_1) = \frac{22}{64} = \frac{11}{32}$

$P(X=x_1 \cap Y=y_1) = \frac{7}{64}$

$P(X=x_1) * P(Y=y_1) \neq P(X=x_1 \cap Y=y_1)$

© 2024 H. Chuang All rights reserved

$\frac{1}{6}$ $\frac{2}{6}$ $\frac{3}{6}$

2.9 Covariance & Correlation

共變異數：兩個數字的共同變化關係 你變大，我變大

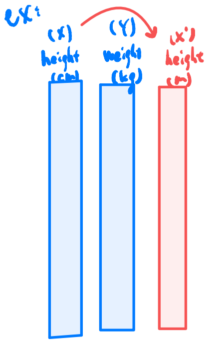
- **Covariance for two random variables *X* and *Y***

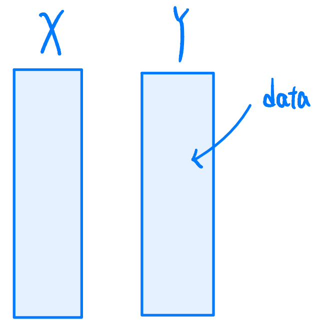
>0, <0, =0

$$\mathrm{COV}(X, Y) = \sum_{i=1}^{n} p_i (x_i - \mu_X)(y_i - \mu_Y)$$

範圍沒限制 取決於尺度

where $p_i = P(X=x_i, Y=y_i)$ is the joint probability distribution function
e.g., outcome of rolling two fair dies
***X* & *Y* are independent if P(*X*=*x*, *Y*=*y*)=P(*X*=*x*)*P(*Y*=*y*)**





不受單位影響

- **Correlation of two random variables *X* and *Y***

$$\mathrm{CORR}(X, Y) = \frac{\mathrm{COV}(X, Y)}{\sigma_X \sigma_Y}$$

COV標準化後的數值，不受單位影響

完美負相關 ("正")

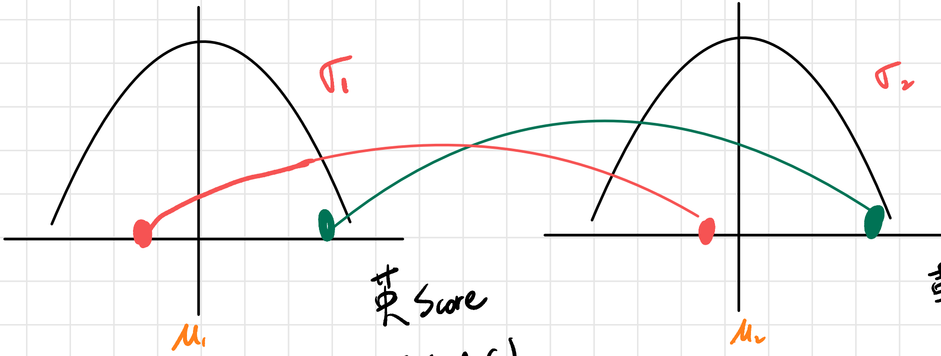
ranging in [-1, 1]. If COV(*X*, *Y*)=0, CORR(*X*, *Y*)=0
any examples on random variables with positive/negative COV/CORR?

$\begin{cases} \mathrm{cov}(X, Y) \\ \mathrm{cov}(X', Y) \end{cases}$ different

- **What are the covariance & correlation of the example above?**

see pages 85-86 in Chapter 2

© 2024 H. Chuang All rights reserved



英 Score (TSMC) Demand A $\longleftrightarrow$ 數 Score (MTK) Demand B

$$\begin{cases} Cov(X_1, X_1) = \Sigma (x - \bar{x})^2 \\ \sigma^2 = \dfrac{\Sigma (x-\bar{x})^2}{n} \\ Cov(X_1, X_1) = n\sigma^2 \end{cases}$$

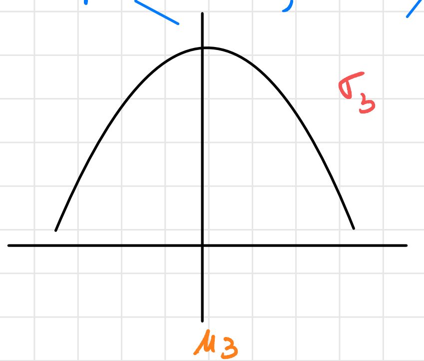
$$\underbrace{\overset{cov}{\Sigma}_{m \text{ by } m}}_{\text{矩陣}} = \begin{array}{c|cc} & x_1 & x_2 \\ \hline X_1 & \overset{\sigma_1^2}{Var}(X_1), & cov(X_2, X_1) \\ X_2 & cov(X_1, X_2), & \overset{\sigma_2^2}{Var}(X_2) \end{array}$$

$$CORR_{m \text{ by } m} = \begin{bmatrix} 1 & CORR(X_2, X_1) \\ CORR(X_1, X_2) & 1 \end{bmatrix}$$

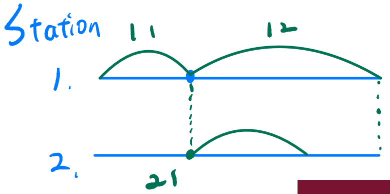
$\to -1 \sim 1$

$$CORR(X_1, X_1) = 1 = \frac{Cov(X_1, X_1)}{\sigma_{X_1} \cdot \sigma_{X_1}} = \frac{Var(X_1)}{\sigma_{X_1}^2} = 1$$

if



$$\Sigma_{3 \text{ by } 3} = \begin{array}{c|ccc} & X_1 & X_2 & X_3 \\ \hline X_1 & \overset{\sigma_1^2}{Var}(X_1), & cov(X_2, X_1), & cov(X_3, X_1) \\ X_2 & cov(X_1, X_2), & \overset{\sigma_2^2}{Var}(X_2), & cov(X_3, X_2) \\ X_3 & cov(X_1, X_3), & cov(X_2, X_3), & \overset{\sigma_3^2}{Var}(X_3) \end{array}$$



The Multi-Variate Normal (MVN) Distribution

- ***X* ~ Multi-Variate Normal(*mu*, *Sigma*)**

都常態

X is an array of *m*-dimensional random variables (X_1, X_2,…, X_m)

m個隨機變數

***Sigma* (Σ) is a variance-covariance matrix**

Simulate *S* random realizations: multivariate_normal.rvs in scipy

Revisit Example 3.8 in page 130

Issuing an insurance policy requires two steps – *underwriting* the policy with process time (a normal RV X_1 with μ_1=150 mins & σ_1=30 mins) and *rating* the policy with process time (a normal RV X_2 with μ_2=75 mins & σ_2=25 mins).

Assuming X_1 and X_2 are <u>NOT independent</u> of each other, and CORR(X_1 , X_2)=0.37

What is the covariance of X_1 and X_2?

$0.37 * \overset{\sigma_1}{30} * \overset{\sigma_2}{25}$

$CORR(X_1, X_2) = \frac{COV(X_1, X_2)}{\sigma_{X_1} \times \sigma_{X_2}}$

?

$\boxed{\text{What is the probability distribution of } X\text{?}}$

What is the probability that it takes 180 minutes or less to issue a policy?

What is the 95th percentile of the time for issuing a policy?

What is the probability of issuing three policies in an 8 hour day?

© 2024 H. Chuang All rights reserved

X是二維多元常態分布

$X \sim N(\mu, \Sigma)$

服從於…分布情況

$\mu = (150, 75)$

$\Sigma = \begin{pmatrix} 30^2, & 277.5 \\ 277.5, & 25^2 \end{pmatrix}$

協方差矩陣: covariance of x_1 and x_2

$\begin{bmatrix} \sigma^2 & cov \\ cov & \sigma^2 \end{bmatrix}$

The Multi-Variate Normal (MVN) Distribution

- ***X* ~ Multi-Variate Normal(*mu*, *Sigma*)**

  X **is an array of *m*-dimensional random variables (X_1, X_2,…, X_m)**

  Sigma **(Σ) is a variance-covariance matrix**

  Simulate *S* random realizations: multivariate_normal.rvs in scipy

  Example: Capturing stochastic dependencies

  Below is the end-of-year price for four stock funds from 2007-2017

  Fund1: 65, 79, 85, 78, 107, 108, 124, 156, 195, 181, 216

  Fund2: 47, 61, 73, 60, 89, 86, 104, 120, 140, 134, 175

  Fund3: 38, 37, 39, 40, 47, 46, 57, 71, 74, 72, 87

  Fund4: 61, 64, 74, 72, 95, 89, 114, 147, 146, 127, 152

  $\vec{a} = (a_1, a_2, a_3)$

  $\vec{b} = (b_1, b_2, b_3)$

  $\vec{a} \times \vec{b} = \left(\begin{vmatrix} a_2 & a_3 \\ b_2 & b_3 \end{vmatrix}, \begin{vmatrix} a_3 & a_1 \\ b_3 & b_1 \end{vmatrix}, \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \right)$

  What are the annual growth ratios respectively?

  How about estimating *Sigma* (Σ) from data?

  Assuming end-of-year price is a RV *X* ~ MVN(*mu*, *Sigma*), how can we simulate 2018 growth ratios for the four stock funds?

© 2024 H. Chuang All rights reserved

if Not Normal ?

Amazon

Sigma (Σ)

$$\underset{m \text{ by } m}{\Sigma} = \begin{array}{c|cccc} & z_1 & z_2 & \cdots & z_m \\ \hline z_1 & Var(z_1) & & & \\ z_2 & & Var(z_2) & & \\ \vdots & & & \ddots & \\ z_m & & & & Var(z_m) \end{array}$$

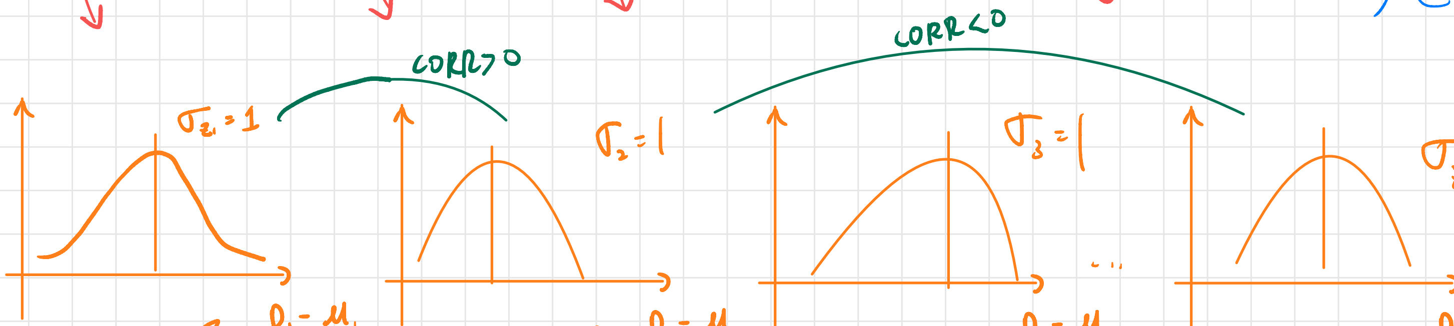
Observed Varience

Why (Muhat, Sigmahat, S)

?

Why Cov ✓ X Corr

Demand$_1$ D_1 Demand$_2$ D_2 Demand$_3$ D_3 Demand$_m$ D_m



$$z_1 = \frac{D_1 - \mu_1}{\sigma_1} \qquad z_2 = \frac{D_2 - \mu_2}{\sigma_2} \qquad z_3 = \frac{D_3 - \mu_3}{\sigma_3} \qquad z_m = \frac{D_m - \mu_m}{\sigma_m}$$

To Hedge or Not to Hedge?

Exchange loss 匯損
Put option 賣權

© 2024 H. Chuang All rights reserved

To Hedge or Not to Hedge? 避險

- Company D has $2.1 billion revenue in 1997
  - see its global revenue distribution below

Region	Revenues ($ million)	Percent of all Revenues (%)
United States	1,050	50
Germany	420	20
Great Britain	336	16
Rest of Europe	84	4
Asia	210	10

  - USD$420 million (Germany) from 645 million DM (Deutsche Mark)
    - Exchange rate 0.6513 USD$/DM ; 645x0.6513=420 million
    - If USD goes strong, e.g., 0.6072 USD$/DM, only $392 million
  - D is considering 賣權(put options) to hedge against strong USD in 1998
    - use guaranteed dollar price for DM-USD exchange in 1998
    - expiration date (98年的到期日)
    - strike price *k*
    - unit cost *c*

© 2024 H. Chuang All rights reserved

To Hedge or Not to Hedge?

- **DM put options**
  - Random exchange rate in the future
  - DM1 (DM-USD rate) on expiration date, the **net payoff** for put

$$
\begin{array}{ll}
k - DM1 - c & if\ DM1 \le k \\
-c, & if\ DM1 > k
\end{array}
$$

  - equivalent $max(k - DM1, 0) - c$

    可以賣在 0.62

    賣權

  - for a 1-year DM-USD put with *k*=\$0.62 & *c*=\$0.0137
    - If USD strong, DM1=0.5 USD\$/DM. The net payoff will be

$$\underbrace{max(0.62 - 0.5, 0)}_{\text{賺}} - \underbrace{0.0137}_{\text{成本}} = \$0.1063$$

    - If USD weak，DM1=0.64 USD\$/DM. The net payoff will be

$$max(0.62 - 0.64, 0) - 0.0137 = \underbrace{-\$0.0137}_{\text{為了歸避風險付出的代價}}$$

  - Uncertainty: DM Rate (USD\$/DM)
    - suppose R_{DM} ~ Normal(mu=0%, sigma=9%)
    - 0.6513 USD\$/DM in 1997

$$DM1 = 0.6513 \times (1 + \boxed{\frac{R_{DM}}{100}})$$

匯率的波动性

© 2024 H. Chuang All rights reserved

To Hedge or Not to Hedge?

- **1989-97 USD$/DM year-to-year change**

$$\text{Change from previous year} = 100 \times \left(\frac{\text{DM Rate in year } (t+1)}{\text{DM Rate in year } t} - 1 \right)$$

Year	*Average DM rate (US$/DM)*	*Change from previous year (%)*
1989	0.4914	—
1990	0.4562	−7.16
1991	0.5368	17.66
1992	0.5789	7.84
1993	0.5326	−7.99
1994	0.5542	4.05
1995	0.6285	13.40
1996	0.6394	1.73
1997	0.6513	1.86

© 2024 H. Chuang All rights reserved

To Hedge or Not to Hedge?

- **Company D has \$2.1 billion revenue in 1997**
  - **see its global revenue distribution below**

Region	*Revenues (\$ million)*	*Percent of all Revenues (%)*
United States	1,050	50
Germany	420	20
Great Britain	336	16
Rest of Europe	84	4
Asia	210	10

  - **\$336 million (UK) from 272 million BP (British Pound)**
    - **What will happen to 272 million, if USD goes strong?**
    - **suppose R_{BP} ~ Normal(mu=0%, sigma=11%)**
    - **1.234 USD\$/BP in 1997**
    - **BP1 in 1998**

$$BP1 = 1.234 \times (1 + \frac{R_{BP}}{100})$$

770

  - **historical data suggests CORR(R_{BP}, R_{DM})=0.675**

© 2024 H. Chuang All rights reserved

$COV(R_{DM}, R_{BP}) = CoRR \times \sigma_{DM} \times \sigma_{BP} = 0.675 \times 9 \times 11$

$$\Sigma_{aby^2} = \begin{bmatrix} \sigma_{DM}^2 & , \\ cov(R_{DM}, R_{BP}), & \sigma_{BP}^2 \end{bmatrix}$$

To Hedge or Not to Hedge?

- **D revenue forecast for 1998 = 1997 earning (NO growth)**
  - **98 forecasted revenue in Germany is 645 million DM**

利率的變化

$$\text{Unhedged Revenue} = 645 \times DM_1 = 645 \times \boxed{0.6513 \times \left(1 + \frac{R_{DM}}{100}\right)}$$

  - **if D buys n_{DM}(in million) put options (strike price k_{DM} & cost c_{DM})**

$$\text{Hedged Revenue} = \text{Unhedged Revenue} + \text{Number of Options} \times (\text{Net Payoff of the put option})$$

換回來賺的

- **take 1998 for instance**

$$\begin{aligned} \text{Hedged Revenue} &= 645 \times DM_1 \\ &\quad + n_{DM} \times [\max(k_{DM} - DM_1, 0) - c_{DM}] \\ &= 645 \times 0.6513 \times \left(1 + \frac{R_{DM}}{100}\right) \\ &\quad + n_{DM} \times \left[\max\left(k_{DM} - 0.6513\left(1 + \frac{R_{DM}}{100}\right), 0\right) - c_{DM}\right] \end{aligned}$$

選擇權價格

成本

© 2024 H. Chuang All rights reserved

To Hedge or Not to Hedge?

- **D revenue forecast for 1998 = 1997 earning (NO growth)**
  - **98 forecasted revenue in Germany is 645 million DM**
  - **if n_{DM} 500 million bought**
    - **k_{DM}=0.62 USD$/DM; cost c_{DM}=$0.00137**
  - **97 exchange rate 0.6513 USD$/DM. If DM is down by 23.23% in 1998**

$$\text{Unhedged Revenue} = 645 \times 0.6513 \times \left(1 + \frac{-23.23}{100}\right) = \$322.502 \text{ million}$$

口數

$$\begin{aligned}\text{Hedged Revenue} &= \text{Unhedged Revenue} \\ &\quad + \underline{500} \times \left[\max\left(0.62 - 0.6513 \times \left(1 + \frac{-23.23}{100}\right), 0\right) - 0.0137\right] \\ &= \$322.502 + \underline{\$53.148} \\ &= \$375.650 \text{ million.}\end{aligned}$$

損失變少

保護

  - **DM down by 23.23%, but exchanged $USD is only down by 10.6%**
    - **From $USD 420 million in 97 to $USD 375.65 million in 98**

© 2024 H. Chuang All rights reserved

To Hedge or Not to Hedge?

- D revenue forecast for 1998 = 1997 earning (NO growth)
  - 98 forecasted revenue in Germany is 645 million DM
  - if n_{DM} 口數 500 million bought
    - k_{DM}=0.62 USD$/DM; cost c_{DM}=$0.00137
  - 97 exchange rate 0.6513 USD$/DM. If DM is up by 10.84% in 1998

$$\text{Unhedged Revenue} = 645 \times 0.6513 \times \left(1 + \frac{10.84}{100}\right) = \$465.626 \text{ million.}$$

$$\begin{aligned}\text{Hedged Revenue} &= \text{Unhedged Revenue} \\ &\quad + \underline{500} \times \left[\max\left(0.62 - 0.6513 \times \left(1 + \frac{10.84}{100}\right), 0\right) - 0.0137\right] \\ &= \$465.626 - \$6.85 \\ &= \$458.776 \text{ million.}\end{aligned}$$

直接模擬各數字

獲利降低

  - DM up by 10.84%, but exchanged $USD is only up by 9.23%
    - from $USD 420 million in 97 to $USD 458.776 million in 98

© 2024 H. Chuang All rights reserved

To Hedge or Not to Hedge?

Bottom line

- **CEO Bill sees high exchange rate risk in 98 in west Europe**
  - **USD$756 million from Germany & UK in 97 (most of overseas earnings)**
  - **he wishes to minimize the chance earnings < USD$706 million in 1998**
    - **1-year put options for DM and BP in the market are listed below**

對沖

不想因匯損讓$ < 706 million

Strike Price for DM (in $)	*Cost (in $)*
0.66	0.085855
0.65	0.032191
0.64	0.020795
0.63	0.017001
0.62	0.013711
0.61	0.010851
0.60	0.008388
0.59	0.006291
0.55	0.001401

Strike Price for BP (in $)	*Cost (in $)*
1.30	0.137213
1.25	0.082645
1.20	0.045060
1.15	0.028348
1.10	0.016146
1.05	0.007860
1.00	0.003277
0.95	0.001134
0.90	0.000245

  - **suppose D intends to get n_{DM} 500 million & n_{BP} 500 million**
    - **which put options should be bought to meet Bill's objective?**
  - **suppose n_{DM} and n_{BP} can be bought in one of [100, 300, 500] million**
    - **which put options should be bought to meet Bill's objective?**

why 直接平均而不用看所有組合?

© 2024 H. Chuang All rights reserved