01/11

Probability ①

① Basic Concepts

nnett + Welsh obability n Introduction

1·1 Sample Space

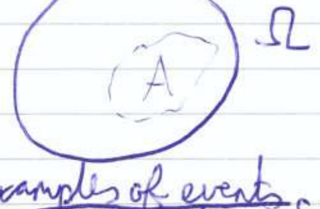
Experiment with uncertain outcomes.

1) A coin toss   3) A roulette wheel spin   5) Spin a pointer
2) A die throw   4) Pick for the national lottery

Ω, a set of all possible outcomes.

1) $\Omega = \{H, T\}$ 2) $\Omega = \{1, 2, 3, 4, 5, 6\}$
3) $\Omega = \{1, 2, \ldots, 36, 0 \ldots\}$ 4) $\Omega = \{\text{all 6-subsets of } \{1, 2, 3, \ldots, 49\}\}$
5) $\Omega = [0, 2\pi)$

The <u>sample space</u> is the set of all possible outcomes.



A subset of Ω is called an event.

<u>Examples of events</u>

1) heads $A = \{H\}$
2) prime $A = \{2, 3, 5\}$
3) even $A = \{2, 4, \ldots\}$
4) runs $A = \{(k, k+1, k+2, \ldots, k+5) : 1 \le k \le 44\}$
5) ≤ 6 o'clock $A = [0, \pi]$

Outcome: $\omega \in \Omega$ elementary event

If ω occurs in the experiment, we say "A occurs" iff $\omega \in A$

<u>sets</u> $\longleftrightarrow$	<u>events</u>
$A \cup B$	either A or B occurs
$A \cap B$	both A and B
$A \setminus B = A \cap \bar{B}$	A, but not B
$A \subseteq B$	If A, then B
$A = B$	equivalence
$A \cap B = \emptyset$	Cannot have both A and B, mutually exclusive.

1·2 Combinatorial Probability

Ω is finite, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$

Assume each ω_i is equally likely.

Let $P : \{\text{events}\} \to [0, 1]$ $P(A) = \frac{|A|}{n}$ for $A \subseteq \Omega$

Example: A hand of 13 cards is dealt from 52. What is the probability that it contains;

i) Exactly one ace   ii) Exactly one ace and two kings

i) $|A| = \binom{4}{1}\binom{48}{12}$ total # hands $= \binom{52}{13}$, answer $\frac{\binom{4}{1}\binom{48}{12}}{\binom{52}{13}}$

ii) $|A| = \binom{4}{1}\binom{4}{2}\binom{44}{10}$ answer $\frac{\binom{4}{1}\binom{4}{2}\binom{44}{10}}{\binom{52}{13}}$

22/01/11

Probability (2)

Example Table of random integers, of which we pick r.

$\Omega = \{0,1,2,3,4,\dots 9\}^r$

Assume each $\omega \in \Omega$ is equiprobable

What is the probability that:

a) No digit exceeds k? $\quad k \in \{0,1,\dots,9\}$

b) Greatest digit $= k$?

Solution a) $\frac{(k+1)^r}{10^r} = \alpha_k \qquad$ b) $\alpha_k - \alpha_{k-1}$

1.3 Permutations and Combinations

Perm: n objects, choose r to form an ordered subset. nPr, ${}^nP_r = \frac{n!}{(n-r)!}$

Combinations: An unordered subset. n objects, a set of r. nCr, ${}^nC_r$ (n Choose r) $= \frac{n!}{r!(n-r)!} = ($

Questions Urn contains b blue balls and r red. Remove them at random without replacement. Find the probability that the first red ball that it is the $(k+1)^{th}$ ball overall? "B^kR"

Solution Let R be the index of the first red ball. What is the probability that $R = k+1$?

$$p(R = k+1) = P(B^kR) = \frac{\text{\# such sequences, length } (b+r)}{\text{total \# of sequences}}$$

$$= \frac{\binom{b+r-(k+1)}{r-1}}{\binom{b+r}{r}}$$

Ménages Problem M/W couples seated randomly at a circular table, alternating MWMW.... Find p(noone is seated beside their partner)

$$= \frac{1}{n!}\sum_{k=0}^{n} (-1)^k \frac{2n}{2n-k}\binom{2n-k}{k}(n-k)!$$

Example n keys in pocket, 1 lock, pick keys at random until success.

a) With replacement:

$$p(\text{succeed on } r^{th} \text{ attempt}) = \frac{(n-1)^r \times 1}{n^r} = \frac{1}{n}\left(1-\frac{1}{n}\right)^r \simeq \frac{1}{n}e^{-\frac{r}{n}}$$

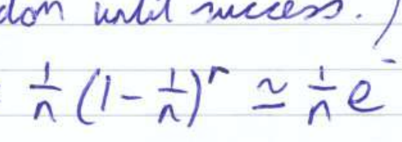
for large n and r

b) Without replacement $p(\text{succeed on } r^{th} \text{ attempt}) = \frac{(n-1)\dots(n-r+1)\times 1}{n(n-1)\dots(n-r+1)} = \frac{1}{n}$

Two Facts ~~Stirling's Formula~~ Stirling's Formula $\quad \frac{n!}{n^n e^{-n}\sqrt{2\pi n}} \to 1$ as $n \to \infty$

How fast does $n!$ grow as $n \to \infty$?

Weak Version $\frac{\log n!}{n \log n} \to 1$ as $n \to \infty$



$$\log n! = \sum_{k=1}^{n} \log k \qquad \int_1^n \log x\,dx \le \sum_1^n \log k \le \int_1^{n+1} \log x\,dx$$

$$[x\log x - x]_1^n \le \log n! \le [x \log x]_1^{n+1}$$

$$\frac{n\log n - n + 1}{n \log n} \leq \frac{\log n!}{n \log n} \leq \frac{(n+1)\log(n+1) - (n+1) + 1}{n \log n}$$

tends to $\quad 1 \leq \frac{\log n!}{n \log n} \leq 1$

<u>Binomial Expansion</u> $(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \dots x^n$

$= \sum_{k=0}^{\infty} x^k \binom{n}{k}$ assuming: $x \in \mathbb{R}$, $n \in \{1,2,3,\dots\}$

$(1+x)^\alpha = f(x)$ for $\alpha \in \mathbb{R}$? Expand $f(x)$ as a power series in x.
If the Taylor Expansion converges correctly then

$f(x) = f(0) + f'(0)x + f''(0)\frac{x^2}{2} + \dots$

$f'(0) = \alpha(1+0)^{\alpha-1}$

$f^{(k)}(0) = \alpha(\alpha-1)\dots(\alpha-k+1) \qquad f(x) = \sum_{k=0}^{\infty} x^k \binom{\alpha}{k}$

$\binom{\alpha}{k} = \frac{\alpha(\alpha-1)\dots(\alpha-k+1)}{k!}$ Fine providing $|x| < 1$

Example. Toss a fair coin $2n$ times. $H = \#$ heads $\quad \Omega = \{H, T\}^n$

$p(H=n) = \frac{\binom{2n}{n}}{2^{2n}}$

(Stirling) $\approx \frac{(2n)^{2n} e^{-2n}\sqrt{4\pi n}}{2^{2n}\left[n^n e^{-n}\sqrt{2\pi n}\right]^2} \approx \frac{1}{\sqrt{\pi n}}, \quad \frac{n!}{n^n e^{-n}\sqrt{2\pi n}} \to 1$

<u>Exercise</u> $p(H=n)$ when $3n$ coins are tossed.

$$p(H=n) = \frac{\binom{3n}{n}}{2^{3n}}$$

$$= \frac{(3n)!}{n!(2n)!}\frac{1}{2^{3n}} \approx \frac{(3n)^{3n}e^{-3n}\sqrt{6\pi n}}{n^n e^{-n}\sqrt{2\pi n}\,(2n)^{2n}e^{-2n}\sqrt{4\pi n}}\frac{1}{2^{3n}}$$

$\sqrt{\frac{6}{8}} = \frac{\sqrt{3}\times\sqrt{2}}{2\sqrt{2}} = \frac{\sqrt{3}}{2}$

$$= \frac{(3n)^{3n}}{n^n(2n)^{2n}} \times \frac{\sqrt{6\pi}}{\sqrt{2\pi}\sqrt{4\pi}} \times \frac{\sqrt{n}}{n} \times \frac{1}{2^{3n}}$$

$$= \frac{3^{3n}}{2^{2n}}\frac{\cancel{n^{3n}}}{\cancel{n^n n^{2n}}} \times \sqrt{\frac{6}{8\pi}} \times \frac{1}{\sqrt{n}} \times \frac{1}{2^{3n}}$$

$$= \frac{3^{3n}}{2^{5n}} \times \sqrt{\frac{6}{8}} \times \frac{1}{\sqrt{\pi n}} = \frac{3^{3n}}{2^{5n}}\sqrt{\frac{3}{4\pi n}}$$

25/01/11

Probability ③

2 Probability Space

2.1 (i) Sample Space Ω
(ii) {Collection of events} = event space
(iii) Probability function

The power set of Ω is the set of all subsets of Ω. Denoted 2^Ω or $\{0,1\}^\Omega$ = {vector indexed by Ω with elements 0, 1}

In general, the event space $\subseteq 2^\Omega$ but not equal, in general

Reason If Ω is uncountable, 2^Ω is too big.

Reasonable conditions on events

If A, B are events, then so are $A \cup B$, $A \cap B$, $\Omega \setminus A$

Definition An event space (or σ-field or σ-algebra) is a collection $\mathcal{F}$ of subsets of the sample space Ω such that:

a) $\emptyset \in \mathcal{F}$

b) If $A_1, A_2, \ldots \in \mathcal{F}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

c) If $A \in \mathcal{F}$ then $\bar{A} = \Omega \setminus A \in \mathcal{F}$

Notes (i) $\Omega = \Omega \setminus \emptyset \in \mathcal{F}$, by (a), (c) ($\mathcal{F}$ is an event space)

(ii) Finite unions of events lie in $\mathcal{F}$ ($A_i = \emptyset$ for $i \geq n+1$)

(iii) $\bigcap_i A_i = \overline{\bigcup_i \bar{A_i}}$ $\therefore$ $\mathcal{F}$ is closed under countable intersectio[ns] and also finite intersections ($A_i = \Omega$ for $i \geq n+1$)

(iv) $A \setminus B = A \cap \bar{B} \in \mathcal{F}$ if $A, B \in \mathcal{F}$

Similarly $A \triangle B = (A \setminus B) \cup (B \setminus A)$

(v) (a) is equivalent to requiring $\mathcal{F} \neq \emptyset$

(Since if $A \in \mathcal{F}$, then $\bar{A} \in \mathcal{F}$ $\therefore$ $A \cap \bar{A} = \emptyset \in \mathcal{F}$)

Definition Let Ω be a set and $\mathcal{F}$ an event space of Ω. The pair $(\Omega, \mathcal{F})$ is a 'measurable pair'. A <u>probability measure</u> is a function, $P: \mathcal{F} \to \mathbb{R}$ such that

a) $0 \leq P(A) \leq 1$ for $A \in \mathcal{F}$

b) $P(\Omega) = 1$, $P(\emptyset) = 0$

c) If $A_1, A_2, \ldots \in \mathcal{F}$ are ~~pairwise~~ disjoint.

Then $P\left(\bigcup_1^{\infty} A_i\right) = \sum_1^{\infty} P(A_i)$ 'countable additivity'

<u>(Non-examinable) The problem with event spaces</u>

<u>Theorem</u> Assuming the Continuum Hypothesis, there is no measure μ on the set of all subsets of $I = [0,1]$ with $\mu(I) = 1$, and $\mu(\{x\}) = 0 \ \forall x \in I$

<u>Notes</u> (i) P is finitely additive. $P(A_1 \cup A_2 \cup \ldots \cup A_n \cup \emptyset \cup \emptyset \ldots) = \sum_1^{n} P(A_i) + ($

(ii) $P(\emptyset) = 0$ follows by $\Omega \cup \emptyset = \Omega$ $\Rightarrow P(\emptyset) = 0$

$\Omega \cup \emptyset \cup \emptyset \cup \emptyset \cup \ldots = \Omega$ $\therefore P(\Omega) + \ldots = P(\Omega)$

<u>Definition</u> A probability space is $(\Omega, \mathcal{F}, P)$ with

a) Ω is a set

Experiment $\longleftrightarrow$ Probability Space

b) $\mathcal{F}$ is an event space in Ω

c) P is a probability ~~space~~ measure on $(\Omega, \mathcal{F})$

<u>Example</u> (i) <u>Bernoulli Distribution</u> $\Omega = \{0,1\}$, $\mathcal{F} = 2^{\Omega}$

$0 \leq p \leq 1$

A	$\emptyset$	$\{0\}$	$\{1\}$	Ω
$P(A)$	0	$1-p$	p	1

"Toss of a possibly biased coin"

25/01/11

Probability ③

(ii) <u>Combinatorial Probability</u>

$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$, $\mathcal{F} = 2^{\Omega}$

$A \in \mathcal{F}$, $P(A) = \frac{|A|}{n}$

(iii) <u>Poisson Distribution</u> $\mathcal{F} = 2^{\Omega}$

$\Omega = \{\omega_1, \omega_2, \dots\}$ $(p_i : i \geq 1)$ a real sequence with $p_i \geq 0$ $\sum_i p_i = 1$

$P(A) = \sum_{i = \omega_i \in A} p_i$

<u>E.g.</u> $p_i = \frac{c\lambda^i}{i!}$, $c = e^{-\lambda}$ Poisson Probabilities, parameter λ

<u>Theorem:</u> Let $(\Omega, \mathcal{F}, P)$ be a probability space.

If $A, B \in \mathcal{F}$ then a) $P(A) + P(\bar{A}) = 1$

b) If $B \subseteq A$, $p(A) = p(B) + p(A \setminus B)$, $\geq p(B)$

c) $P(A \cup B) = p(A) + p(B) - p(A \cap B)$

<u>Proof c)</u> $A \cup B = A \cup (B \setminus A)$

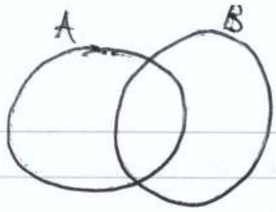
$B \setminus A = B \setminus (A \cap B)$

since $A \cap B \subseteq B$

$p(A \cup B) = p(A) + p(B \setminus A \cap B) = p(A) + (p(B) - p(A \cap B))$

27/01/11

Probability ④



Venn

$p(A \cup B) = p(A) + p(B) - p(A \cap B)$

Theorem: Inclusion - Exclusion Principle

$$p(A_1 \cup A_2 \cup A_3 \cup \dots \cup A_n) = \sum_i p(A_i) - \sum_{i<j} p(A_i \cap A_j) + \sum_{i<j<k} p(A_i \cap A_j \cap A_k) - \dots + (-1)^{n+1} p(A_1 \cap \dots \cap A_n)$$

Proof - By Induction on n

True for $n = 2$

We assume truth for $n = k$

$$p(\underbrace{A_1 \cup \dots \cup A_k}_{} \cup A_{k+1}) = p(A_1 \cup \dots \cup A_k) + p(A_{k+1}) - p\left[(A_1 \cup A_2 \cup A_3 \cup \dots \cup A_k) \cap A_{k+1}\right]$$

Expand using the induction hypothesis, and collect term. $\square$

Boole's Inequality (Sub-additivity of probability)

$$p\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} p(A_i)$$

Proof: Trivial for $n = 1$, and use a proof by induction $\square$
(NOTE! Also true for a countable union if $n = \infty$, but can't use induction)

Bonferroni Inequality

$$p\left(\bigcap_{i=1}^{n} A_i\right) \geq 1 - \sum_{i=1}^{n} p(\bar{A}_i) = 1 - n + \sum_i p(A_i)$$

Proof: $p(\bigcap A_i) = 1 - p(\bigcup \bar{A}_i)$
$\geq 1 - \sum_i p(\bar{A}_i)$ by Boole

neo-Bonferroni Inequality

$$p(A_1 \cup A_2 \cup \dots \cup A_n) \gtrless \sum_i p(A_i) - \sum_{i<j} p(A_i \cap A_j) \dots \dots + (-1)^{r+1} \sum_{i_1 < \dots < i_r} p(A_{i_1} \cap \dots \cap A_{i_r})$$

with $\geq$ if r is even
$\leq$ if r is odd.

Example (Derangements)

After dinner, the porter hands the hats to guests at random. There are n hats and n guests. What is the probability that nobody recieves the correct hat?

Solution $\Omega = \{$permutations of $1, 2, \ldots, n\}$

where the permutation $(i_1, i_2, \ldots, i_n)$ means that guest j recieves the hat of guest i_j. $(\omega_1, \ldots, \omega_n)$

Let $A_i = \{\omega \in \Omega : \omega_i = i\} = \{i^{th}$ person recieves the correct hat$\}$

We want $p(\bigcap_i \bar{A}_i) = 1 - p(\bigcup_i A_i)$

$$p(A_{i_1} \cap \ldots \cap A_{i_r}) = \frac{(n-r)!}{n!}$$

$$\sum_{i_1 < \ldots < i_r} \quad = \frac{(n-r)!}{n!}\binom{n}{r} = \frac{1}{r!}$$

$$p(\cup A_i) = 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n+1}\frac{1}{n!} \qquad = 1 - p(\cap \bar{A}_i)$$

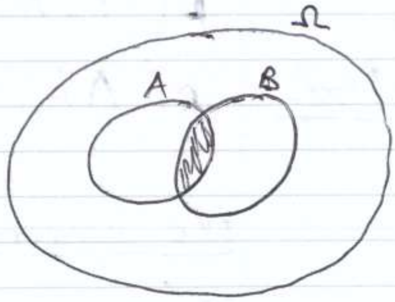
As $n \to \infty$, $= 1 - \frac{1}{e}$

Let $p_m(n) = p($exactly m people recieve the correct hat$)$

$$= \binom{n}{m}\frac{p_0(n-m)\,(n-m)!}{n!} = \frac{p_0(n-m)}{m!} = \frac{e^{-1}}{m!} \text{ as } n \to \infty$$

the Poisson distribution with parameter $\lambda = 1$.

2.2 Conditional Probability



$[(\Omega, \mathcal{F}, P)$ is a probability space $]$

Event 'A', probability $p(A)$

New information: A certain event 'B' has occured.

What now is the probability of A?

The definition is $c\,P(A \cap B)$ for some c

27/01/11

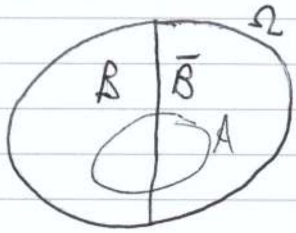
Probability ④

The probability of Ω given 'B' must be 1

$$cP(\Omega \cap B) = 1, \text{ but } p(\Omega \cap B) = p(B) \Rightarrow c = \frac{1}{p(B)}$$

<u>Definition</u> The "conditional probability of A given B" denoted $p(A \mid B)$ is

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)}$$



Note: Well defined if and only if $p(B) \neq 0$

<u>Theorem</u> Let B satisfy $0 < p(B) < 1$

Then $p(A) = p(A|B)\,p(B) + p(A|\bar{B})\,p(\bar{B})$

<u>Proof</u> $A = (A \cap B) \cup (A \cap \bar{B})$ a disjoint union.

$$p(A) = p(A \cap \bar{B}) + p(A \cap B)$$
$$= p(A|\bar{B})\,p(\bar{B}) + p(A|B)\,p(B)$$

$$p(B_i | A) = \frac{p(A \cap B_i)}{p(A)}$$

$$= \frac{p(A \cap B_i)}{\sum_j p(A|B_j)p(B_j)} \times \frac{p(B_i)}{p(B_i)}$$

$$= \frac{p(A|B_i)\,p(B_i)}{\sum_j p(A|B_j)p(B_j)}$$

01/02/11

Probability 5

$$p(A \mid B) = \frac{p(A \cap B)}{p(B)} \qquad (\text{for } p(B) > 0)$$

More generally:

Theorem If $B_1, B_2, \ldots$ is a partition of Ω with $p(B_i) > 0 \ \forall i$ then

$$p(A) = \sum_i p(A \mid B_i)\, p(B_i) \qquad (\text{the Law of Total Probability})$$

"2 stage experiment"

Example A fair coin is tossed once. If heads, 1 die is tossed; if tails 2 dice are tossed. What is the probability that the sum of any die values is four?

Solution $[\Omega = \{0,1\} \times \{1,2,\ldots,6\}^2 \text{ for example}]$

Let $A = \{\text{total is } 4\}$, $B = \{\text{coin shows heads}\}$

$$p(A) = p(A \mid B)\, p(B) + p(A \mid \bar{B})\, p(\bar{B}) = \frac{1}{6} \times \frac{1}{2} + \frac{3}{36} \times \frac{1}{2} = \frac{1}{8}$$

Properties of Conditional Probability

a) $p(A \cap B) = p(A \mid B)\, p(B)$

b) $p(A \mid B) = \dfrac{p(B \mid A)\, p(A)}{p(B)}$

c) $p(A \cap B \cap C) = p(A \mid B \cap C)\, p(B \mid C)\, p(C)$

d) $p(A \mid B \cap C) = \dfrac{p(A \cap B \mid C)}{p(B \mid C)}$

Theorem (Bayes's Formula)

Let $B_1, B_2, \ldots$ be a partition of Ω, $p(B_i) > 0 \ \forall i$.

Then *

$$p(B_i \mid A) = \frac{p(A \mid B_i)\, p(B_i)}{\sum_j p(A \mid B_j)\, p(B_j)}$$

Map: $\underset{\text{prior distribution}}{p(B_i)} \rightarrow \underset{\text{posterior distribution}}{p(B_i \mid A)}$

Example : False Positives

There is a rare disease with incidence in the population 1 in 100,000.
The test is fairly reliable.
If you have the disease, the test ~~says~~ is positive with probability 0·95.
If not, the test is positive with probability 0·005.

If the test is positive, what is the probability the patient has the disease?

Solution $D = \{\text{has disease}\}$ $T = \{\text{Test is positive}\}$

$$p(D|T) = \frac{p(T|D)p(D)}{p(T|D)p(D) + p(T|\bar{D})p(\bar{D})}$$

$$= \frac{\frac{1}{100\,000} \times 0{\cdot}95}{\frac{1}{100\,000} \times 0{\cdot}95 + \frac{99999}{100\,000} \times 0{\cdot}005} \simeq 0{\cdot}002$$

Test is essentially useless

Principle of Uniformity:

In the absence of information, take the prior to be uniform.

e.g. If there are 2 possibilities, take $\frac{1}{2}$ and $\frac{1}{2}$. ⇒ "Bayes Postulate"
1939 Harold Jeffreys

Simpson's Paradox (British Medical Journal 1986, Kidney Stone Removal)

Before 1980, Open Surgery was performed and afterwards, a new operation PN. In 1972 - 1980 [OS], $\frac{273}{350} \simeq 78\%$ were successful.
In 1980 - 1985 [PN], $\frac{289}{350} \simeq 83\%$ were successful.

percutaneous nephrolithotomy

We deduce that PN is better than OS.

	Small (<2cm)	Large (>2cm)
OS	93%, $\frac{81}{87}$	73% $\frac{192}{263}$
PN	87% $\frac{234}{270}$	69% $\frac{55}{80}$

We then see that OS is better in BOTH subcases.

02/01/11

Probability ⑤

The following are not inconsistent:

$p(R|A) > p(R|B)$

$p(R|A \cap S) < p(R|B \cap S)$

$p(R|A \cap L) < p(R|B \cap L)$

$R = \{\text{success}\}$

A : PN
B : OS
S : small
L : large

03/02/11

Probability (6)

2·3 Independence

Intuition: if $p(A) = p(A \mid B)$ then B is of little relevance to A.

Definition

Events are independent if $p(A \cap B) = p(A)\,p(B)$

More generally, a family (I as the index set, could be countable, uncountable

$(A_i : i \in I)$ is independent if

$p\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} p(A_i)$ for all finite subsets $J \subseteq I$,

and pairwise independent if $p(A_i \cap A_j) = p(A_i)\,p(A_j)\ \forall\, i,j \in I$, $i \neq j$

Independence $\Rightarrow$ Pairwise Independence. The converse is false.

Example

$\Omega = \{1,2,3,4\} \qquad p(A) = |A|/4$

$A_1 = \{1,2\},\ A_2 = \{2,3\},\ A_3 = \{1,3\}$

'Independence' $\longleftrightarrow$ 'repeated trials'

Example Two dice are thrown, equiprobable outcomes

A: attribute of the number on the first die.

B: attribute of the number on the 2nd die

$p(A \cap B) = \frac{\#\text{ outcomes: A on 1st, B on 2nd}}{36} = \frac{\#\text{ outcomes with A on 1}}{6} \times \frac{\#\text{ outcomes with B on 2}}{6}$

More general: Product probability space

$(\Omega_1, \mathcal{F}_1, P_1), (\Omega_2, \mathcal{F}_2, P_2)$

$\Omega_1 = \{\alpha_1, \alpha_2, \ldots\} \quad P_1(\alpha_i) = p_i$

$\Omega_2 = \{\beta_1, \beta_2, \ldots\} \quad P_2(\beta_j) = q_j$

Let $\Omega = \Omega_1 \times \Omega_2 = \{(\alpha_i, \beta_j) : i, j \geq 1\}$

$\mathcal{F}$ = something suitable

$p[(\alpha_i, \beta_j)] = p_i q_j \quad , \quad i, j \geq 1$

Then $A_1 \subseteq \Omega_1$, $A_2 \subseteq \Omega_2$

$$p(A_1 \times A_2) = \sum_{\substack{i \in A_1 \\ j \in A_2}} p_i q_j = \sum_{i \in A_1} p_i \sum_{j \in A_2} q_j = \underbrace{p_1(A_1)}_{p(A_1 \times \Omega_2)} \underbrace{p_2(A_2)}_{p(\Omega_1 \times A_2)}$$

Language

Flips of a coin, Throws of a die } Interpreted to imply independence between outcomes.

Example

n flips of a coin that shows heads with probability p each time.

Let S_n be the number of heads. Find $p(S_n = k)$.

Solution 1

$$p(S_n = k) = \binom{n}{k} p(H^k T^{n-k}) = \binom{n}{k} p^k (1-p)^{n-k}$$

The Binomial Distribution.

Solution 2

Let X be the outcome of the first flip

$$p(S_n = k) = p(S_n = k \mid X = H)\, p(X = H) + p(S_n = k \mid X = T)\, p(X = T)$$

$$p(S_n = k) = p(S_{n-1} = k-1)\, p + p(S_{n-1} = k)(1-p)$$

Valid for $k \geq 0$

Let $p_n(k) = p(S_n = k)$

$$p_n(k) = p\, p_{n-1}(k-1) + (1-p)\, p_{n-1}(k)$$

A discrete recurrence relation.

03/02/11

Probability ⑥

$$p_n(k) = p\left[p\,p_{n-2}(k-2) + (1-p)\,p_{n-2}(k-1)\right] + (1-p)\left[p\,p_{n-2}(k-1) + (1-p)\,p_{n-2}(k)\right]$$

$$= \sum_{r=0}^{s} p^r (1-p)^{s-r} \binom{s}{r} p_{n-s}(k-r) \qquad s \geq 0$$

(Prove by induction)

$$= \sum_{r=0}^{n} p^r (1-p)^{n-r} \binom{n}{r} p_0(k-r) \qquad p_0(k-r) = \delta_{kr}$$

Kronecker Delta

$$= p^k (1-p)^{n-k} \binom{n}{k}$$

<u>Geometric Distribution</u>

Same coin is tossed repeatedly until the first head appears. Let R be the number of flips required.

$$p(R=r) = (1-p)^{r-1} p \quad , \quad r = 1, 2, 3, \ldots$$

Take care: Can see the geometric distribution as

$$p_r = (1-p)^r p \quad , \quad r = 0, 1, 2, \ldots$$

<u>Example 2</u>

<u>Random walks</u>

$$0 \quad 1 \quad 2 \quad \xleftarrow{q}\ k\ \xrightarrow{p} \quad N$$

$q = 1-p$

Walk on $\{0, 1, \ldots, N\}$

Start at k. At each step, move one step right with probability p, or left with probability q. Different steps are independent. Assume there are absorbing barriers at 0 and N. This can be compared to a Gambler who plays a game until either he reaches

0 – he leaves bankrupt $\qquad N$ – leaves and buys a large car.

What is the probability of ultimate bankruptcy?

$= p(\text{we reach } 0 \text{ before we reach } N)$

Let $A = \{\text{absorbed at } 0\}$

$B = \{\text{1st step is to the right}\}$

$p(A) = p(A|B)\,p(B) + p(A|\bar{B})\,p(\bar{B})$

Let $p_k = p(A \mid \text{start at } k)$

$\Rightarrow p_k = p_{k+1}\,p + p_{k-1}\,q \;, \text{ for } 0 < k < N$

07/02/11

Probability ⑦

Random walk:

$p_k = p\, p_{k+1} + q\, p_{k-1}$

$\Rightarrow \quad p(p_{k+1} - p_k) = q(p_k - p_{k-1})$

Boundary conditions $p_0 = 1$, $p_N = 0$. Try $p_k = \theta^k$

$\Rightarrow p\theta^2 - \theta + q = 0 \qquad (p\theta - q)(\theta - 1) = 0$

$\theta = \frac{q}{p}, 1$ If $q \neq p$, roots are distinct

$(p \neq \frac{1}{2})$ General solution $p_k = A\left(\frac{q}{p}\right)^k + B \cdot 1^k$

$p_k = \dfrac{\left(\frac{q}{p}\right)^k - \left(\frac{q}{p}\right)^N}{1 - \left(\frac{q}{p}\right)^N}$, $p_k = A + Bk$

If $q = p = \frac{1}{2}$ (symmetric) $p_k = 1 - \frac{k}{N}$

3 Discrete Random Variables $(\Omega, \mathcal{F}, P)$

Concept definition A random variable is a function $X : \Omega \to \mathbb{R}$

Example 1 $X = \#$ heads after two coin tosses

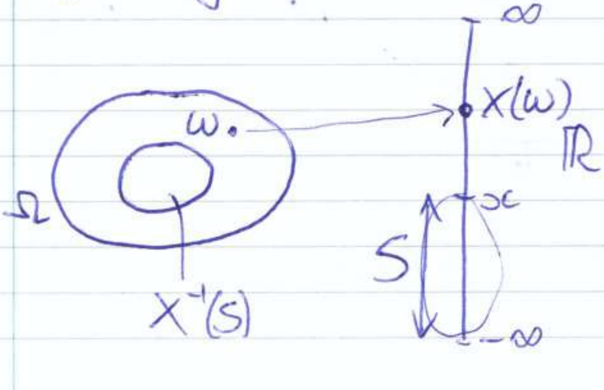
$\Omega = \{0, 1\}^2 \qquad \omega = \{\omega_1, \omega_2\} \in \Omega$

$X(\omega) = \omega_1 + \omega_2$

Example 2 Throw 3 dice. X is the largest number shown.

Definition The distribution function of a random variable X is defined by $F : \mathbb{R} \to [0, 1]$, $F(x) = P(\{\omega \in \Omega : X(\omega) \leq x\})$

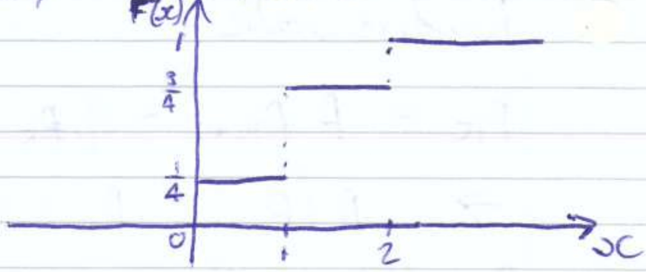
$= p(X \leq x)$



Note: If X is countable $F_x(x) = \sum_{u \leq x} f_x(u)$

<u>Example 1</u> Fair coin, two tosses, $X = \#$ heads

X takes values in $\{0, 1, 2\}$



<u>Definition:</u> The mass function of the random variable X is the function $f : \mathbb{R} \to [0, 1]$ given by $f(x) = p(X = x)$

<u>Definition</u> The ~~ms~~ random variable X is discrete if there exists a countable set $S = \{x_1, x_2, \ldots\}$ such that $p(X \in S) = 1$

If X is discrete we usually work with the mass function.

<u>1. Bernoulli Distribution</u> (coin toss)

$p_0 = 1 - p,\ p_1 = p$, where $p \in [0, 1]$ $\quad f(0) = 1 - p,\ f(1) = p$

<u>2. Binomial Distribution</u> n, p. $\mathrm{bin}(n, p)$

$f(k) = \binom{n}{k} p^k (1-p)^{n-k},\ k = 0, 1, \ldots, n$

<u>3. Poisson Distribution</u> $Po(\lambda)$ $\quad f(k) = \frac{\lambda^k e^{-\lambda}}{k!},\ k = 0, 1, \ldots$

<u>Relationship between Binomial and Poisson distributions</u>

A book: One page has $n = 10^5$ characters, each of which is misprinted with probability $p = 10^{-5}$ independently of the others.

The distribution of N, the total number of misprints is $\mathrm{bin}(n, p)$.

As $n \to \infty$, $p \downarrow 0$ such that $pn \to \lambda$ as $n \to \infty$:

$$p(n = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\ldots(n-k+1)}{n^k} \frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k} \frac{\lambda^k}{k!}$$

As $n \to \infty$ $\quad p(n = k) \to \frac{e^{-\lambda}\lambda^k}{k!}$

So $\mathrm{bin}(n, \frac{\lambda}{n}) \to Po(\lambda)$ as $n \to \infty$

08/02/11

Probability (8)

Example 4 Geometric distribution

$f(k) = A\beta^k, \quad k \geq 1, 0<\beta<1$

$\sum_k f(k) = \sum_1^\infty A\beta^k = \frac{A\beta}{1-\beta} = 1, \quad A = \frac{1-\beta}{\beta}$

$f(k) = (1-\beta)\beta^{k-1}$

Example 5 Negative binomial

Toss a coin, $p(\text{heads}) = p$, until exactly r heads are tossed.

What is the probability that this requires k tosses $(= p_k)$.

$p_k = p(\underset{1 \quad 2 \; \cdots \cdots \; k}{r-1 \text{ heads}, k-r \text{ tails} \quad H}) = \binom{k-1}{r-1}(1-p)^{k-r}p^{r-1}p$

$= p^r(1-p)^{k-r}\binom{k-1}{r-1}, \quad k = r, r+1, \ldots$

$L = k-r$
$k = L+r$

$\sum_k p_k = p^r \sum_{k \geq r}(1-p)^{k-r}\binom{k-1}{r-1}$

$= p^r \sum_{L=0}^\infty (1-p)^L \binom{L+r-1}{L} \quad = \frac{(L+r-1)(L+r-2)\cdots(L+1)}{\cancel{(r-1)!}\, L!}$

$= p^r \sum_{L=0}^\infty (1-p)^L \binom{-r}{L}(-1)^L \quad = \frac{(-1)^L}{L!}(-r)\cdots(-r-L+1) = (-1)^L\binom{-r}{L}$

$= p^r[1-(1-p)]^{-r} = 1$

$\mathbb{R} \quad \mathbb{P} \quad \mathbb{E}$

3.2 Expectation (of Discrete Random Variables)

$(\Omega, \mathcal{F}, P)$, discrete random variable X

Definition The expectation (or mean value) of X is

$E(X) = \sum_{\{x : P(X=x)>0\}} xP(X=x) = \sum_x x f_X(x)$ whenever this sum converges absolutely.

Composition of Functions Composition creates a new random variable

$\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}$

$Y = g(X)$

$[Y(\omega) = g(X(\omega)), \; \omega \in \Omega]$

1

Theorem (Law of the unconscious statistician)

$E(g(X)) = \sum_x g(x) f_X(x)$

Proof

$Y = g(X)$, $E(Y) = \sum_y y\, p(Y=y)$

$= \sum_y y \left[\sum_{x: g(x)=y} p(X=x)\right] = \sum_x \cancel{\sum_y}\, g(x)\, p(X=x)$

Properties of Expectation

1) If $X \geq 0$, then $E(X) \geq 0$ ['almost surely', $p(X \geq 0) = 1$]

2) If $X \geq 0$, $E(X) = 0$, then $p(X=0) = 1$

Proof If $X \geq 0$, $E(X) = \sum_{x \geq 0} x\, P(X=x) = 0$

$\Rightarrow x\, P(X=x) = 0$ for all $x > 0$.

3) $E(\alpha X + \beta) = \sum_x (\alpha x + \beta) p(X=x) = \alpha \sum_x x\, p(X=x) + \beta \sum_x p(X=$

$= \alpha E(X) + \beta$ $\qquad \alpha, \beta \in \mathbb{R}$

4) $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$

$E(\alpha X + \beta Y) = \sum_{x,y} (\alpha x + \beta y)\, p(X=x, Y=y)$

$= \alpha \sum_{x,y} x\, p(X=x, Y=y) + \beta \sum \ldots$

$= \alpha \sum_x x\, p(X=x) + \ldots$

$= \alpha E(X) + \beta E(Y)$

"Expectation is a linear operator" $\quad E(X)$: measure of distribution's centr

Variance: a measure of dispersion.

Definition The Variance of X is:

$Var(X) = E[(X - E(X))^2]$

and the standard deviation $\sigma(X) = \sqrt{Var(X)}$

08/02/11

Probability (8)

a) Variance is non linear

$Var(\alpha X + \beta) = \alpha^2 Var(X)$

and hence $\sigma(\alpha x + \beta) = |\alpha| \sigma(x)$

b) $Var(X) = E[(X - EX)^2] = E(X^2 - 2X(EX) + (EX)^2)$

$= E(X^2) - 2E(X)E(X) + [E(X)]^2$

$= E(X^2) - (EX)^2$

Warning: Be careful with parentheses, e.g. what does EX^2 mean?

$(EX)^2$ or $E(X^2)$?

<u>Definition</u> The k^{th} <u>moment</u> of X is $m_k = E(X^k)$, $k = \mathbb{N}_0$

<u>Note</u> a) $Var(X) = m_2 - (m_1)^2$

b) $Var(X) \geq 0$

c) $Var(X) = 0$ if and only if $P(X = c) = 1$ for some $c \in \mathbb{R}$

10/02/11

Probability (9)

$M_k = E(X^k)$ moments

Example 1 Bernoulli's Distribution

$p(X=0) = q$, $p(X=1) = p$, $p+q=1$

$E(x) = 0q + 1 \times p = p$

$Var(x) = E(X^2) - (EX)^2 = p - p^2 = pq$

Probability Space $(\Omega, \mathcal{F}, p)$ event $A \subseteq \Omega$

Indicator function of A is the random variable $I_A : \Omega \to \{0,1\}$

by $I_A(\omega) = \begin{cases} 0 & ; \omega \notin A \\ 1 & ; \omega \in A \end{cases}$

I_A is a Bernoulli random variable. $p(I_A = 0) = p(\bar{A})$

$p(I_A = 1) = p(A)$. $E(I_A) = p(A)$

Example 2 Binomial Distribution

X is $bin(n,p)$

$E(X) = \sum_k kp(X=k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np$

or X is the sum of n $Bern(p)$ variables each with mean p, so

$E(X) = np$, and in fact $Var(X) = npq$

Example 3 Poisson Distribution, $Po(\lambda)$

$E(X) = \sum_k kp(X=k) = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda$

Example 4 Geometric, parameter p

$p(X=r) = (1-p)^{r-1} p \quad r \geq 1$

$E(x) = \sum_{r=1}^{\infty} r(1-p)^{r-1} p$

$\sum_{r=0}^{\infty} x^r = \frac{1}{1-x}$ if $|x|<1$, $\sum_{r=0}^{\infty} r x^{r-1} = \frac{1}{(1-x)^2}$, $|x|<1$

$E(x) = p \cdot \frac{1}{p^2} = \frac{1}{p}$

$f\left(\frac{x+y}{2}\right) \leq \frac{f(x)+f(y)}{2}$

and $\mathrm{Var}(X) = \frac{q}{p^2}$ $(q = 1-p)$

3.3 Probability Generating Functions

Definition Random variable X taking values in $\{0, 1, 2, \ldots\}$

The probability generating function (PGF) of X is the function

$G: S \to \mathbb{R}$, $G(s) = \sum_{k=0}^{\infty} s^k p(X=k) = E(s^X)$

whenever this sum converges absolutely, and as big an S as possible. Note: This sum converges absolutely whenever $|s| < 1$. $(-1 < s \leq 1)$. Sometimes we write G_X for G.

$G_X(0) = p(X=0)$, $G_X(1) = 1$

Theorem

The distribution of X is uniquely determined by its PGF G.

Proof $p_k = p(X=k)$

$G(s) = p_0 + sp_1 + s^2 p_2 + \ldots$

Converges on $(-1, 1]$. $s = 0 \Rightarrow G(0) = p_0$

$G'(0) = p_1 \quad \ldots \quad G^{(k)}(0) = k!\, p_k$

Why?

1) An elegant method for handling sums of random variables.

2) A good method for calculating moments

$G(s) = \sum_k s^k p(X=k) \qquad s \in (-1, 1]$

$G'(s) = \sum_k k s^{k-1} p(X=k)$

10/02/11 Probability ⑨

<u>Theorem</u> $E(X) = G_X'(1)$ ↙ Ok for IA probability

Non rigorous: $G(s) = E(s^X)$, $G'(s) = E(Xs^{X-1}) \Rightarrow G'(1) = E(X)$

Optional (<u>Problem</u> $s=1$ <u>might</u> be on the edge of the domain of convergence of G. We need Abel's Lemma.

<u>Further such results</u>

$E(X) = G'(1)$ $\quad G^{(k)}(1) = E[X(X-1)(X-2)\dots(X-k+1)]$ $\quad k=1,2,\dots$

$G''(1) = E[X(X-1)] = E(X^2 - X) = E(X^2) - E(X)$

$\therefore \mathrm{Var}(X) = G''(1) + G'(1) - [G'(1)]^2$

<u>Examples</u> $\mathrm{Bern}(p)$

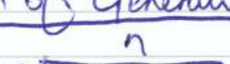
$G(s) = qs^0 + ps^1 = q + ps$

$\mathrm{Bin}(n,p)$ $\quad G(s) = (q+ps)^n$

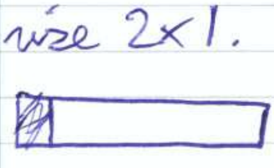
$\mathrm{Po}(\lambda)$ $\quad G(s) = e^{\lambda(s-1)}$ $\quad \sum_{k=0}^{\infty} s^k \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(s\lambda)^k}{k!}$

$\mathrm{Geom}(p)$ $\quad G(s) = \frac{ps}{1-qs}$

<u>Application of Generating functions to tiling a bathroom</u>

$n \times 2$ 

Let f_n be the number of possible ways to tile this with tiles of size 2×1. $\quad f_n = f_{n-1} + f_{n-2}$



$f_0 = 1$, $f_1 = 1$

Let $F(s) = \sum_{n=0}^{\infty} f_n s^n$

$$\sum_{n\geq 2} s^n f_n = \sum_{n\geq 2} f_{n-1}s^n + \sum_{n\geq 2} s^n f_{n-2}$$

$$F(s) - f_0 - f_1 s = s[F(s) - f_0] + s^2 F(s)$$

$$F(s) = \frac{f_0(1-s) + f_1 s}{1-s-s^2} = \frac{1}{1-s-s^2}$$

12/02/11

Probability (10)

Number of 2×1 tilings of a $2\times n$ area. $f_n = \#$ tilings

$(n\geq 2)$ $f_n = f_{n-1} + f_{n-2}$, $f_0 = f_1 = 1$ $\qquad \alpha_1 = \frac{1+\sqrt{5}}{2}$

$F(s) = \sum_{n\geq 0} s^n f_n = \frac{1}{1-s-s^2} = \frac{1}{(1-\alpha_1 s)(1-\alpha_2 s)}$ $\qquad \alpha_2 = \frac{1-\sqrt{5}}{2}$

$$= \frac{1}{\alpha_1-\alpha_2}\left(\frac{\alpha_1}{1-\alpha_1 s} + \frac{\alpha_2}{1-\alpha_2 s}\right)$$

$f_n = \text{coefficient of } s^n = \frac{1}{\alpha_1-\alpha_2}\left(\alpha_1^{n+1} - \alpha_2^{n+1}\right)$

The method of generating functions is robust:

$f_n = nf_{n-1} + f_{n-2}$, $s^n f_n = sn f_{n-1} + s^n f_{n-2}$

$F(s) - f_0 - f_1 s = s \ldots\ldots$

3.4 Independent Random Variables

Definition Discrete random variables X, Y are independent if

$$p(X=x, Y=y) = p(X=x)p(Y=y) \quad \forall x, y \in \mathbb{R}$$

This can be extended to families of random variables.

$\{X_i : i \in I\}$ is independent if

$$p(X_i = x_i \ \forall i \in J) = \prod_{i\in J} p(X_i = x_i) \quad \forall \text{ finite } J \subseteq I$$

The function $f_{X,Y}(x,y) = p(X=x, Y=y)$ is called the joint (probability) mass function of the pair x, y:

Definition

The covariance of X and Y:

$$\mathrm{cov}(X,Y) = E[(X-EX)(Y-EY)]$$

and the correlation coefficient is

$$\rho(X,X) = \frac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

1

X, Y are called <u>uncorrelated</u> if $\rho(X, Y) = 0$

* Note $\text{cov}(X, Y) = E(XY - YEX - XEY + EXEY)$

$$= E(XY) - E(X)E(Y)$$

<u>Theorem</u>

a) If X, Y are independent then $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$ for $g, h : \mathbb{R} \to \mathbb{R}$

b) If X, Y are independent, $\text{cov}(X, Y) = 0$, hence $\text{Var}(X)\text{Var}(Y) \neq 0$

c) There exist random variables X, Y that are dependent and uncorrelated.

<u>Proof</u>

a) $E[g(X)h(Y)] = \sum_{x,y} g(x)h(y)\, p(X=x, Y=y)$

$= \sum_{x,y} g(x)h(y)\, p(X=x)\, p(Y=y)$ by independence

$= \sum_{x} g(x)\, p(X=x) \sum_{y} h(y)\, p(Y=y) = E[g(X)]E[h(Y)]$

b) If independent, $E(XY) = E(X)E(Y) \Rightarrow \text{Cov}(X, Y) = 0$

c) U, V are $\text{Bern}(\frac{1}{2})$, independent.

$X = U + V$, $Y = |U - V|$

<u>Exercise</u> Show

$p(X=2, Y=1) \neq p(X=2)\, p(Y=1)$

and X, Y are uncorrelated.

12/02/11 Probability 10

Correlation as a measure of dependence

a) It is a single number

b) $-1 \leq \rho(X, Y) \leq 1$ (assume variances $\neq 0$)

Theorem (Schwarz's or Cauchy-Schwarz inequality)

$[E(XY)]^2 \leq E(X^2)E(Y^2)$ True for all random variables

Proof Let $Z = X + tY$ where $t \in \mathbb{R}$

$$0 \leq E(Z^2) = E(X^2 + 2tXY + t^2Y^2)$$

$$= E(X^2) + 2tE(XY) + t^2E(Y^2) \geq 0 \quad \forall t \in \mathbb{R}$$

$\Rightarrow$ Quadratic in t^2, discriminant ≤ 0 as there is at most 1 real root.

$$4[E(XY)]^2 - 4E(X^2)E(Y^2) \leq 0$$

$$\rho(X,Y)^2 = \frac{\mathrm{cov}(X,Y)^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} \leq 1$$

Use Inequality on $E[(X-EX)(Y-EY)]^2$

c) $\rho^2 = 1$ iff $(X - EX) + t(Y - EY) = 0$ for some $t \in \mathbb{R}$

iff $X + tY = c$ for some $t, c \in \mathbb{R}$ $\quad *$

and $\rho = 1$ iff t in $(*)$ satisfies $t < 0$, $\rho = -1$ iff $t > 0$

d) $\rho(X, Y) = 0$ if X, Y are independent

e) $\rho(aX + b, cY + d) = \rho(X, Y)$ if $ac > 0$

15/02/11

Probability 11

Theorem a) $Var(X+Y) = Var(X) + 2Cov(X,Y) + Var(Y)$

b) If X and Y are independent $Var(X+Y) = Var(X) + Var(Y)$

Proof

$$Var(X+Y) = E\left[(X+Y-E(X+Y))^2\right]$$
$$= E\left[(X-EX)^2 + 2(X-EX)(Y-EY) + (Y-EY)^2\right]$$
$$= Var(X) + 2Cov(X,Y) + Var(Y)$$

Examples

a) Variance of $bin(n,p)$ is $np(1-p)$

b) Negative Binomial, parameters k, p, has the same distribution as the sum of k independent $Geom(p)$ random variables $k\frac{(1-p)}{p^2}$

Sums of random variables

Theorem $p(X+Y=Z) = \sum_x p(X=x, y=z-x)$

Proof $\{X+Y=z\} = \bigcup_x \{X=x, Y=z-x\}$, a ~~countable~~ disjoint union

Since X, Y are discrete, $p(X+Y=z) = \sum_x p(X=x, Y=z-x)$

Corollary

If X, Y are independent then for $Z = X+Y$

$$f_Z(z) = \sum_x f_X(x) f_Y(z-x)$$

Convolution $f_Z = f_X * f_Y$

Theorem If X and Y are independent,

$$G_{X+Y}(s) = G_X(s)G_Y(s)$$

Proof $G_{X+Y}(s) = E(s^{X+Y}) = E(s^X s^Y) \overset{\text{independent}}{=} E(s^X)E(s^Y) = G_X(s)G_Y(s)$

Example

1. Let X be $Po(\lambda)$, Y be $Po(\mu)$ which are independent

$G_X(s) = e^{\lambda(s-1)}$

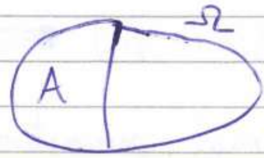
$G_{X+Y}(s) = e^{\lambda(s-1)} e^{\mu(s-1)} = e^{(\lambda+\mu)(s-1)}$ so $X+Y$ is $Po(\lambda+\mu)$

2. What is the pgf of the negative binomial distribution with parameters k, p?

It is the k^{th} power of the pgf of $Geom(p)$

i.e. $\left(\frac{ps}{1-qs}\right)^k$

3.5 Indicator Functions



$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

Note $E(I_A) = p(A)$, $Var(I_A) = p(A)p(\bar{A})$

Basic Facts

i) $I_{A\cap B} = I_A I_B$ ii) $I_{\bar{A}} = 1 - I_A$

$I_{A\cup B} = 1 - I_{\overline{A\cup B}} = 1 - I_{\bar{A}\cap\bar{B}} = 1 - I_{\bar{A}} I_{\bar{B}} = 1 - (1-I_A)(1-I_B)$

$= I_A + I_B - I_{A\cap B}$

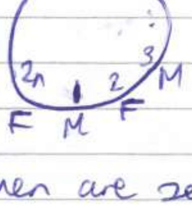
Take expectations: $p(A\cup B) = p(A) + p(B) - p(A\cap B)$

Example Inclusion-Exclusion Formula

$I_{\bigcup A_i} = 1 - \prod_{i=1}^{n}(1 - I_{A_i}) = 1 - \left[1 - \sum_i I_{A_i} + \sum_{i<j} I_{A_i} I_{A_j} \;\ldots\right]$

Take Expectations $p(\bigcup A_i) = \sum_i p(A_i) - \ldots$

Example



etc

$n \geq 2$ Male/Female Couples. The n men are seated randomly in the odd positions at a round table, and the women randomly in the even positions.

15/02/11

Probability (11)

Let $N_\circ :=$ # Men seated beside the right woman.

What are the mean and variance of N ?

Solution

Let A_i be the event that the i^{th} couple are next to each other.

$$N = \sum_i 1_{A_i} \quad , \quad E(N) = \sum_i E(1_{A_i}) = \sum_i p(A_i)$$

$$E(N) = n\, p(A_1) \quad \text{by symmetry}$$

$$E(N) = n \cdot \frac{2}{n} = 2 \qquad 1_A^2 = 1_A$$

$$E(N^2) = E\left(\sum_i 1_{A_i}^{\cancel{2}} + 2\sum_{i<j} 1_{A_i} 1_{A_j}\right)$$

$$= E(N) + 2\sum_{i<j} p(A_i \cap A_j)$$

$$= E(N) + 2\, p(A_1 \cap A_2)\binom{n}{2} \quad \text{by symmetry}$$

$$= E(N) + n(n-1)\, p(A_1 \cap A_2)$$

$$= E(N) + n(n-1)\, p(A_1)\, \underbrace{p(A_2 \mid A_1)}_{\text{so far unknown}}$$

$$= E(N) + n(n-1)\, p(A_1) \times \left(\frac{1}{n-1} \cdot \frac{1}{n-1} + \frac{n-2}{n-1} \cdot \frac{2}{n-1}\right)$$

$$= 2 + \frac{2(2n-3)}{n-1}$$

$$Var(X) = E(N^2) - (EN)^2 = \frac{2(n-2)}{n-1}$$

In fact $p(N = k) = f_n(k)$

$$p(N = k) \xrightarrow{(n \to \infty)} \frac{2^k e^{-2}}{k!} \qquad Po(2)$$

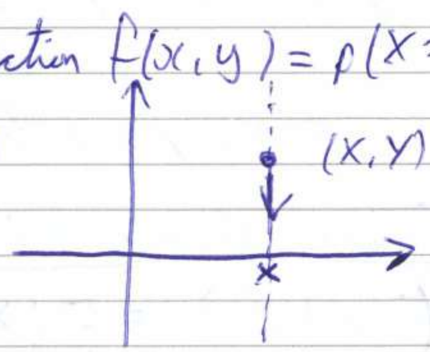
3

17/02/11 Probability (12)

3·6 Joint Distributions, Conditional Distributions

For X, Y discrete, the joint mass function $f(x, y) = p(X = x, Y = y)$

The marginal mass functions are:



$f_X(x) = \sum_y f(x, y)$

$f_Y(y) = \sum_y f(x, y)$

The conditional mass function of X given Y

$f_{X|Y}(x|y) = p(X = x \mid Y = y) = \dfrac{f(x,y)}{f_Y(y)} = \sum_x f(x, y)$

It is well defined if and only if $f_Y(y) \neq 0$

The conditional expectation of X given $Y = y$

is $E(X \mid Y = y) = \sum_x x f_{X|Y}(x|y)$

Writing $\varphi(y) = E(X \mid Y = y)$ we normally define the conditional expectation of X given Y as $\varphi(Y)$. This is a random variable. $\varphi(Y) = E(X|Y)$

Example $X_1, X_2, \ldots, X_n$ are independent Bern (p)

Find $E(X_1 | Y)$, $Y = X_1 + X_2 + \ldots + X_n$ Independent

Solution $E(X_1 \mid Y = y) = p(X_1 = 1 \mid Y = y)$

$$E(X \mid Y = y) = \frac{p(X=1, \cancel{Y} = y)}{p(Y=y)} = \frac{p(X_1 = 1)\, p(X_2 + X_3 + \ldots + X_n)}{p(Y=y)}$$

$$= \frac{p\binom{n-1}{y-1} p^{y-1}(1-p)^{n-y}}{\binom{n}{y} p^y (1-p)^{n-y}} = \frac{y}{n}$$

$$E(X|Y) = \frac{Y}{n}$$

A more Clever Method $X_1, X_2, \ldots, X_n$ iid (independent, identically distribute

$Y = X_1 + X_2 + \ldots + X_n$

$E(Y \mid Y) = Y$

$E(X_1 + \dots + X_n \mid Y) = E(X_1 \mid Y) + E(X_2 \mid Y) + \dots + E(X_n \mid Y)$

$= n E(X_1 \mid Y)$ by symmetry

$\therefore E(X_1 \mid Y) = Y/n$

Theorem (Properties of Conditional Expectation)

a) $E(E(X \mid Y)) = E(X)$ $\boxed{\text{Very Useful !!}}$

b) If X and Y are independent, $E(X \mid Y) = E(X)$, a constant

Proof

a) $E(E(X \mid Y)) = \sum_y E(X \mid Y = y) p(Y = y)$

$= \sum_y \left[\sum_x x \, p(X = x \mid Y = y) \right] p(Y = y)$

$= \sum_{x,y} x \, p(X = x, Y = y) = \sum_x x \, p(X = x) = E(X)$

b) Obvious from the definition of conditional expectation.

Reminder $X_1, X_2, \dots$ iid, taking values in $\{0, 1, 2, \dots\}$

$S = X_1 + X_2 + \dots + X_n$

$G_S(s) = G_{X_1}(s) \dots G_{X_n}(s) = G_X(s)^n$

Theorem The random sum formula

Let $N, X_1, X_2, \dots$ be independent, taking values in $\{0, 1, 2, \dots\}$

Suppose the X_i are iid with pgf G. Then

$T = X_1 + X_2 + \dots + X_N$ has pgf $G_T(s) = G_N(G(s))$

Example $p(N = k) = 1$ $G_N(s) = s^k$

Proof $G_T(s) = E(s^T) = E(E(s^T \mid N))$

$G_T(s) = E[G(s)^N] = G_N(G(s))$

17/02/11

Probability (12)

$E(S^T) = E(E(S^T|N)) = \sum_n E(S^T|N=n)\,p(N=n)$

$= \sum_n G(s)^n p(N=n) = G_N(G(s))$

$N=m$

$E(T|N=n) = nE(X_1$

Example $p(N=k)=1$, $G_N(s) = s^k$

Theorem $E(T) = E(N)E(X_1)$

Proof $E(T) = G_T'(1) = G_N'(G(1))\,G'(1) = E(N)E(X_1)$

Exercise Find $Var(T)$

$Var(T|N=n) = n\,Var(X_1)$

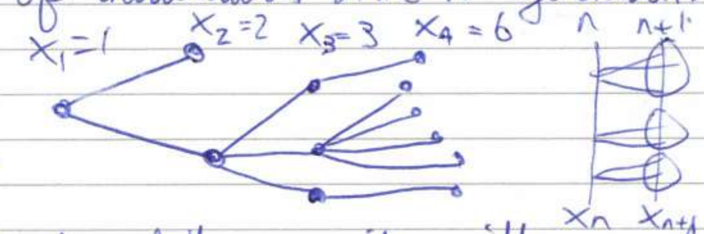
In general $Var(T) \neq E(N)\,Var(X_1)$

19/02/11

Probability (13)

3·7 Branching Process

A model for population growth (bacterial, spread of a family name
Sometimes known as the Goulton-Watson (-Bienaymé) process.
It deals with growth in generations.
Let X_n be the number of individuals in the n^{th} generation.



Assumptions

a) $X_0 = 1$, a progenitor

b) X_1 is the number of offspring of the progenitor with mass function $f(k) = p(X_1 = k)$.

c) Each member of the process has a family whose size has mass function f.

d) All offspring have family sizes which are independent of one another (iid

We can draw a family tree of the branching process, a random tree.
$X_{n+1} = Y_1 + Y_2 + Y_3 + \dots + Y_{X_n}$ where the Y_i are iid, mass function f, and independent of X_n. X_{n+1} is a sum of a random number, X_n, of independent family sizes.

Let $G_N(s) = E(S^{X_n})$, the pgf of X_n.

Theorem

$G_{n+1}(s) = G_n(G(s))$, where $G(s)$ is the pgf of a family size.
i.e. $G(s) = E(S^{X_1}) = \sum_k S^k f(k)$

$X_{n+1} = A_1 + \dots + A_{X_1}$ and the A_i are iid with distribution of X_n.

$G_{n+1}(s) = G(G_n(s))$

Proof

By decomposition of the tree, and the random sum formula.

Hence $G_n(s) = G(G_{n-1}(s)) = G(G(\ldots(G(s))\ldots))$, n times.

Corollary

Let $\mu = E(X_1) < \infty$, $\sigma^2 = Var(X_1) < \infty$, then:

$$E(X_n) = \mu^n \quad , \quad Var(X_n) = \begin{cases} n\sigma^2 & \text{if } \mu = 1 \\ \dfrac{\sigma^2\mu^{n-1}(\mu^n - 1)}{\mu - 1} & \text{if } \mu \neq 1 \end{cases}$$

Proof

$$G_n'(1) = G_{n-1}'(G(1))\,G'(1) \qquad\qquad G_n(s) = G_{n-1}(G(s))$$

$$= G_{n-1}'(1)\,G'(1)$$

$$E(X_n) = E(X_{n-1})\mu = E(X_0)\mu^n = \mu^n$$

The calculation of variance is left as an exercise.

Example

Let X_1 have the geometric distribution. $p(X_1 = k) = pq^k$

$k = 0, 1, 2, \ldots, \; p + q = 1, \; p \neq q, \; pq \neq 0$

$$G(s) = \sum_0^\infty s^k pq^k = \frac{p}{1 - qs} \qquad (\text{if } |qs| < 1)$$

$$G_n(s) = p\,\frac{(q^n - p^n) - qs(q^{n-1} - p^{n-1})}{(q^{n+1} - p^{n+1}) - qs(q^n - p^n)} \qquad |s| \leq 1$$

Proof

By induction

19/02/11

Probability (13)

Hence $p(X_n = k)$ for general k, n.

$\mu = E(X_1) = q/p \neq 1$

$\therefore E(X_n) = \mu^n = \left(q/p\right)^n$

<u>Extinction</u>

$$p(X_n = 0) = G_n(0) = \frac{\mu^n - 1}{\mu^{n+1} - 1} \xrightarrow{n \to \infty} \begin{cases} 1 & \mu < 1 \\ 1/\mu & \mu > 1 \end{cases}$$

22/02/11

Probability (14)

Problem of extinction

$G(s) = E(s^{X_1})$

Let $A_n = \{X_n = 0\} \subseteq A_{n+1}$ $\quad A_n \subseteq A_{n+1} \subseteq \dots \subseteq \lim_{m\to\infty} A_m$

$\lim_{m\to\infty} A_m := \bigcup_n A_n = \{\text{ultimate extinction}\}$

Theorem

If $B_1, B_2, \dots$ is an increasing sequence of events then

$p(\bigcup_1^\infty B_i) = \lim_{i\to\infty} p(B_i)$ i.e. $p(\lim B_i) = \lim p(B_i)$

"Probability measures are continuous set functions"

A similar statement for intersections of decreasing sequences exists.

Proof

$B_n \setminus B_{n-1} =: C_n$, $C_1 = B_1$, then the C_n are disjoint

$\lim_{i\to\infty} P(B_i) = \lim_{i\to\infty} P(C_1 \cup C_2 \cup \dots \cup C_i)$, a disjoint union

$= \lim_{i\to\infty} \sum_{j=1}^{i} p(C_j) = \sum_{j=1}^{\infty} p(C_j) = p(\bigcup_j C_j) = p(\bigcup_j B_j)$ $\square$

Branching Processes

Let $\eta = p(\text{ultimate extinction}) = \lim_{n\to\infty} p(X_n = 0)$ by the last theorem

Theorem

η is the smallest non-negative root of the equation $x = G(x)$

Proof

Let $\eta_n = p(X_n = 0)$, so $\eta_n \to \eta$

$\eta_n = G_n(0) = G(G_{n-1}(0)) = G(\eta_{n-1})$

As $n \to \infty$, $\eta_n \to \eta$, $G(\eta_{n-1}) \to G(\eta)$ by continuity of G

$\eta = G(\eta)$

Let e be any non negative root of $x = G(x)$

$\eta_1 = G(0) \leq G(e) = e$ $\qquad G(x) = \sum_k x^k p_k$

$\eta_2 = G(\eta_1) \leq G(e) = e$

By induction $\eta_n \leq e \ \forall e$, hence $\eta \leq e$ $\square$

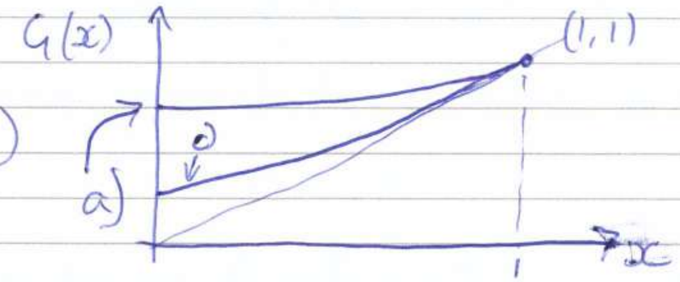
<u>Theorem</u>

a) If $\mu < 1$, then $\eta = 1$

b) If $\mu > 1$, then $\eta < 1$

c) If $\mu = 1$, and $Var(X_1) > 0$, then $\eta = 1$

<u>Proof</u> (by picture)



$G(1) = 1$

$G(0) = p(X_1 = 0)$

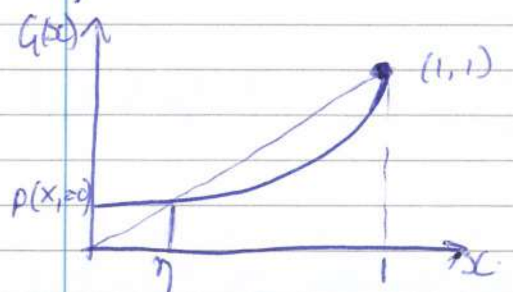
G is increasing

G is convex because $G''(x) = \sum k(k-1) x^{k-2} p_k \geq 0$

$G'(1) = \mu$

a) When $\mu < 1$, the only solution to $x = G(x)$ is $x = 1$, $\therefore \eta = 1$

b) $\mu > 1$ $\qquad$ Then there exists another root in $[0,1)$ and this is η.



2

/02/11 Probability (14)

3·8 Random Walk

$\overset{q=1-p}{\longleftarrow} \bullet \overset{p}{\longrightarrow}$

Consider a random walk on $\{0, 1, \ldots, N\}$, with absorbing barriers at 0 and N.

Let $M :=$ # steps up to the moment of absorption at either 0 or N

Let $e_k = E(M \mid \text{start at } k)$

$e_k = E\big(E(M \mid 1^{st} \text{ step})\big) = p(e_{k+1} + 1) + q(e_{k-1} + 1)$ $\quad 1 \le k \le N-1$

$e_0 = 0 = e_N$

(*) $p e_{k+1} - e_k + q e_{k-1} = -1$

General solution $= \begin{cases} A\left(\frac{q}{p}\right)^k + B & q \neq p \\ A + Bk & q = p \end{cases}$

Particular Solution $\begin{cases} -\frac{k}{p-q} & p \neq q \\ \end{cases}$

$p \neq q$, $e_k = -\frac{k}{p-q} + G.S.$, $A = \frac{N}{(p-q)\left(\left(\frac{q}{p}\right)^N - 1\right)}$, $B = -A$

$p = q$, $e_k = k(N-k)$

4. Continuous Random Variables

4·1 Density functions

$(\Omega, \mathcal{F}, p)$, $X: \Omega \to \mathbb{R}$, Distribution functi

Distribution Function $F_X(x) = p(X \le x)$

Definition X is called "continuous" if there exists $f: \mathbb{R} \to \mathbb{R}$ such that a) $p(X \le x) = \int_{-\infty}^{x} f(u)\,du$, $x \in \mathbb{R}$

b) $f(u) \geq 0 \quad \forall u$

If this holds, f is called the probability density function of X.

24/02/11 Probability (15)

If $F_X(x) = \int_{-\infty}^{x} f(u)\,du$, $f \geqslant 0$

f is "the" pdf of X.

<u>Note</u>

i) If F_X is differentiable we take the pdf to be $f_X = F_X'$

ii) Assume henceforth that X has pdf f.

$p(X = x) = 0 \quad \forall x \in \mathbb{R}$

<u>Proof</u>

$\{X = x\} = \bigcap_n \{X \in (x - \frac{1}{n}, x)\}$

$p(X = x) = \lim_{n \to \infty} p(x - \frac{1}{n} < X \leq x)$ since P is a continuous set function

$= \lim \left[\int_{-\infty}^{x} f\,du - \int_{-\infty}^{x-\frac{1}{n}} f\,du \right]$

$= \lim_{n \to \infty} \int_{x-\frac{1}{n}}^{x} f(u)\,du = 0$

iii) $p(a \leq X \leq b) = \int_a^b f(u)\,du$

(<u>Proof</u> $p(X \leq X \leq b) = p(X = a) + p(a < X \leq b)$

$= 0 + \int_{-\infty}^{a} f(u)\,du - \int_{-\infty}^{b} f(u)\,du = \int_a^b f(u)\,du$)

iv) the pdf f is characterised by: $f(u) \geqslant 0 \quad \forall u$

f is integrable with $\int_{-\infty}^{\infty} f(u)\,du = 1$

v) For a mass function, the element of probability is $f(x)$; for a density function, it is $f(x)\,dx$

<u>Note</u> Proofs for discrete distributions are often valid also for continuous distributions with $f(u) \to f(u)\,du$

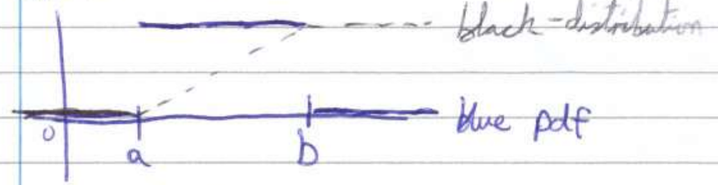
$\sum \to \int$

Examples

1. Uniform Distribution Unit $[a, b]$

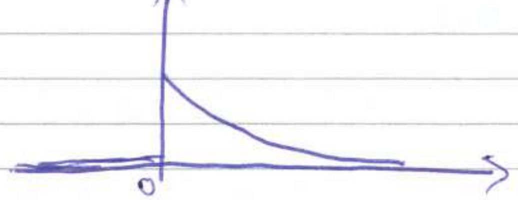
$$f(u) = \begin{cases} 0 & \text{if } u \notin [a,b] \\ c & \text{if } u \in [a,b] \end{cases} \quad \text{for some } c$$

$$\int_{-\infty}^{\infty} f(u)\, du = c(b-a) = 1 \quad \therefore \; c = \frac{1}{b-a}$$



2. Exponential Distribution $Exp(\lambda)$

$$f(u) = \begin{cases} 0 & u \leq 0 \\ \lambda e^{-\lambda u} & u > 0 \end{cases}$$



$$F(x) = \int_{-\infty}^{x} f(u)\, du = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

Important Property 'lack of memory' or 'memoryless' property

Let X be $Exp(\lambda)$. We need $p(X > y + z \mid x > y)$

$$p(X > y+z \mid x > y) = \frac{p(X > y+z)}{p(X > y)} = \frac{e^{-\lambda(y+z)}}{e^{-\lambda y}} = e^{-\lambda z} \qquad y, z \geq 0$$

Conversely, if F is a distribution function with pdf f, and

$$\frac{1 - F(y+z)}{1 - F(y)} = 1 - F(z), \quad y, z \geq 0, \text{ take } F(0) = 0$$

then F is the distribution function for an exponential distribution.

This is a key property in the theory of Markov processes and more generally, to random/stochastic processes.

24/02/11

3. Normal / Gaussian Distribution

$N(0,1)$: $f(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$, $u \in \mathbb{R}$

More generally : $N(\mu, \sigma^2)$: $g(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(u-\mu)^2}{2\sigma^2}\right]$, $u \in$

i.e. $N(0,1)$ changed by location μ and scaled by σ

Change of Variables 4.2 If X has pdf f, and $h : \mathbb{R} \to \mathbb{R}$, what is the pdf of $h(X)$?

More generally : What is the distribution of $h(X)$ in terms of the distribution function of X ?

$p[h(x) \le y] = p[X \in h^{-1}(-\infty, y]]$ + calculation

$= F_{h(x)}(y)$ $= p(X \le h^{-1}(y))$ if h is strictly increasing

$= F_X[h^{-1}(y)]$

assuming h is sufficiently smooth

Take X to have pdf f. $f_{h(x)}(y) = f[h^{-1}(y)] \frac{d}{dy}[h^{-1}(y)]$

28/02/11

Probability (16)

$Y = h(X)$. If all functions are sufficiently smooth

$$f_Y(y) = f_X[h^{-1}(y)]\left|\frac{d}{dy}[h^{-1}(y)]\right|$$

Example

If X is Unif$[0,1]$, $h(x) = -\log x$, $Y = h(X)$

$$p(Y \leq y) = p(-\log X \leq y) = p(\log X \leq -y)$$

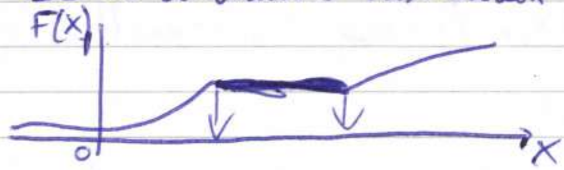
$$= p(X \geq e^{-y}) = 1 - e^{-y}, \ y > 0$$

$$f_Y(y) = e^{-y}, \ y > 0$$

Important Method

X is Unif$[0,1]$, let F be a continuous distribution function

Let $Y = F^{-1}(X)$



[For the sake of rigour,

$F^{-1}(V)$ is the infemum of $\{x : F(x) = V\}$]

$$p(Y \leq y) = p(F^{-1}(X) \leq y) = p(X \leq F(y)) = F(y)$$

Y has distribution F.

See Monte Carlo Methods

Example If $X \sim N(0,1)$

Let $Y = \sigma X + \mu$ $\qquad \sigma, \mu \in \mathbb{R}$

$$h(x) = \sigma x + \mu = y, \quad x = \frac{y-\mu}{\sigma} = h^{-1}(\sigma)$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]\frac{1}{|\sigma|}$$

4.3 Expectation

X discrete, $E(X) = \sum_x x P(X=x)$

X continuous, $E(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$ whenever the integral is absolutely convergent

Theorem

If X has pdf f, and $\int_{-\infty}^{\infty} |g(x)|\, f(x)\, dx < \infty$ then

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

Proposition If X is a continuous random variable

$$E(X) = \int_0^{\infty} P(X > x)\, dx - \int_0^{\infty} P(X < -x)\, dx$$

If $E(X)$ exists. This may be used as a definition of $E(X)$ for any X, regardless of type.

If X has pdf f
$p(X \in A) = \int_A f(x)\, dx$

Proof

$$\int_0^{\infty} p(X > x)\, dx = \int_0^{\infty} \left[\int_x^{\infty} f(u)\, du \right] dx$$

$$= \int_0^{\infty} du\, f(u) \int_0^{u} dx = \int_0^{\infty} u f(u)\, du$$

and similarly $\int_0^{\infty} p(X < -x)\, dx = -\int_{-\infty}^{0} u f(u)\, du$

Proof

$$\int_0^{\infty} P(g(X) > y)\, dy = \int_0^{\infty} dy \int_{\{x : g(x) > y\}} f(x)\, dx$$

$$= \int_{\{x : g(x) > 0\}} dx\, f(x) \int_0^{g(x)} dy = \int g(x) f(x)\, dx$$

The 2nd integral is $-\int_{\{x : g(x) < 0\}} g(x) f(x)\, dx$

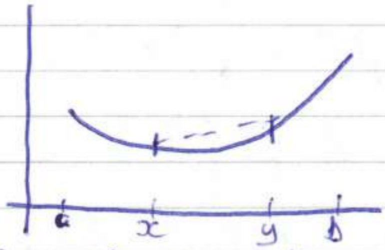
hence the claim is proved.

Interchanging orders of integration is validated by a result called Fubini's Theorem.

Note: Using discrete theory, one now defines mean, variance, moments, covariance

01/03/2011

Probability 17



5.1 Three Inequalities

5.1 Jensen's Inequality

Definition A function $u : (a,b) \to \mathbb{R}$ is called convex if

$u[px + (1-p)y] \leq pu(x) + (1-p)u(y)$

$\forall x, y \in (a,b)$, $p \in [0,1]$

u is concave if $-u$ is convex. N.B. Convexity $\Rightarrow$ continuity

Examples

$u(x) = -\log(x)$, $u(x) = \frac{1}{x}$ on $(0, \infty)$

Fact If u'' exists and satisfies $u''(x) \geq 0$ for $x \in (a,b)$ then u is convex.

Theorem

Let X be a random variable taking values in some open interval (a,b) and let u be convex on (a,b). Then

$u(EX) \leq E[u(X)]$

Example AM - GM inequality

Let $x_1, x_2, \ldots, x_n > 0$

Let $p(X = x_i) = \frac{1}{m}$, $i = 1, 2, \ldots, m$

$u(x) = -\log(x)$

By Jensen's Inequality $u\left(\frac{1}{m}\sum_1^m x_i\right) \leq \sum_{i=1}^m \frac{1}{m} u(x_i)$

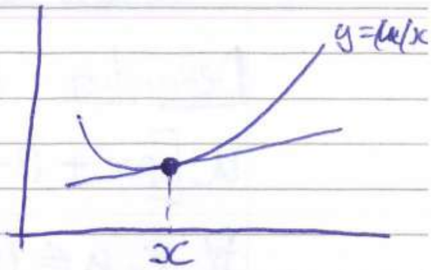
$-\log(AM) \leq -\log(GM)$

$AM \geq GM$

Example $p(X>0)=1$, $u(x)=\frac{1}{x}$

$\frac{1}{EX} \leq E(1/X)$. In general $E(1/X) \neq \frac{1}{EX}$

unless $Var(X)=0$.



Proof of Jensen's Inequality

Theorem (Supporting hyperplane theorem)

u is convex on (a,b) if and only if

$\forall x \in (a,b)$, $\exists \lambda \in \mathbb{R}$ such that $u(y) \geq \lambda(y-x)+u(x)$ $\forall y \in (a,b)$

The proof is to be discussed later.

Let $x=EX$. By the supporting hyperplane theorem,

$\exists \lambda$ such that $u(y) \geq \lambda(y-EX) + u(EX)$

$\therefore u(X) \geq \lambda(X-EX)) + u(EX)$

$E[u(X)] \geq 0 + u(EX)$

More on Jensen

Lemma If u is convex then

$u\left(\sum_i^m p_i x_i\right) \leq \sum_i^m p_i u(x_i)$ for $x_i \in (a,b)$

$p_i > 0$ with $\sum p_i = 1$.

This is equivalent to Jensen's Inequality for discrete random variables taking finitely many values.

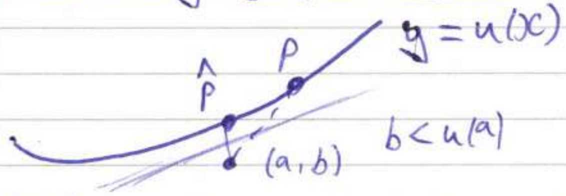
01/03/11 Probability (17)

Proof (by induction on m)

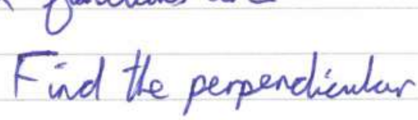
$m=2$ holds by the definition of continuity. Assume this is true for $m=k\geq 2$

$$u\left(\sum_1^{k+1} p_i x_i\right) = u\left((1-p_{k+1})\frac{\sum_1^k p_i x_i}{1-p_{k+1}} + p_{k+1}x_{k+1}\right)$$
$$\leq (1-p_{k+1})u(\ldots) + p_{k+1}u(x_{k+1}) \text{ by convexity}$$
$$\leq (1-p_{k+1})\sum_1^k \frac{p_i}{1-p_{k+1}}u(x_i) + p_{k+1}x_{k+1} \text{ by the induction hypothesis}$$

Supporting hyperplane theorem



$d(P)$ = Euclidean distance P to (a,b). Find $\hat{p}$ such that $d(\hat{P})\leq d(P)$ $\forall$ p on the curve. Since convex functions are continuous, $\inf_p d(P)$ is attained at some $\hat{p}$. Find the perpendicular bisector of the line from (a,b) to $\hat{p}$. This bisector does not intersect the curve (by convexity) as this would cause a contradiction. This line is a "separating hyperplane".



From separating, to supporting hyperplanes

Find (a_i, b_i) such that $b_i < u(a_i)$ and $a_i\to c$, $b_i\to u(c)$

For each i, there exists a separating line $y=\alpha_i x+\beta_i$

$\{(\alpha_i,\beta_i): i\geq 1\}$ possesses some limit point (α,β). The line $y=\alpha x+\beta$ is the required supporting hyperplane.

03/03/11

Probability (18)

5-2 Chebyshev's Inequality

If $Var(X)$ is small, in what sense is X near to a constant?

Theorem (Markov's Inequality)

If $E(X)$ exists, then $p(|X| \geq a) \leq \frac{E(|X|)}{a}$ for $a > 0$

Proof

Let $A = \{|X| \geq a\}$ ($a>0$). Then $|X| \geqslant a\, I_A$ (Check on A and $\bar{A}$)

$\therefore E(|X|) \geqslant E(a I_A) = a\, P(A)$ $\square$

Theorem (Chebyshev's Inequality)

$$p(|X - E(X)| \geq a) = \frac{Var(X)}{a^2}, \quad a > 0$$

Proof

$$p(|X - E(X)| \geqslant a) = p([X - E(X)]^2 \geqslant a^2)$$
$$\leq \frac{1}{a^2} E[(X - EX)^2] \text{ by Markov}$$

$$p(|X - EX| \geqslant a) \leq \frac{Var(X)}{a^2}$$

$$p(X \geqslant a) = p(e^{\theta x} \geqslant e^{\theta a}) \qquad \theta > 0$$
$$\leq \frac{E(e^{\theta x})}{e^{\theta a}}$$

$$\therefore p(X \geqslant a) \leq \inf\{e^{-\theta a} E(e^{\theta x}) : \theta > 0\}$$

This leads to the theory of large deviations.

5.3 Law of large numbers (relating to repeated experimentation)

Theorem Let $X_1, X_2, \ldots$ be iid random variables with finite variance and mean μ. Let $S_n = \sum_{i=1}^{n} X_i$

a) $E\left[\left(\frac{S_n}{n} - \mu\right)^2\right] \to 0$ as $n \to \infty$

b) $p\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon \right) \to 0$ as $n \to \infty$ $\forall \varepsilon > 0$

a) "mean square convergence" "convergence in L^2"

b) The weak law of large numbers. There is also a strong law.

Language

$X_n \to X$ in mean square, or L^2, if $E[(X_n - X)^2] \to 0$

$X_n \to X$ in probability if $\forall \varepsilon > 0$

$p(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$

There are also other forms of convergence

Repeated Experimentation

Repeat an Experiment. Each time, we observe whether A occurs or not. $A_i = \{A$ occurs on the i^{th} experiment$\}$

$\frac{1}{n} \sum_{i=1}^{n} 1_{A_i}$ should converge to something which we can interpret as $p(A)$.

Proof

a) $E\left(\frac{S_n}{n}\right) = \frac{1}{n} E(S_n) = \frac{1}{n} n\mu = \mu$

$E\left[\left(\frac{S_n}{n} - \mu\right)^2\right] = Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2} Var(S_n) = \frac{1}{n} Var(X) \to 0$ as $n \to \infty$

03/03/11 Probability 18

b) $P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{Var\left(\frac{S_n}{n}\right)}{\varepsilon^2}$ by Chebyshev

$= \frac{Var(X)}{n\varepsilon^2} \to 0$ as $n \to \infty$

05/03/11 Probability (19)

4.4 Functions of Random Variables

We may have a collection $\underline{X} = (X_1, X_2, \ldots, X_n)$ on $(\Omega, \mathcal{F}, p$

We use a joint distribution function:

$F_{\underline{x}}(\underline{x}) = p(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n)$

If $F_{\underline{x}}(\underline{x}) = \underbrace{\int\int \ldots \int}_{n \text{ times}}_{\underline{u} \leq \underline{x}} f_{\underline{x}}(\underline{u})\, du$ $\qquad \underline{x} \in \mathbb{R}^n$, $\underline{x} = (x_1, x_2, \ldots, x_n)$

and $f_{\underline{x}}(\underline{x}) \geq 0$, $f_{\underline{x}}$ is the joint distribution function of $\underline{X}$.

Normally $f_{\underline{x}}(\underline{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \ldots \partial x_n} F_{\underline{x}}(\underline{x})$

Notes If $F_{X,Y}(x, y) = p(X \leq x, Y \leq y)$ $\qquad \left[\lim_{y \to \infty} F(x,y)\right.$

a) The marginal distribution function of X is $F_X(x) = F_{X,Y}(x, y)$

b) The marginal pdf of X is $f_X(x) = \frac{d}{dx} F_{X,Y}(x, \infty)$

$f_X(x) = \frac{d}{dx} \int_{-\infty}^{x} \int_{-\infty}^{\infty} f(u,v)\, du\, dv = \int_{-\infty}^{\infty} f_{X,Y}(x, v)\, dV$

c) The "basic element of probability" is

$p(x < X < x+dx, y < Y < y+dy) \approx f_{X,Y}(x, y)\, dx\, dy$

$p[(X, Y) \in A] = \iint_A f_{X,Y}(x, y)\, dx\, dy$

d) X, Y are independent if the joint distribution function factorises as $F_{X,Y}(x, y) = F_X(x) F_Y(y)$, $x, y \in \mathbb{R}$

i.e. in the continuous case $f_{X,Y}(x, y) = f_X(x) f_Y(y)$, $x, y \in \mathbb{R}$

Reminder $A_1, A_2, \ldots, A_n$

are independent events if and only if

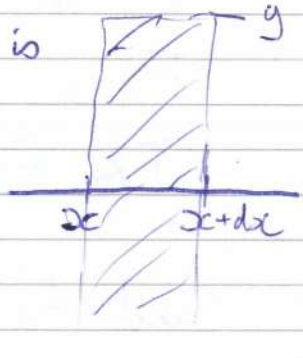
$p(A_1 \cap A_2 \cap \ldots \cap A_n) = \prod_{i=1}^{n} p(A_i)$. We can discuss the independence of a family $\{X_1, X_2, \ldots, X_n\}$ of random variables in a similar fashion.

Application

If X, Y are independent, $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$

and hence $Var(X+Y) = Var(X) + Var(Y)$

e) The conditional ~~distribution~~ density function of Y given X is



$$\lim_{dx \to 0} \frac{\partial}{\partial y} P(Y = y \mid x < X < x+dx)$$

$$\approx \lim_{dx \to 0} \frac{\partial}{\partial y} \frac{\int_{-\infty}^{y} f(x,v)\,dx\,dv}{f_X(x)\,dx} = \frac{f(x,y)}{f_X(x)}$$

Definition

The conditional density function of Y given X is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

f) The conditional expectation of Y given X is

$\psi(X) = E(Y|X)$ given by

$$\psi(x) = \text{"}E(Y|X=x)\text{"} = \int_{-\infty}^{\infty} y f_{Y|X}(y|x)\,dy$$

Theorem $E[E(Y|X)] = E(Y)$

4.5 Changes of variable

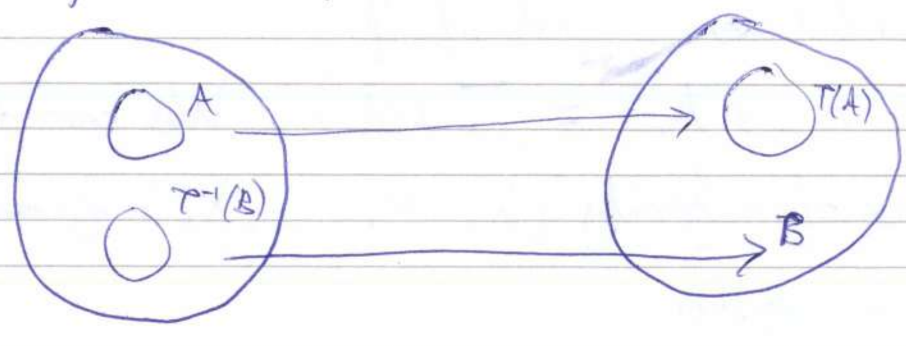
General Question If X, Y have joint pdf f

$U = u(X,Y)$, $V = v(X,Y)$

What is the joint pdf of (U, V)?

$T: \mathbb{R}^2 \to \mathbb{R}^2$, $T: (x,y) \mapsto (u(x,y), v(x,y))$

$(U,V) = T(X,Y)$



05/03/11

Probability (19)

$$p[(U,V) \in B] = P[(X,Y) \in T^{-1}(B)] = \iint_{T^{-1}(B)} f(u,v)\,du\,dv$$

$$D = \{(x,y) : f(x,y) > 0\}$$

Let S be $T(D)$. $T : D \Rightarrow S$

Assume that T is invertible on S, i.e. T is bijective on D.

$$p[(u,v) \in B] = \iint_{T^{-1}(B)} f(x,y)\,dx\,dy$$

$$= \iint_{B} f[x(u,v), y(u,v)]\,|J|\,du\,dv$$

$$J = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

$$\therefore f_{u,v}(u,v) = \begin{cases} f[x(u,v), y(u,v)]\,|J| & , u,v \in S \\ 0 & u,v \notin S \end{cases}$$

Example

X, Y are independent, $Exp(1)$.

Let $U = X + Y$, $V = \frac{X}{X+Y}$

$f(x,y) = e^{-x-y}$ for $x, y > 0$

$u = x+y$, $v = \frac{x}{x+y}$, so $x = uv$, $y = u(1-v)$

$T : (0,\infty)^2 \Rightarrow (0,\infty) \times (0,1)$, a bijection

$$\text{Jacobian} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 1-v & -u \end{vmatrix} = -u$$

$\therefore f_{u,v}(u,v) = e^{-uv - u(1-v)}|u|$ $u > 0, 0 < v < 1$

$f_{u,v}(u,v) = ue^{-u}$

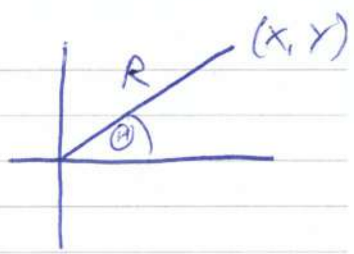
Hence: $f_u(u) = ue^{-u}, u > 0$. $f_v(v) = 1$ $0 < v < 1$

and $f_{u,v}(u,v) = f_u(u)\, f_v(v)$

so u and v are independent

08/03/10

Probability (20)



<u>Example</u> X, Y are independent, $N(0,1)$

$R = \sqrt{X^2+Y^2}$ $\quad \Theta = \arctan(Y/X)$

Use $x = r\cos\theta$, $y = r\sin\theta$

$$f_{R,\Theta}(r,\theta) = f_{X,Y}(r\cos\theta, r\sin\theta) = \frac{r}{2\pi} e^{-\frac{1}{2}r^2} \quad r>0, \ \theta \in [0,2\pi)$$

Therefore R, Θ are independent, Θ is Unif $[0,2\pi)$

R has pdf $re^{-\frac{1}{2}r^2}$, $r>0$

4.6 Bivariate (multivariate) normal distribution

$$N(\mu, \sigma^2) \qquad f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

<u>Exercise</u> The mean is $\int_{-\infty}^{\infty} x f(x)\,dx = \mu$

The variance is $\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = \sigma^2$

If $X \sim N(\mu, \sigma^2)$, $Y = \frac{X-\mu}{\sigma}$ is $N(0,1)$.

<u>Bivariate Case</u>

$$f(x,y) = c_1 \exp\left[-c_2 Q(x,y)\right]$$

where Q is a quadratic form in x and y. We take

$$(1-\rho^2)\,Q(x,y) = \left(\frac{x-\mu_1}{\sigma_1}\right)^2 + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right)$$

$$f(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}Q(x,y)\right], \quad x,y \in \mathbb{R}$$

Parameters: $\sigma_1, \sigma_2 > 0$, $\mu_1, \mu_2 \in \mathbb{R}$, $|\rho| < 1$

<u>Note</u> $Q(x,y) = (\underline{x} - \underline{\mu})^T \Sigma^{-1} (\underline{x} - \underline{\mu})$

$\underline{x} = (x, y)$, $\underline{\mu} = (\mu_1, \mu_2)$, $\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$

$$\Sigma = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix}$$

Let $U = \frac{X-\mu_1}{\sigma_1}$, $V = \frac{Y-\mu_2}{\sigma_2}$

$$f_{U,V}(u,v) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{u^2-2\rho uv+v^2}{2(1-\rho^2)}\right]$$

$$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u,v)\,dv$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left[(v-\rho u)^2 + u^2(1-\rho^2)\right]\right] dv$$

$$= \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left[-\frac{(v-\rho u)^2}{2(1-\rho^2)}\right] dv}_{N(\rho u,\, 1-\rho^2)}$$

$$= \frac{e^{-\frac{u^2}{2}}}{\sqrt{2\pi}} \times 1$$

U is normal with parameters 0 and 1, V is also $N(0,1)$.

$f_{U,V}(u,v) = f_U(u)\, f_{V|U}(v|u)$. Given $U=u$, V is $N(\rho u, 1-\rho^2)$

$Q(u,v) = (u, v)\, A \begin{pmatrix} u \\ v \end{pmatrix}$

$$A = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}^{-1} = \begin{pmatrix} Var(U) & cov(U,V) \\ cov(U,V) & Var(V) \end{pmatrix}^{-1}$$

Correlation and Covariance

$$E(UV) = E\left[E(UV|U)\right] = E\left[U\,E(V|U)\right]$$

$$E(UV) = E(U \cdot \rho U) = \rho E(U^2) = \rho\, Var(U) = 1$$

$\therefore$ ρ is $Cov(U,V) = corr(U,V)$

Very Important Properties

1. U, V are ~~uncorrelated~~ independent if and only if they are uncorrelated (i.e $\rho = 0$)
2. If U and V have a bivariate normal distribution then $\alpha U + \beta V$ has a normal distribution, for any given $\alpha, \beta \in \mathbb{R}$.

Actually, this characterises the normal distribution.

08/03/10

Probability 20

Hence, for example: $(X_1, X_2, \dots, X_n)$ is said to have a multivariate normal distribution, MVN, if:

$\sum_1^n a_i X_i$ is univariate normal, for any $a_1, a_2, \dots, a_n \in \mathbb{R}$.

In general:

For $\underline{X} = (X_1, X_2, \dots, X_n)$

the mean vector is $\underline{\mu} = (EX_1, EX_2, \dots, EX_n)$

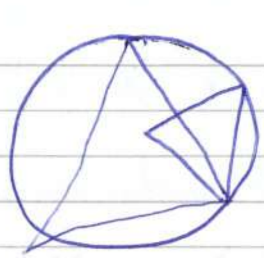
the covariance matrix is $V = (V_{ij}) \qquad (n \times n)$

$$= \mathrm{Cov}(X_i, X_j)$$

$$E\left[(\underline{X}-\underline{\mu})(\underline{X}-\underline{\mu})^T\right]$$

3

10/03/11 Probability (21)



6. Geometrical Probability

6.1 Bertrand's Paradox

A chord of the unit circle is picked at random. What is the probability that an equilateral triangle with the chord as its base, fits within the circle?

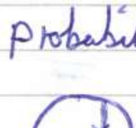
a) 

Assume D is Unif $[0,1]$. $D = \frac{1}{2}$ gives the largest allowed triangle, so the triangle lies in the circle if and only if $D \geq \frac{1}{2}$. $p(D \geq \frac{1}{2}) = \frac{1}{2}$

b) 

Assume the acute angle A between the chord and tangent at an endpoint is Unif $[0, \frac{\pi}{2}]$

probability $= p(A \geq \frac{\pi}{3}) = \frac{\pi/3}{\pi/2} = \frac{2}{3}$

c) 

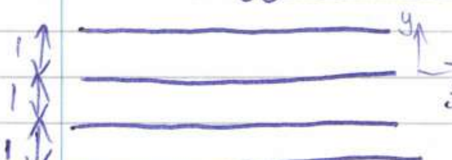
Pick a point uniformly on the disc. Draw a chord with this point on centre.

$p(D \leq d) = \frac{\pi d^2}{\pi} = d^2$ for $d \in (0,1)$

Answer $= p(D \geq \frac{1}{2}) = 1 - p(D \leq \frac{1}{2}) = 1 - \frac{1}{4} = \frac{3}{4}$

d) Choose p and q as independent points on the circumference, each having the uniform distribution. Answer $= \frac{2}{3}$

6.2 Buffon's Needle



Parallel lines

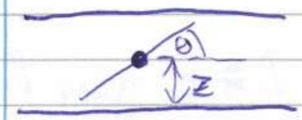
A unit needle is dropped at random onto a plane, ruled by parallel straight lines, each of which is a unit distance apart.

What is the probability the needle intersects some line?

Let (X, Y) be the coordinates of the centre of the needle, Θ be the inclination to the x-axis. Assume:

- $Z = Y - \lfloor Y \rfloor$ is Unif$[0,1]$
- Y, Θ are independent
- Θ is Unif$[0, \pi]$

$f_{Z,\Theta}(z, \theta) = \frac{1}{\pi}$ for $0 \le z \le 1$, $0 \le \theta \le \pi$



For what pairs (Z, Θ) is there an intersection?

This intersection occurs if $Z \le \frac{1}{2}\sin\Theta$ or $Z \ge 1 - \frac{1}{2}\sin\Theta$ (*)

$p(\text{intersection}) = \iint_B f_{Z,\Theta}(z, \theta)\, dz\, d\theta$

$B = \{(z,\theta) \in [0,1] \times [0,\pi] : (*) \text{ holds}\}$

$p(\text{intersection}) = \frac{1}{\pi}\int_0^{\pi} d\theta \left(\int_0^{\frac{1}{2}\sin\theta} dz + \int_{1-\frac{1}{2}\sin\theta}^{1} dz \right)$

$= \frac{1}{\pi}\int_0^{\pi} \sin\theta\, d\theta = \frac{1}{\pi}\left[-\cos\theta\right]_0^{\pi} = \frac{2}{\pi}$

Therefore, Buffon's Needle can be used to estimate π.

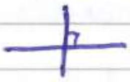
By repeated experimentation one may obtain a numerical estimate for π.

The rate of convergence depends on the variance of the number of intersections in n throws. Mathematically, let $I = \{\text{intersection}\}$

Let 1_I be the indicator function of I.

$E(1_I) = \frac{2}{\pi}$. $Var(1_I) = \frac{2}{\pi}\left(1 - \frac{2}{\pi}\right)$

<u>Buffon's Cross</u>



Throw n times. $Z := \#$ intersections overall

$E\left(\frac{Z}{2}\right) = n\frac{2}{\pi}$, $\frac{1}{n} Var\left(\frac{Z}{2}\right) = \frac{3-\sqrt{2}}{\pi} - \frac{4}{\pi^2}$

10/03/11

Probability (21)

Buffon's cross provides a better estimate as this estimate converges faster.

What happens if we use a needle of length $L < 1$?

Intersections occur if $Z \leq \frac{L}{2}\sin\theta$ or $Z \geq 1 - \frac{L}{2}\sin\theta$

The probability turns out to be $\frac{2L}{\pi}$

<u>Buffon's Noodle</u>

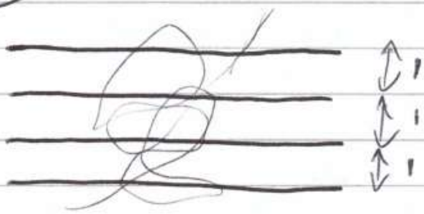
What can we say about dropping a flexible 'needle'?

3

12/03/10

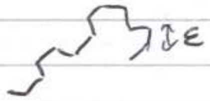
Probability (22)

Buffon's Noodle



Take a ruled plane

Drop a noodle of length L onto the grid.



$I :=$ # intersections with the lines

$$\simeq \sum_{\substack{\text{segments}\\ \text{length } \epsilon}} 1_{\{\text{segment crosses a line}\}}$$

$$E(I) = \sum_{\text{segments}} \frac{2\epsilon}{\pi} \underset{\epsilon \downarrow 0}{\simeq} \frac{2L}{\pi}$$

6.3 Broken Sticks

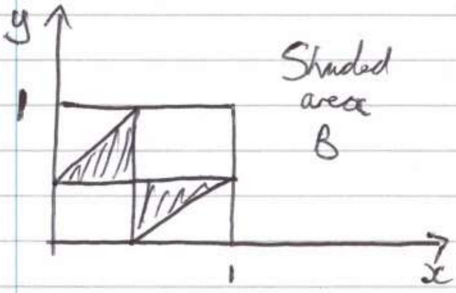
Take a stick of unit length and break it in two places, X, Y, chosen uniformly on $[0, 1]$, independently of each other. What is the probability that the three small sticks can form a triangle?

$U = \min\{X, Y\}$ $\quad V = |X - Y|$ $\quad W = 1 - U - V$

Condition to be able to construct a Triangle:

$U < V + W \quad V < U + W \quad W < U + V$

$\Leftrightarrow U, V, W < \frac{1}{2}$ [what about equality?]



Shaded area B

either $\cancel{X} < \cancel{Y}$

$X < \frac{1}{2}$
$X - Y < \frac{1}{2}$
$1 - Y < \frac{1}{2}$

or $X > Y$

$X > Y$ so $Y < \frac{1}{2}$
$X - Y < \frac{1}{2}$
$1 - X < \frac{1}{2}$

$$p[(X, Y) \in B] = |B| = \frac{1}{4}$$

Note: Generalise to n breaks and the answer is $1 - \frac{n+1}{2^n}$

7. Central Limit Theorem

Consider $X_1, X_2, \dots$ iid. $E(X_1) = \mu$, $Var(X_1) = \sigma^2 \in (0,\infty)$

$$S_n = \sum_{i=1}^{n} X_i$$

Law of Large Numbers $S_n \approx n\mu$

Central Limit Theorem $S_n \approx n\mu + \sqrt{n}(\sigma N)$, N is Normal $(0,1)$

"normalise" $\dfrac{S_n - n\mu}{\sqrt{n\sigma^2}}$

Central Limit Theorem:

Under the above assumptions, $P\left(\dfrac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) \to \Phi(x)$ as $n \to \infty$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$

"$\dfrac{S_n - n\mu}{\sqrt{n\sigma^2}}$" is asymptotically $N(0,1)$.

Definition

The Moment Generating Function (MGF) is defined as, for a random variable X: $M_X(t) = E(e^{tX})$ for any t for which this is finite

Note If X takes values in $\{0,1,2,\dots\}$:

$$M_X(t) = E[(e^t)^X] = G_X(e^t)$$

Examples

a) $Exp(\lambda)$ $\quad M_X(t) = E(e^{tX}) = \int_0^\infty e^{tx} \lambda e^{-\lambda x} dx$

$$M_X(t) = \int_0^\infty \lambda e^{-x(\lambda - t)} dx = \begin{cases} \frac{\lambda}{\lambda - t} & t < \lambda \\ \infty & t \geq \lambda \end{cases}$$

b) $N(0,1)$

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{2\pi}} dx = \int_{-\infty}^{\infty} \exp\left[-\tfrac{1}{2}(x-t)^2\right] \frac{e^{\frac{1}{2}t^2}}{\sqrt{2\pi}} dx$$

$$= e^{\frac{1}{2}t^2}, \quad t \in \mathbb{R}$$

12/03/11 Probability (22)

c) Cauchy Distribution $f(x) = \frac{1}{\pi(1+x^2)}$

$$M(t) = \int_{-\infty}^{\infty} \frac{e^{tx}}{\pi(1+x^2)}\,dx \quad \text{diverges if } t \neq 0$$

The Cauchy distribution has infinite mean and variance.

Properties of the Moment Generating Function

A) Uniqueness Theorem If $M(t) < \infty$ on some neighbourhood of the origin 0, then there is a unique distribution with MGF M.

B) Continuity Theorem

15/03/11

Probability (23)

For a random variable X: the mgf is $M(t) = E(e^{tX})$

A. Uniqueness

If M is the mgf of some distribution, and $M(t) < \infty$ for $|t| < \varepsilon$ and some $\varepsilon > 0$, then this is the unique distribution with mgf M.

B Continuity Theorem

If $Y_1, Y_2, \dots$ are random variables, such that $\forall t$:

$M_{Y_n}(t) \to e^{\frac{1}{2}t^2}$ as $n \to \infty$

then $P(Y_n \leq x) \to \Phi(x)$ as $n \to \infty$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$

C $M_{aX+b}(t) = E\left(e^{t(aX+b)}\right) = e^{tb} E\left(e^{(at)X}\right) = e^{tb} M_X(at)$

D $M_{X+Y}(t) = E\left[e^{t(X+Y)}\right] = E\left[e^{tX} \cdot e^{tY}\right] = M_X(t) M_Y(t)$ if X, Y are independent

E $M_X(t) = E(e^{tX})$

$$= E\left(1 + tX + \frac{t^2X^2}{2!} + \dots\right) \text{ "=" } 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \dots$$

Generating function. $G_{\underline{a}}(t) = \sum_n t^n a_n$

Exponential generating function of (a_n): $\sum_n \frac{t^n a_n}{n!} = E_{\underline{a}}(t)$ — Exponential GF of the moments of X.

The above is ok if $M < \infty$ on some neighbourhood of 0.

Central Limit Theorem

$X_1, X_2, \dots$ iid, mean μ, variance $\sigma^2 \neq 0$

$S_n = \sum_1^n X_i$

$P\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq x\right) \to \Phi(x)$

<u>Proof</u> WLOG, take $\mu = 0$, $\sigma^2 = 1$ (Let $U_i = \frac{X_i - \mu}{\sigma}$)

$$M_{\frac{S_n}{\sqrt{n}}}(t) = M_{S_n}\left(\tfrac{t}{\sqrt{n}}\right) \text{ by C}$$
$$= M_U\left(\tfrac{t}{\sqrt{n}}\right)^n \text{ by D}$$
$$= \left(1 + \tfrac{t}{\sqrt{n}} \cdot 0 + \tfrac{t^2}{2n} \cdot 1 + o\left(\tfrac{t^2}{n}\right)\right)^n \text{ by E}$$
$$= \left(1 + \tfrac{t^2}{2n} + o\left(\tfrac{1}{n}\right)\right)^n \to e^{\frac{1}{2}t^2} \text{ as } n \to \infty$$

$\therefore$ the claim holds, by the continuity theorem

<u>Example</u>

An unknown fraction p of the population vote for unlimited Higher Education fees. It is desired to estimate p by asking a sample of size n. Allow an error in estimate ≤ 0.05.

What n should be used?

Assume each individual votes yes with probability p independently of all others. Let $X_i = \begin{cases} 1 & \text{if the } i^{th} \text{ says yes} \\ 0 & \text{if no} \end{cases}$

$S_n = \sum_1^n X_i$. Use $\overline{X} = \frac{S_n}{n}$ to estimate p.

$$P\left(\left|\tfrac{S_n}{n} - p\right| < 0.005\right)$$
$$= P\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| < \frac{0.005\sqrt{n}}{\sqrt{p(1-p)}}\right) \qquad p(1-p) \leq \tfrac{1}{4}$$
$$\geq P\left(\left|\frac{S_n - np}{\sqrt{np(1-p)}}\right| < 0.005\sqrt{4n}\right)$$

We agree to tolerate mistakes that have probability $\leq 5\%$, say.

as $n \to \infty$, this is approximately

$$\approx \int_{-0.005\sqrt{4n}}^{+0.005\sqrt{4n}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, du = 2\Phi(0.005\sqrt{4n}) - 1 \approx 0.95$$

if $n \approx 40{,}000$

15/03/11

Probability (23)

An application of the Central Limit Theorem

a) $X_i \sim Bern(p)$ $\quad \to \mu = p$

$S_n = \sum_{i=1}^{n} X_i \qquad P\left(S_n - n\mu \leq \alpha\sqrt{np(1-p)}\right) \to \Phi(\alpha)$

$$= \sum_{k : k \leq n\mu + \alpha\sqrt{np(1-p)}} \binom{n}{k} p^k (1-p)^{n-k}$$

$$\to \int_{-\infty}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

b) $p = \frac{1}{2}$

$$\sum_{|k-\frac{1}{2}n| \leq \frac{1}{2}x\sqrt{n}} \binom{n}{k} \approx 2^n \int_{-x}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

c) $X_1, X_2, \ldots$ iid, Poisson (1)

$S_n = \sum_{1}^{n} X_i$ is Poisson (n), $\mu = \sigma^2 = n$

$$P\left(\frac{S_n - n}{\sqrt{n}} \leq \alpha\right) \approx \Phi(\alpha)$$

$$\sum_{k : |k-n| \leq \alpha\sqrt{n}} \frac{n^k}{k!} \approx e^n \int_{-\alpha}^{\alpha} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

Also

$$e^{-n}\left(1 + n + \frac{n^2}{2!} + \ldots + \frac{n^n}{n!}\right) \xrightarrow[n\to\infty]{} \frac{1}{2}$$