RANDOM VARIABLES AND EVENTS

Definition: Random Variable
A random variable is a variable that is random. They are denoted *X*, *Y*, *Z*.

Note
A random variable which is constant will be called a constant random variable.

Example
Roll a "fair" die and let *X* denote the number of dots.
- *X* is a random variable. Its possible values are 1, 2, 3, 4, 5, 6.
- Observing X leads to data or observed value. We will denote a typical observation by x.

Definition: Event
In general, an event is a statement involving random variables. They are denoted *A*, *B*, *C*, *D*.

Notation
{*X* is even} means "the event that *X* is even".

Note
Events either occur or they don't.

Notes
1) Events may be thought of as collections/sets of outcomes.
2) The sure event (denoted *S* or Ω) is the set of all possible outcomes.
3) *S* includes all the outcomes so that any event *A* is made up of outcomes "drawn" from *S*. *A* is a subset of *S* and we write $A \subset S$.
4) The impossible event or empty set (denoted ϕ) is the event which consists of no possible outcomes and hence never occurs.
   - Notice: $\phi \subset A \subset S$.
5) We can talk about any function of *X*, say, $g(X)$. Examples: X^2, $\sin X$, $e^X = \exp(X)$, etc.

Definition: The Indicator Random Variable or Bernoulli Random Variable
If *A* is an event, then we define the indicator random variable or Bernoulli random variable for/of *A* to be

$$I_A = \begin{cases} 1 \text{ if } A \text{ occurs} \\ 0 \text{ otherwise} \end{cases}.$$

Example
Let *X* be a random variable. Measure it an infinite number of times to get $x_1, x_2, x_3, \ldots$.
- The average will be the expected value or expectation of *X*, denoted $E(X)$.
- If *X* denotes marks, then $A = \{\text{pass}\} = \{X \ge 0\}$.
- The indicator random variable might look like: $0,0,1,0,\ldots$.

Definition

The proportion or the relative frequency of 1s will be the probability of A, denoted $P(A)$.

- Notice $P(A) = E(I_A)$.

More On Events and Random Variables

- Events consist of outcomes.
- Every event is a subset of S.

Sometimes the same event has different descriptions, so we might want to show something like:

$$\left.\begin{array}{l} A \subset B \\ B \subset A \end{array}\right\} A = B.$$

How to show $A \subset B$?

- Take an arbitrary element of A.
- Show that it is an element of B, i.e. let $x \in A$ and show $x \in B$.

From our point of view, $A \subset B$ means $A \Rightarrow B$.

Two events are equal (i.e. the same) if $A \Rightarrow B$ and $B \Rightarrow A$.

Random Variables

How to show two random variables X and Y are equal? They are equal if they are equal all the time.

Example

Let A_1 and A_2 be disjoint events (i.e. no outcomes in common, no overlap, etc.). Look at the event $A_1 \cup A_2$ (the event that at least one of the A's occurs).

- The indicator random variable of this new event is $I_{A_1 \cup A_2}$ or $I(A_1 \cup A_2)$. "Clearly", $I_{A_1 \cup A_2} = I_{A_1} + I_{A_2}$.

Notes

1) $\bigcup_{i=1}^{\infty} A_i = A_1 \cup A_2 \cup \cdots$ is the event that at least one of the A's occurs.
2) If the A's are disjoint, then a common notation for $\bigcup_{i=1}^{\infty} A_i = A_1 + A_2 + \cdots$.
3) $\bigcap_{i=1}^{\infty} A_i = A_1 \cap A_2 \cap \cdots = A_1 A_2 \cdots$ is the event that all of the A's occur.

Complements

Definition

The complement of an event A is the event consisting of all outcomes not in A. It is denoted A^c or $\overline{A}$.

Example

Toss a coin n times and set $Y = \text{number of heads}$.

- $A = \{\text{at least one head}\} = \{Y \geq 1\}$.
- $A^c = \{\text{no head}\} = \{\text{all tails}\} = \{Y = 1\}$.

Set $A_1 = \{\text{head on the first toss}\}$, $A_2 = \{\text{head on the second toss}\}$, etc.

- $A_1 A_2 \cdots A_m = \{Y = m\}$.
- $A_1 \cup A_2 \cup \cdots \cup A_m = \{Y \geq 1\}$.
- $(A_1 \cup A_2 \cup \cdots \cup A_m)^c = \{Y = 0\} = A_1{}^c A_2{}^c \cdots A_m{}^c$.

In General

- $\left(\bigcup_i A_i\right)^c = \bigcap_i A_i{}^c$ and $\left(\bigcap_i A_i\right)^c = \bigcup_i A_i{}^c$ – Morgan's Laws.
- $(A^c)^c = A$.

In terms of indicator random variables,

- $I_{A^c} + I_A = 1$.
- $I_{\bigcap_{i=1}^{\infty} A_i} = I_{A_1} \cdots I_{A_m}$.
- $I_{A_1 \cup A_2} = I_{A_1} + I_{A_2} - I_{A_1 \cap A_2}$.

Bernoulli (p) Random Variables

Definition

X is a Bernoulli (p) random variable if X can only take on either 0 or 1 and $P(X = 1) = p$.

Note

1) Let X be Bernoulli (p) (X ~ Bernoulli (p)). Now measure it an infinite number of times. You end up with a list of 0's and 1's. The proportion of 1's will be p and the proportion of 0's will be q.
2) Now,
   - $E(X) = p$ (1st moment of X), $E(X^2) = p$ (2nd moment of X), $E(X^n) = p$.
   - $E(2^X) = p(2) + q(1)$, $E(3^X) = p(3) + q(1)$.
3) If s is a dummy variable, then $E(s^X) = ps + q$. Call $E(s^X)$ the <u>probability generation function</u> denoted by $G(s)$. This function is in fact defined for any count random variable.

Kolmogorov Axioms/Laws of Probabilty

1) $P(A) \geq 0$; $P(S) = 1$.

2) If A_1 and A_2 are disjoint, then $P(A_1 \cup A_2) = P(A_1) + P(A_2)$.

3) If $A_1, A_2, \ldots$ are disjoint, then $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

Remark

1) If the A_i's are disjoint, we often write $\bigcup_{i=1}^{\infty} A_i = \sum_{i=1}^{\infty} A_i$.

2) $P\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$ and $P\left(\sum_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$ is the countable additivity property of probability.

Some Consequences

1) $P(A) + P(A^c) = 1$.
2) If $A \Rightarrow B$, then $P(A) \le P(B)$.
3) $0 \le P(A) \le 1$.
4) $P(A \cup B) = P(A) + P(B) - P(AB)$.

AXIOMS/LAWS OF EXPECTATION

1) $X \ge 0 \Rightarrow E(X) \ge 0$; $E(1) = 1$.
2) $E(cX) = cE(X)$; $E(X+Y) = E(X) + E(Y)$.
3) If $X_1, X_2, \ldots \ge 0$, then $E\left(\sum_i X_i\right) = \sum_i E(X_i)$.

Note

1) $E(X)$ I also called the <u>mean</u> of X and is often denoted by μ.
2) $E\left[(X-\mu)^2\right] \ge 0$ is called the <u>variance</u> of X, denoted $\text{Var}(X)$ or σ^2.
3) $E(X_1 + X_2 + X_3) = E[(X_1 + X_2) + X_3] = E(X_1 + X_2) + E(X_3) = E(X_1) + E(X_2) + E(X_3)$. Notice that using induction, we would have $E(X_1 + \cdots + X_m) = E(X_1) + \cdots + E(X_m), \forall m \ge 2$.
4) $\text{Var}(X) = E\left[(X-\mu)^2\right] = E\left(X^2 - 2X\mu + \mu^2\right) = E\left(X^2\right) - 2\mu E(X) + \mu^2 = E\left(X^2\right) - [E(X)]^2$.
5) $\sqrt{\text{Var}(X)}$ is the <u>standard deviation</u> and is denoted by $SD(X)$ or σ.

Some Consequences

1) If $Y \ge X$, then $E(Y) \ge E(X)$.
2) Boole's Inequality: $P\left(\bigcup_i A_i\right) \le \sum_i P(A_i)$.

PROBABILITY AND COMBINATORICS

Example

Toss a fair coin and let X be the number of heads. Then X is either 0 or 1. Then $P(X=0)=P(X=1)=\frac{1}{2}$.

Proof: $S=\{X=0, X=1\}=\{X=0\}\cup\{X=1\}$, so $P(S)=P(X=0)+P(X=1)$. But $P(S)=1$. So $P(X=0)+P(X=1)=1$. Now $P(X=0)=P(X=1)$ since the coin is fair, so $P(X=0)=P(X=1)=\frac{1}{2}$.

In General

Suppose there is a finite number of outcomes, say N of them. Assume outcomes are equally likely. Let A be the event consisting with n of those outcomes, then $P(A)=\frac{n}{N}$.

Basic Principles in Counting

- Suppose there are n distinct items, the number of arrangements is $n\times(n-1)\times\cdots\times 2\times 1=n!$.
- Suppose there are n items that can be categorized into two types, the number of arrangements is $\frac{n!}{n_1!n_2!}=\binom{n}{n_1}=\binom{n}{n_2}$ (binomial coefficient).
- Suppose there are n items that can be categorized into k types, the number of arrangements is $\frac{n!}{n_1!n_2!\cdots n_k!}=\binom{n}{n_1,\ldots,n_k}$ (multinomial coefficient).

Binomial Theorem

$$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}.$$

Note

$2^n=\sum_{k=0}^{n}\binom{n}{k}$ is the number of subsets of n elements.

THE UNIFORM

Definition

Let B be a finite set of numbers. X is <u>uniform</u> on B if $P(X=x)$ is the same for all $x\in B$. We write $X\sim\text{uniform}(B)$.

For any "nice" B, we can say $X\sim\text{uniform}(B)$ if the probability is "evenly smeared" out over B.

Example

Let $B=[0,1]$ and let $X \sim \text{uniform}([0,1])$. If r is a ration number, then $P(X=r)$ doesn't change for $r \in [0,1]$. Since **Q** is countable, the number of rational numbers in $[0,1]$ is countable. So $P(X \in \mathbf{Q} \cap [0,1]) \le 1 = \sum_{r\in[0,1]} P(X=r)$, and so $P(X=r)=0$ for all rational numbers in $[0,1]$, and hence for all numbers in $[0,1]$.
Now, let $B_1,\ldots,B_n$ partition $[0,1]$. Then $P(X\in B_1)=P(X\in B_2)=\cdots=P(X\in B_n)$. Since $B_1,\ldots,B_n$ partition $[0,1]$, $P(X\in B_1)+\cdots+P(X\in B_n)=1 \Rightarrow P(X\in B_k)=\frac{1}{n}$. This is identical to the finite space when outcomes are equally likely.

Note

Let $B \subset \mathbf{R}$ and define $g(x)=\begin{cases}1, x\in B\\ 0, x\notin B\end{cases}$. Then $g(X)$ is a random variable. In fact, $g(X)=\begin{cases}1, X\in B\\ 0, x\notin B\end{cases}=I_B(x)$ is the indicator random variable that $x\in B$, i.e. $g(X)=I_{\{X\in B\}}$.
Starting with $U \sim \text{uniform}([0,1])$, we get a rich family of random variables by taking functions of U.

RESULTS USEFUL FOR CALCULATING PROBABILITIES AND EXPECTATIONS

Expectation

If X is a random variable with possible values $a_1,\ldots,a_k$. Then $E(X)=a_1P(X=a_1)+\cdots+a_kP(X=a_k)$ and $E(X^2)=a_1^2P(X=a_1)+\cdots+a_k^2P(X=a_k)$.

Definitions

1) X is called a discrete random variable if there exists a function $f(x)\ge 0$ such that $E[g(X)]=\sum_{\text{all } x} g(x)f(x), \forall g$.
2) X is called a continuous random variable if there exists a function $f(x)\ge 0$ such that $E[g(X)]=\int g(x)f(x), \forall g$.

Note

Why did we define the discrete random variable the way we did?
Imagine a discrete random variable X with possible values $x_1,x_2,\ldots$ and suppose $g(x)\ge 0, \forall x$. Then $g(X)=\sum g(x_i)I_{\{X=x_i\}}$. So $E[g(X)]=E\left[\sum g(x_i)I_{\{X=x_i\}}\right]=\sum g(x_i)E\left(I_{\{X=x_i\}}\right)=\sum g(x_i)P(X=x_i)$. So if we let $f(x)=P(X=x_i)$, then we get $g(X)=\sum g(x_i)P(X=x_i)=\sum_{\text{all } x} g(x)f(x)$.

Note

1) $f(x)=P(X=x)$.
2) $f(x)=0$ except at a countable number of values.
3) $E(1)=1=\sum_{\text{all } x} f(x)$.

Note

1) $\left.\begin{array}{r} f(x) \geq 0 \\ \sum_{\text{all } x} f(x) = 1 \end{array}\right\}$ is the condition for a function to be a probability function. So a discrete random variable has $P(X \in B) = \sum_{x \in B} f(x)$.

2) $\left.\begin{array}{r} f(x) \geq 0 \\ \int_{-\infty}^{\infty} f(x) dx = 1 \end{array}\right\}$ is the condition for a function to be a probability density function. So continuous random variable has $P(X \in B) = \int_B f(x) dx$.

Definitions

1) $m(t) = E\left(e^{tX}\right)$ is the moment generating function.
2) $c(t) = E\left(e^{itX}\right)$ is the characteristic function.

Note

$G(s) = E\left(s^X\right) = \sum_{k=0}^{\infty} P(X = k)s^k = P(X = 0) + P(X = 1)s + P(X = 2)s^2 + \cdots$ is a probability generating function for counting random variables only.

Definition

The probability distribution of X is the collection of all probabilities of X-events, i.e. the collection of all expectations of functions of X.

Note

If you know $m(t)$, $c(t)$, or $G(s)$, then you know the distribution. So they are the "representatives" for the distribution.

Definition

The function $F(x) = P(X \leq x)$ is called the (cumulative) distribution function.

Note

If you know F, you know the distribution.

Example

Suppose X has distribution function F and let $a < b$. Then $P(a < X \leq b) = P(X \in (a, b]) = F(b) - F(a)$.
Proof: $\{-\infty < x \leq b\} = \{-\infty < x \leq a\} \cup \{a < x \leq b\} \Rightarrow P(\{-\infty < x \leq b\}) = P(\{-\infty < x \leq a\}) + P(\{a < x \leq b\}) \Rightarrow$ $F(b) = F(a) + P(a < X \leq b) \Rightarrow P(a < X \leq b) = F(b) - F(a)$.

MONOTONE SEQUENCES

- $\left\{-\infty < X \le b - \frac{1}{n}\right\}, n = 1,2,\ldots$ gets bigger as n gets bigger (an increasing set). So

  $\lim_{n\to\infty}\left\{-\infty < X \le b - \frac{1}{n}\right\} = \{-\infty < X < b\}$, the union of all the sets.
- $\left\{-\infty < X \le b + \frac{1}{n}\right\}, n = 1,2,\ldots$ gets smaller as n gets bigger (an decreasing set). So

  $\lim_{n\to\infty}\left\{-\infty < X \le b + \frac{1}{n}\right\} = \{-\infty < X \le b\}$, the intersection of all the sets.

Definitions

If $A_1 \subset A_2 \subset \cdots$, then $\lim_{n\to\infty} A_n = \bigcup_{k=1}^{\infty} A_k$ and write $A_n \uparrow A$ where $A = \lim_{n\to\infty} A_n$.

If $A_1 \supset A_2 \supset \cdots$, then $\lim_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} A_k$ and write $A_n \downarrow A$ where $A = \lim_{n\to\infty} A_n$.

Examples

1) $\left\{-\infty < X \le b - \frac{1}{n}\right\} \uparrow \{-\infty < X < b\}$.
2) $\left\{-\infty < X \le b + \frac{1}{n}\right\} \downarrow \{-\infty < X \le b\}$.

Remarks

1) If $A_n \uparrow A$, it is called an <u>increasing sequence of events</u>.
2) If $A_n \downarrow A$, it is called a <u>decreasing sequence of events</u>.
3) We only talk about the limit of <u>monotone sequence of events</u> (increasing or decreasing).
4) If $A_n \uparrow A$, then $A_1 \cup A_2 \cup \cdots \cup A_N = A_N$. If $A_n \downarrow A$, then $A_1 \cap A_2 \cap \cdots \cap A_N = A_N$.
5) $A_n \uparrow A \Leftrightarrow A_n{}^c \downarrow A^c$, $A_n \downarrow A \Leftrightarrow A_n{}^c \uparrow A^c$.

Theorem: Monotone Convergence Theorem

Let $0 \le X_1 \le X_2 \le \cdots$ and suppose $\lim_{n\to\infty} X_n = X$. Then $\lim_{n\to\infty} E(X_n) = E(X)$.

Proof: $X_n = X_1 + (X_2 - X_1) + \cdots + (X_n - X_{n-1}) \Rightarrow \lim_{n\to\infty} X_n = X_1 + (X_2 - X_1) + \cdots \Rightarrow$

$E\left[\lim_{n\to\infty} X_n\right] = E[X_1 + (X_2 - X_1) + \cdots] = E(X_1) + E(X_2 - X_1) + \cdots \Rightarrow$

$E\left[\lim_{n\to\infty} X_n\right] = \lim_{n\to\infty} E(X_1) + E(X_2 - X_1) + \cdots + E(X_n - X_{n-1}) = \lim_{n\to\infty} E(X_n)$. So $\lim_{n\to\infty} E(X_n) = E(X)$.

Lemma

$A_n \uparrow A \Leftrightarrow I_{A_n} \uparrow I_A$ and $A_n \downarrow A \Leftrightarrow I_{A_n} \downarrow I_A$.

Note

$X_n \uparrow X$ means $X_1 \le X_2 \le \cdots$ and $\lim_{n\to\infty} X_n = X$.

Corollary: Continuity Property of Probability

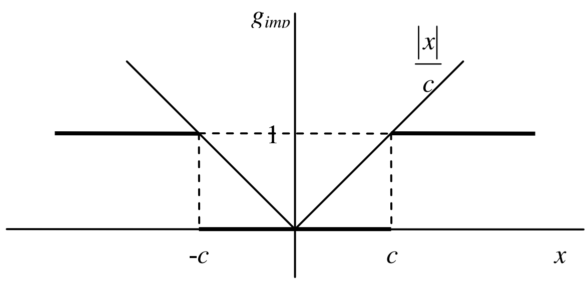
$A_n \uparrow A \Rightarrow \lim_{n\to\infty} P(A_n) = P(A)$ and $A_n \downarrow A \Rightarrow \lim_{n\to\infty} P(A_n) = P(A)$.

Proof:

- $A_n \uparrow A \Rightarrow I_{A_n} \uparrow I_A \Rightarrow \lim_{n\to\infty} E(I_{A_n}) = E(I_A) \Rightarrow \lim_{n\to\infty} P(A_n) = P(A)$.
- $A_n \downarrow A \Rightarrow {A_n}^c \uparrow A^c \Rightarrow \lim_{n\to\infty} P({A_n}^c) = P(A^c) \Rightarrow \lim_{n\to\infty} 1 - P(A_n) = 1 - P(A) \Rightarrow \lim_{n\to\infty} P(A_n) = P(A)$.

MARKOV'S INEQUALITY

Let $c > 0$ and define g_{imp} as in the picture.



Then $g_{imp}(x) = I_{\{|x| \ge c\}}$.

Markov's Inequality

If $c > 0$ and X is a random variable, then $P(|X| > c) \le \frac{E(|X|)}{c}$.

Proof: $g_{imp}(x) = \begin{cases} 1, |x| \ge c \\ 0, \text{otherwise} \end{cases}$, so $g_{imp}(x) \le \frac{|x|}{c}, \forall x \Rightarrow E(g_{imp}(x)) \le \frac{E(|X|)}{c} \Rightarrow E(I_{\{|x| \ge c\}}) \le \frac{E(|X|)}{c} \Rightarrow$ $P(|X| \ge c) \le \frac{E(|X|)}{c}$.

UPDATING PROBABILITY AND INDEPENDENCE

Example

Roll a fair die and let X be the number of dot. Let $A_1=\{X\geq 4\}$ and $A_2=\{X \text{ is even}\}$. Then: $P(A_2)=\frac{3}{6}$, $P(A_1)=\frac{3}{6}$, $P(A_1A_2)=\frac{2}{6}$, $\frac{P(A_1A_2)}{P(A_1)}=\frac{2}{3}$. If you know A_1 has occurred then you update then probability of A_2 to 2/3.

Definition

The <u>conditional probability</u> of A_2 given A_1 is $P(A_2 \mid A_1)=\frac{P(A_1A_2)}{P(A_1)}$.

Notes

1) For fixed A, $P(A \mid A_1)$ satisfies the laws of probability.
2) $P(A_1A_2)=P(A_1)P(A_2 \mid A_1)$.

Definition

A_1 and A_2 are independent events if $P(A_1A_2)=P(A_1)P(A_2)$.

Definition

$A_1, A_2, \ldots$ are independent if for any finite or infinite collection $A_{i_1}, A_{i_2}, \ldots$, $P(A_{i_1}A_{i_2}\cdots)=P(A_{i_1})P(A_{i_2})\cdots$.

Definition

X_1 and X_2 are independent if X_1-events and X_2-events are.

Fact

X_1 and X_2 are independent if and only if for any finite collection $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$,
$E[g(X_{i_1})h(X_{i_2})\cdots w(X_{i_k})]=E[g(X_{i_1})]E[h(X_{i_2})]E[w(X_{i_k})]$, "$\forall$" $g, h, \ldots, w$.

BINOMIAL DISTRIBUTION

Let $X_1, X_2, \ldots$ be independent random variables each with the same probability distribution. We say they are independent identically distributed random variables (idd).

Example

Let $X_1, X_2, \ldots$ be idd Bernoulli(p) random variables. So they are independent, $\begin{cases} P(X_1=1)=p \\ P(X_1=0)=q=1-p \end{cases}$, and $p+q=1$. Let $Y=X_1+\cdots+X_n$ be a binomial(n, p) random variable.

Then $P(Y=k)=\binom{n}{k}p^k q^{n-k}, k=0,1,\ldots$. So the probability function of Y is

$$f(y)=P(Y=y)=\begin{cases}\binom{n}{y}p^y q^{n-y}, n=0,1,\ldots \\ 0, \text{otherwise}\end{cases}.$$

Expectation and Variance of the Binomial Distribution

Let $Y \sim \text{binomial}(n,p)$. Then:

- $E(Y)=\mu=np$.
- $\text{Var}(Y)=\sigma^2=npq$.

Some Problems

Chebyshev's Inequality

Let X be a random variable with mean μ and variance σ^2, and suppose $k \geq 0$. Show $P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$.

Solution: $P(|X-\mu| \geq k\sigma)=P((X-\mu)^2 \geq (k\sigma)^2) \leq \frac{E[(X-\mu)^2]}{k^2\sigma^2}=\frac{\sigma^2}{k^2\sigma^2}\frac{1}{k^2}$.

Variance of Independent Random Variables

Let X_1, X_2 be independent with mean 0. Show that $\text{Var}(X_1+X_2)=\text{Var}(X_1)+\text{Var}(X_2)$.
Solution:
$\text{Var}(X_1+X_2)=E[(X_1+X_2-0)^2]=E(X_1{}^2)+E(X_2{}^2)+2E(X_1X_2)=E(X_1{}^2)+E(X_2{}^2)+2E(X_1)(X_2)$. Since X_1, X_2 independent, $E(X_1X_2)=E(X_1)E(X_2)$. Since the mean is 0, $E(X_1)E(X_2)=0$. So $\text{Var}(X_1+X_2)=\text{Var}(X_1)+\text{Var}(X_2)$.
Note: This can be extended to n by induction. It can also be extended to non-zero means.

Variance

Let X have mean μ and variance σ^2. Then:

- $E(aX+b)=a\mu+b$.
- $\text{Var}(aX+b)=E[((aX+b)-(a\mu+b))^2]=a^2E[(X-\mu^2)]=a^2\sigma^2$.

Geometric Random Variable

Toss a coin with $P(\{H\})=p$ and $q=1-p$ until we obtain a head. Let Y be the number of tosses. Calculate the pgf of Y.

- We have the pgf $G(s)=E(s^Y)=\sum_{\text{all } y} s^Y f(y)=\sum_{k=1}^{\infty} s^k P(Y=k)$.
- $P(Y=k)=P(\{k-1 \text{ tails, then head}\})=q^{k-1}p, k=1,2,\ldots$.

- So $G(s)=\sum_{k=1}^{\infty} s^k q^{k-1} p = \frac{p}{q}\sum_{k=1}^{\infty}(sq)^k = \frac{p}{q}\frac{sq}{1-sq}, |sq|<1 \Rightarrow G(s)=\frac{sp}{1-sq}, |s|<\frac{1}{q}$.

Note: Y is called a geometric random variable.

Poisson Random Variable

Suppose $Y \sim \text{binomial}(n,p)$ and we set $\lambda = np$. Think of n large, p small, but λ reasonable. The pgf of Y is $G(s)=(q+ps)^n = (1+p(s-1))^n = \left(1+\frac{\lambda(s+1)}{n}\right)^n \approx e^{\lambda(s-1)}$. What kind of random variable ahs this kind of pgf?

$\text{Poisson}(\lambda)$. The probabilities can be obtained by noting $e^{\lambda(s-1)} = e^{-\lambda}e^{\lambda s} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}s^k$, so $P(\{\text{Poisson}(\lambda)\,\text{rv} = k\}) = \frac{e^{-\lambda}\lambda^k}{k!}, k=0,1,2,\ldots$.

More On Conditional Probability and Independence

Note

X and Y are independent $\Leftrightarrow$ $P(\{X\in B_1\}\cap\{Y\in B_2\}) = P(\{X\in B_1\})P(\{Y\in B_2\}), \text{"}\forall\text{"}\, B_1, B_2 \Leftrightarrow$ $E[g(X)h(Y)] = E[g(X)]E[h(Y)], \text{"}\forall\text{"}\, g, h$.

Note

For fixed A, $P(B \mid A) = \frac{P(AB)}{P(A)}$ satisfies the axioms of probability.

Partition

Let $A_1, A_2, \ldots$ partition S, i.e. they are disjoint. Then, clearly, $P(B)=P(A_1B)+P(A_2B)+\cdots$.
But since $P(AB)=P(B\mid A)P(A)$, $P(B)=P(B\mid A_1)P(A_1)+P(B\mid A_2)P(A_2)+\cdots$.

Correlation

Definition

If X and Y are random variables, then $\rho = \rho_{XY} = \frac{E[(X-E(X))(Y-E(Y))]}{SD(X)SD(Y)}$ is the correlation between X and Y.

Facts

1) $\rho^2 \le 1$.
2) $\text{cov}(X,Y) = E[(X-E(X))(Y-E(Y))]$ is the covariance of X and Y. Notice:
   - $\text{cov}(X,Y) = E(XY) - E(X)E(Y)$.
   - $\text{cov}(X,X) = \text{var}(X)$.

- $\operatorname{cov}(aX, bY) = ab\operatorname{cov}(X, Y)$ and $\operatorname{cov}(X + c, Y + d) = \operatorname{cov}(X, Y)$.

Definition

Suppose *X*, *Y* are independent. Then $E(XY) = E(X)E(Y) \Rightarrow \operatorname{cov}(X, Y) = 0$. We say *X*, *Y* are uncorrelated.

NORMAL DISTRIBUTION

The Probability Density Function

Let $Z \sim N(0,1)$. The pdf of *Z* is $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -\infty < z < \infty$.

The Moment Generating Function

The mgf of *Z* is $M_Z(t) = E\left(e^{tZ}\right) = \int_{-\infty}^{\infty} e^{tz} f(z)dz = \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{\infty} e^{\frac{t^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-t)^2}{2}} dz = e^{\frac{t^2}{2}}$.

The mgf of $Y = \mu + \sigma Z, \sigma > 0$ is $M_Y(t) = E\left[e^{t(\mu + \sigma Z)}\right] = e^{\mu t} E\left[e^{(\sigma t)Z}\right] = e^{\mu t} e^{\sigma^2 \frac{t^2}{2}}$.

MORE ON DISCREET RANDOM VARIABLES

Let $X_1, X_2, \ldots$ be Bernoulli iid rv's, taking on 1 with probability *p*, and 0 with probability $q = 1 - p$.
$Y = X_1 + X_2 + \cdots + X_n$ = (number of 1's in the first *n* trials) is a binomial(*n*, *p*) rv.

Y = (number of "tosses" until a head is obtain) is geometric.
Y = (number of "tosses" until the r^{th} head is obtain) = (# of "tosses" until the first head) + … + (# of "tosses" until the r^{th} head) is negative binomial.

Geometric Random Variables

Let $Y \sim \operatorname{geo}(p)$. So $P(Y = k) = q^{k-1} p, k = 1,2,\ldots$.

The pgf $G(s) = E\left(s^Y\right) = \sum_{k=1}^{\infty} s^k q^{k-1} p = \frac{p}{q} \sum_{k=1}^{\infty} (qs)^k = \frac{p}{q}\left(\frac{qs}{1-qs}\right) = \frac{ps}{1-qs}, |s| < \frac{1}{q}$.

So $G'(1) = \frac{1}{p} \Rightarrow E(Y) = \frac{1}{p}$.

Negative Binomial Random Variables

If $Y_1, \ldots, Y_r$ are iid geometric(*p*), then $Y = Y_1 + \cdots + Y_r$ is a negative binomial. Also set $W = Y - r$.

The possible values of *Y* are $r, r+1, \ldots$, and the possible values of *W* are $0,1,2,\ldots$.

So $E(Y)=E(Y_1)+\cdots E(Y_r)=\frac{r}{p}$.

The pgf $G_Y(s)=E\left(s^Y\right)=E\left(s^{Y_1+\cdots+Y_r}\right)=E\left(s^{Y_1}\cdots s^{Y_r}\right)=E\left(s^{Y_1}\right)\cdots E\left(s^{Y_r}\right)$ since $Y_1,\ldots,Y_r$ are independent, so $G_Y(s)=\left(\frac{ps}{1-qs}\right)^r, |s|<\frac{1}{q}$.

The pfg $G_W(s)=E\left(s^W\right)=E\left(s^{Y-r}\right)=\frac{1}{s^r}E\left(s^Y\right)=\frac{p^r}{(1-qs)^r}, |s|<\frac{1}{q}$.

Now, $P(W=i)=P(Y=i+r), i=0,1,2,\ldots$, so getting the *W*-probabilities gets you the *Y*-probabilities. Note $G_X(s)=P(X=0)+P(X=0)s+P(X=0)s^2+\cdots$, so $G_X(0)=P(X=0)$, $G_X{}^{(1)}(0)=P(X=1)$, $G_X{}^{(2)}(0)=2!\,P(X=2)$, $G_X{}^{(3)}(0)=3!\,P(X=3),\ldots,$ $G_X{}^{(k)}(0)=k!\,P(X=k)$. So $P(X=k)=\frac{G_X{}^{(k)}(0)}{k!}, k=0,1,\ldots$.

So $P(W=k)=\frac{\frac{d^k}{ds^k}\left[\frac{p^r}{(1-qs)^r}\right]_{s=0}}{k!}$. Notice that $\frac{d}{ds}\left[\frac{1}{(1-qs)^r}\right]=\frac{qr}{(1-qs)^{r+1}}$, $\frac{d^2}{ds^2}\left[\frac{1}{(1-qs)^r}\right]=\frac{q^2r(r+1)}{(1-qs)^{r+2}},\ldots,$ $\frac{d^k}{ds^k}\left[\frac{1}{(1-qs)^r}\right]=\frac{q^k r(r+1)\cdots(r+(k-1))}{(1-qs)^{r+k}}$. Therefore, $P(W=k)=\frac{p^r q^k r(r+1)\cdots(r+k-1)}{k!}=P(Y=r+k), k=0,1,2,\ldots$.

Obtaining the probability directly, $P(Y=i)=\binom{i-1}{r-1}p^{r-1}q^{(i-1)-(r-1)}p=\binom{i-1}{r-1}p^r q^{i-r}, i=r,r+1,\ldots$.

Poisson Random Variables

Let $Y\sim\text{Poisson}(\lambda)$, so $P(Y=k)=\frac{e^{-\lambda}\lambda^k}{k!}, k=0,1,\ldots,\lambda>0$.

We know the pgf is given by $G(s)=e^{\lambda(s-1)}$.

Note: Since $G(s)=E\left(s^Y\right)$, $m(t)=E\left(e^{tY}\right)=E\left[\left(e^t\right)^Y\right]=G\left(e^t\right)$. So for $Y\sim\text{Poisson}(\lambda)$, $m(t)=e^{\lambda\left(e^t-1\right)}$.

Recall: $m^{(k)}(0)=E\left(Y^k\right)$. In this case, for $Y\sim\text{Poisson}(\lambda)$, $m'(t)=e^{\lambda\left(e^t-1\right)}\lambda e^t$ and $m''(t)=e^{\lambda\left(e^t-1\right)}\lambda^2e^{2t}+e^{\lambda\left(e^t-1\right)}\lambda e^t$. So $\left.\begin{array}{l} m'(0)=E(Y)=\lambda \\ m''(0)=E\left(Y^2\right)=\lambda^2+\lambda \end{array}\right\}\Rightarrow \operatorname{var}(Y)=E\left(Y^2\right)-(E(Y))^2=\lambda$.

Example

Let $X_1\sim\text{Poisson}(\lambda_1)$ be independent of $X_2\sim\text{Poisson}(\lambda_2)$. Show that $Y=X_1+X_2\sim\text{Poisson}(\lambda_1+\lambda_2)$.

- The pgf of Y is $G_Y(s) = G_{X_1}(s) G_{X_2}(s) = e^{\lambda_1(s-1)} e^{\lambda_2(s-1)} = e^{(\lambda_1 - \lambda_2)(s-1)}$, which is the pgf of a Poisson$(\lambda_1 + \lambda_2)$ rv. Since pgf determines distributions, $Y \sim \text{Poisson}(\lambda_1 + \lambda_2)$.

Note: The same argument yields $X_1 + \cdots + X_n \sim \text{Poisson}(\lambda_1 + \cdots + \lambda_n)$ if $X_i \sim \text{Poisson}(\lambda_i)$ and X's independent.

Hyper-Geometric Probabilities

Example

Suppose there are n_1 b's and m_1 w's, so there are $N = n_1 + m_1$ items. Select n items "at random". Let Y be the number of b's selected.

With replacement, $Y \sim \text{Binomial}\left(n, \frac{n}{N}\right)$, and so $P(Y = k) = \binom{n}{k} p^k q^{n-k}, k = 0, 1, \ldots, n$.

Without replacement, $P(Y = k) = \dfrac{\binom{n_1}{k}\binom{m_1}{n-k}}{\binom{N}{n}}$ is hyper-geometric.

Distributions

Definition

If X is a random variable, the distribution of X refers to the collection of all X-probabilities, or the expected value of the real functions of X.

Notation

Suppose X and Y have the same distribution. We write $X \sim Y$ or $X \overset{d}{\sim} Y$ or $X \overset{d}{=} Y$.

Facts

The following are equivalent:

- $X \overset{d}{=} Y$
- $F_X(u) = F_Y(u), \forall u$
- $c_X(t) = c_Y(t), \forall t$
- $f_X(u) = f_Y(u), \forall u$ (discreet case), $f_X(u) = f_Y(u), \text{"}\forall\text{"} u$ (continuous case)
- $G_X(s) = G_Y(s), \forall s$ (counting case)
- $m_X(t) = m_Y(t), t > 0, t < 0$ if the mgf exists.

Normal Distribution

Definition

$Z \sim N(0,1)$ is a standard normal random variable if its probability density function is $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \forall z$.

Note

The moment generating function of Z is $m(t) = E(e^{tZ}) = e^{-\frac{t^2}{2}}$. So, $\left.\begin{matrix} E(Z) = m'(0) = 0 \\ E(Z^2) = m''(0) = 1 \end{matrix}\right\} \Rightarrow \mathrm{var}(Z) = 1$.

Definition

$Y \sim N(\mu, \sigma^2)$ if $Y = \mu + \sigma Z$ where $Z \sim N(0,1)$.

Notes

1) $E(Y) = E(\mu + \sigma Z) = \mu + \sigma E(Z) = \mu$.
2) $\mathrm{var}(Y) = \mathrm{var}(\mu + \sigma Z) = \sigma^2\ \mathrm{var}(Z) = \sigma^2$.

Probability Density Function of a Normal Random Variable

$Y = \mu + \sigma Z$, and we know μ, σ, pdf of Z. Notice that any Y-probability can be written as a Z-probability (ex: $P(a < Y < b) = P(a < \mu + \sigma Z < b) P\left(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}\right)$), so can get the distribution of Y.

Now, $F_Y(y) = P(Y \le y) = P\left(Z \le \frac{y-\mu}{\sigma}\right) = F_Z\left(\frac{y-\mu}{\sigma}\right) = \Phi\left(\frac{y-\mu}{\sigma}\right)$. So

$f_Y(y) = F_Y'(y) = F_z'\left(\frac{y-\mu}{\sigma}\right) = f_z\left(\frac{y-\mu}{\sigma}\right) \cdot \frac{1}{\sigma}$. Hence $f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}, \forall y$.

Moment Generating Function of a Normal Random Variable

$m_Y(t) = E(e^{tY}) = E(e^{t(\mu + \sigma Z)}) = e^{\mu t} E(e^{\sigma t Z}) = e^{\mu t} e^{-\frac{\sigma^2 t^2}{2}}$.

Example

Let $Y_1 \sim N(\mu_1, \sigma_1^2)$ be independent of $Y_2 \sim N(\mu_2, \sigma_2^2)$. Set $W = Y_1 + Y_2$. Then

$m_W(t) = E(e^{tW}) = m_{Y_1}(t) m_{Y_2}(t) = \left(e^{\mu_1 t} e^{-\frac{\sigma_1^2 t^2}{2}}\right)\left(e^{\mu_2 t} e^{-\frac{\sigma_2^2 t^2}{2}}\right) = e^{(\mu_1 + \mu_2)t} e^{-\frac{(\sigma_1^2 + \sigma_2^2)t^2}{2}}$. So

$W \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

MORE ON GENERAL DISTRIBUTION FUNCTIONS

Definition

$F(x) = P(X \le x)$.

Properties

1) $x_1 < x_2 \Rightarrow F(x_1) \le F(x_2)$ (increasing function).
   Proof: $x_1 < x_2 \Rightarrow \{X \le x_1\} \subset \{X \le x_2\} \Rightarrow P(X \le x_1) \le P(X \le x_2) \Rightarrow F(x_1) \le F(x_2)$.

2) $F(-\infty) = \lim_{x \to -\infty} F(x) = 0$ and $F(\infty) = \lim_{x \to \infty} F(x) = 1$.

3) F is right continuous.
   Proof: Let x_0 be any point. Then as $x \downarrow x_0$, we have $\{X \le x\} \downarrow \{X \le x_0\}$. So,
   $\lim_{x \downarrow x_0} F(x) = \lim_{x \downarrow x_0} P(\{X \le x\}) = P\left(\lim_{x \downarrow x_0} \{X \le x\}\right) = P(\{X \le x_0\}) = F(x_0)$. Therefore F is right continuous at x_0, and since x_0 is arbitrary, F is right continuous.

Note

Note that $x \uparrow x_0 \Rightarrow \{X \le x\} \uparrow \{X < x_0\}$. So $P(x = x_0) = P(X \le x_0) - P(X < x_0) = F(x_0) - \lim_{x \uparrow x_0} F(x)$ is the jump at x_0 (in the discreet case).

Random Vectors

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$ is a random vector; $\mathbf{x} \in \mathbf{R}^n$ is a typical observation.

Distribution

Let $g : \mathbf{R}^n \to \mathbf{R}$ be a real-valued function, and look at $E[g(\mathbf{X})]$. The distribution of $\mathbf{X}$ is the collection of all these (or the collection of $P(\mathbf{X} \in B), B \subset \mathbf{R}^n$.

Theorem

The distribution of $\mathbf{X}$ is determined by $E\left(e^{i\mathbf{t}\cdot\mathbf{X}}\right), \forall \mathbf{t}$.

Notes

1) $i = \sqrt{-1}$.
2) $e^{i\theta} = \cos\theta + i\sin\theta$.
3) $c(\mathbf{t}) = E\left(e^{i\mathbf{t}\cdot\mathbf{X}}\right), \forall \mathbf{t}$ is the characteristic function of $\mathbf{X}$ (or the joint characteristic function of $X_1, \ldots, X_n$).
   Notice that $e^{i\mathbf{t}\cdot\mathbf{X}} = e^{it_1 X_1} \cdots e^{it_n X_n}$.

4) $m(\mathbf{t}) = E\left(e^{\mathbf{t}\cdot\mathbf{X}}\right)$ is the moment generating function (or the joint moment function of $X_1, \ldots, X_n$). By the way, $\frac{\partial}{\partial t_1} m(\mathbf{t}) = E\left(X_1 e^{\mathbf{t}\cdot\mathbf{X}}\right) = E(X_1)$ when $\mathbf{t} = \mathbf{0}$, and $\frac{\partial^2}{\partial t_1 \partial t_2} m(\mathbf{t}) = E\left(X_1 X_2 e^{\mathbf{t}\cdot\mathbf{X}}\right) = E(X_1 X_2)$ when $\mathbf{t} = \mathbf{0}$, so $\operatorname{cov}(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$ can be obtained from the mgf in an "easy" way.
5) If **X** is made up of counting random variables, then $G(s) = E\left(s^{X_1} \cdots s^{X_n}\right)$ is the probability generating function of **X**.
6) $F(\mathbf{x}) = P(\mathbf{X} \le \mathbf{x}) = P(X_1 \le x_1, \ldots, X_n \le x_n)$ is the distribution function.

Continuous vs. Discreet Random Vectors

1) Definition of a continuous random vector: $E[g(\mathbf{X})] = \int g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x}, \text{"}\forall\text{"}\, g$ (the probability density function). Notice that $\int f(\mathbf{x}) d\mathbf{x} = 1$.
2) Definition of a discreet random vector: $E[g(\mathbf{X})] = \sum g(\mathbf{x}) f(\mathbf{x}), \text{"}\forall\text{"}\, g$ (the probability function). Notice that $\sum f(\mathbf{x}) = 1$.

Theorem

$X_1, \ldots, X_n$ are independent if and only if $E\left(e^{i\mathbf{t}\cdot\mathbf{X}}\right) = E\left(e^{it_1 X_1}\right) \cdots E\left(e^{it_n X_n}\right), \forall \mathbf{t}$.

Proposition

In the continuous case, $X_1, \ldots, X_n$ are independent if and only if $f(\mathbf{x}) = f_1(x_1) \cdots f_n(x_n)$. It is the same in the discreet case.

CHANGE OF VARIABLES

Consider random vector **X** and its pdf $f_{\mathbf{X}}(\mathbf{x})$ (known). Let $\mathbf{Y} = h(\mathbf{X})$. We have $E[g(\mathbf{Y})] = \int g(\mathbf{y}) f_{\mathbf{X}}(\mathbf{y}) d\mathbf{y}$.

Also, $E[g(\mathbf{Y})] = E[g(h(\mathbf{X}))] = \int g(h(\mathbf{x})) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} = \int g(\mathbf{y}) f_{\mathbf{X}}\left(h^{-1}(\mathbf{y})\right) \left|\det\left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right)\right| d\mathbf{y}$. So the pdf

$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\left(h^{-1}(\mathbf{y})\right) \left|\det\left(\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right)\right|$.

CONDITIONING

Recall: $P(B \mid A) = \frac{P(AB)}{P(A)} = \frac{E(I_{AB})}{P(A)} = \frac{E(I_A I_B)}{P(A)}$.

Conditional Expectation

The expectation that corresponds to the probability $P(X \mid A)$ is $E(X \mid A) = \dfrac{E(XI_A)}{P(A)}$.

In particular, $E(I_B \mid A) = \dfrac{E(I_B I_A)}{P(A)} = \dfrac{E(I_{AB})}{P(A)} = P(B \mid A)$, which further shows the link between $E(X \mid A)$ and $P(X \mid A)$.

If $A_1, \ldots, A_n$ partition *S*, then $X = XI_{A_1} + \cdots + XI_{A_n} \Rightarrow$ $E(X) = E(XI_{A_1}) + \cdots + E(XI_{A_n}) = E(X \mid A_1)P(A_1) + \cdots + E(X \mid A_n)P(A_n)$. This is the *E*-version of a formula used in Bayes' type problems.

Example

Toss a coin with $P(H) = p \in (0,1)$, and let $Y = \#$ of tosses to first H. Then $Y \sim \text{geometric}(p)$. Now, let $A = \{H \text{ on first toss}\}$. Then $E(Y) = E(Y \mid A)P(A) + E(Y \mid A^c)P(A^c)$. Now let $Y' = Y - 1$. So $E(Y) = 1 \times p + E(Y' + 1 \mid A^c)q = p + \lfloor E(Y' \mid A^c) + 1 \rfloor q = p + q + qE(Y) = 1 + qE(Y) \Rightarrow E(Y) - qE(Y) = 1 \Rightarrow$ $pE(Y) = 1 \Rightarrow E(Y) = \dfrac{1}{p}$.

Conditional Probability (Density) Function

$f(y \mid x) = \dfrac{f(x, y)}{f_X(x)}$ is the "conditional probability function of *Y* given $X = x$". Note that $f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$.

Note

$f(y \mid x) f_X(x) = f(x, y)$ is very useful.

Regression Function

Since $f(y \mid x)$ is a pdf (or pf), we can talk about it's mean. Denoted $r(x) = E(Y \mid X = x) = \int_{-\infty}^{\infty} yf(y \mid x)dx$ is called the "regression function of *Y* on *X*".

Prediction

Suppose *X* is the present and *Y* is the future. We want to predict *Y* using a function of *X*. The "best" function of *X* is called "the conditional expectation of *Y* given *X*" $E(Y \mid X)$ (an rv).
Fact: Under certain conditions, $E(Y \mid X) = r(X)$.

POISSON PROCESS

Example: Point Process

- Throw an infinite points onto $t \geq 0$ so that there is no preference in position.
- Assume density of points is λ.

- This is a Poison Process.

Facts

1) If B is a subset of $\{t \mid t \geq 0\}$ and $N(B)$ be the number of points in B, then $N(B) \sim \text{Poisson}(\lambda|B|)$.
2) If $B_1, B_2, \ldots$ are disjoint, then $N(B_1), N(B_2), \ldots$ are independent.

Note

Recall the distance X to the first point is $\text{exponential}(\lambda)$. In fact the time between points are iid $\text{exponential}(\lambda)$ rv's.

If Y is the time to the r^{th} point, then Y is the sum of r iid $\text{exponential}(\lambda)$ and $f_Y(y) = \dfrac{\lambda^r y^{r-1} e^{-\lambda y}}{(r-1)!}$, $y > 0$, which is the gamma pdf.

Look at $f_Y(y) = \dfrac{\lambda^r y^{r-1} e^{-\lambda y}}{c}$, $y > 0$ where $c = \int_0^\infty \lambda^r y^{r-1} e^{-\lambda y} dy = \int_0^\infty x^{r-1} e^{-x} dx = \Gamma(r)$ is the gamma function. Note that $\left.\begin{aligned}\Gamma(r+1) &= r\Gamma(r)\\ \Gamma(1) &= 1\end{aligned}\right\} \Rightarrow \Gamma(r) = (r-1)!$ if $r \in \mathbf{N}$.

So the pdf $f_Y(y) = \dfrac{\lambda^r y^{r-1} e^{-\lambda y}}{\Gamma(r)}$, $y > 0$ is the gamma pdf (defined for any $r > 0$).

Note

Set $N(t)$ the number of points from $[0, t]$. Then $P(N(t) = k) = \dfrac{e^{-\lambda t} (\lambda t)^k}{k!}$.

Convolution Integral

Let $W = X + Y$. Since $F_W(w) = P(W \leq w) = \int_{-\infty}^{\infty} \int_{-\infty}^{w-x} f_{X,Y}(x, y) dy dx$, so

$f_W(w) = \dfrac{d}{dw} \int_{-\infty}^{\infty} \int_{-\infty}^{w-x} f_{X,Y}(x, y) dy dx = \int_{-\infty}^{\infty} \left[\dfrac{d}{dw} \int_{-\infty}^{w-x} f_{X,Y}(x, y) dy \right] dx = \int_{-\infty}^{\infty} f_{X,Y}(x, w-x) dx$. If X and Y are independent, then $f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w-x) dx$.

Order Statistics

- Let $X_1, \ldots, X_n$ be iid rv's with df F and pdf f. Order them $X_{(1)} < \cdots < X_{(n)}$.
- Also define $F_{(i)}(x)$ the df of $X_{(i)}$, and $f_{(i)}(x)$ the pdf of $X_{(i)}$. Also, $F_{(i)}(x) = P(X_{(i)} > x) = 1 - \overline{F}_{(i)}(x)$.
- Notice that $f_{(i)}(x) = \dfrac{d}{dx} F_{(i)}(x) = -\dfrac{d}{dx} \overline{F}_{(i)}(x)$.

- What is the pdf of $X_{(1)}$? $\overline{F}_{(1)}(x) = P(X_{(1)} > x) = P(X_1 > x, \ldots, X_n > x) = P(X_1 > x) \cdots P(X_n > x) = (\overline{F}(x))^n \Rightarrow$ $f_{(1)}(x) = n(1 - F(x))^{n-1} f(x)$.
- What is the pdf of $X_{(n)}$? $F_{(n)}(x) = P(X_{(n)} \le x) = P(X_1 \le x, \ldots, X_n \le x) = P(X_1 \le x) \cdots P(X_n \le x) = (F(x))^n$ $\Rightarrow f_{(n)}(x) = n(F(x))^{n-1} f(x)$.
- What is the pdf of $X_{(i)}$?

$$f_{(i)}(x) = -\frac{d}{dx} P(X_i > x) = -\frac{d}{dx} \sum_{k=0}^{i-1} \binom{n}{k} (F(x))^k (1 - F(x))^k = \cdots = \frac{n!}{(i-1)!(n-i)!} (F(x))^{i-1} (1 - F(x))^{n-i} f(x).$$

More On Conditioning

Recall

$E[Y \mid X]$ is a function of X used to predict Y. Two important properties are:

1) $E[E[Y \mid X]] = E[Y]$.
2) $X \perp Y$ (independent) if and only if $E[h(Y) \mid X] = E[h(Y)]$.

Example

- Consider a Poisson process of rate λ. Let the X's be iid with mean μ. Let $N(t)$ be the number of X's in $[0, t]$. Note that the X's and $N(t)$ are independent.
- Loot at the compound Poisson process $S(t) = \sum_{i=0}^{N(t)} X_i, X_0 = 0$.
- The expected value $E[S(t)] = E[E[S(t) \mid N(t)]] = E[E(X_1) N(t)] = \mu E[N(t)] = \mu \lambda t$.
- The mgf $m_S(z) = E\left[e^{zS(t)}\right] = E\left[E\left[e^{zS(t)} \mid N(t)\right]\right] = E\left[m_X(z)^{N(t)}\right] = e^{\lambda t (m_X(z) - 1)}$.

More Distributions

Exponential Distribution

- $Z \sim \text{exponential}(1)$ if $f(z) = e^{-z}, z > 0$.
- Let $\theta > 0$ and set $Y = \theta Z$. The pdf is clearly $f(y) = \frac{1}{\theta} e^{-\frac{y}{\theta}}, y > 0$. We write $Y \sim \text{exponential}\left(\frac{1}{\theta}\right)$ (rate $\frac{1}{\theta}$), or sometimes $Y \sim \text{exponential}(\theta)$ (mean θ).

Gamma Distribution

- A sum of iid $\text{exponential}(\theta)$ rv's is a $\text{gamma}(r, \theta)$ rv, if referring to the "mean notation".
- More commonly, a sum of r iid $\text{exponential}(\lambda)$ rv's (rate λ) is a $\text{gamma}(r, \lambda)$ with pdf

$$f(y) = \frac{\lambda^r y^{r-1} e^{-\lambda y}}{\Gamma(r)}, \; y > 0.$$

- Notice that $Y \sim \text{gamma}(r,\lambda) \Leftrightarrow Y = \frac{Z}{\lambda}$ where $Z \sim \text{gamma}(r,1)$. Since $E[Z] = r$, we get $E[Y] = \frac{r}{\lambda}$.
- If $Y_1 \sim \text{gamma}(r_1,\lambda) \perp Y_2 \sim \text{gamma}(r_2,\lambda)$, then $T = Y_1 + Y_2 \sim \text{gamma}(r_1 + r_2, \lambda)$.
- Note: The gamma function $\Gamma(r) = \int_0^\infty y^{r-1} e^{-y} dy, r > 0$, and $\Gamma(r+1) = r\Gamma(r)$, $\Gamma(k) = (k-1)!, n \in \mathbf{Z}$, $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, $Y \sim \text{gamma}(r,1) \Rightarrow E\left[Y^\alpha\right] = \frac{\Gamma(\alpha + r)}{\Gamma(r)}$.

Beta Distribution

- Consider $U = \frac{Y_1}{T} = \frac{Y_1}{Y_1 + Y_2}$. This is a $\text{beta}(r_1, r_2)$ rv.
- The pdf is easily obtained as $f(u) = cu^{r_1}(1-u)^{r_2-1}$ where $c = \frac{\Gamma(r_1 + r_2)}{\Gamma(r_1)\Gamma(r_2)}$.
- Note that *T* and *U* are independent, so $E[U]$ can be obtained easily. $E[Y_1] = E[UT] = E[U]E[T] \Rightarrow$ $E[U] = \frac{E[Y_1]}{E[T]} = \frac{r_1}{r_1 + r_2}$.

Chi-Squared(*n*) Distribution

- $X \sim \text{chi-squared}(n)$ (denoted $X \sim \chi_n^2$ or $X \sim \chi^2(n)$) if $X \stackrel{d}{=} Z_1^2 + \cdots + Z_n^2$, where Z_i iid $N(0,1)$.
- The pdf $f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}, y > 0$, which is $\text{gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$.

Student-*t* Distribution

- $Y \sim t(n)$ if $Y \stackrel{d}{=} \frac{Z}{\sqrt{\frac{X}{n}}}$, where $X \sim \chi^2(n)$.
- The pdf $f(y) \propto \left(1 + \frac{y^2}{n}\right)^{-\frac{n+1}{2}}$, where the constant is $\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}$.

F Distribution

- $Y \sim F(m,n) \Rightarrow f(y) \propto y^{\frac{m}{2}-1}\left(1 + \frac{m}{n} y\right)^{-\frac{m+n}{2}}, y > 0$.

Sequences of Random Variables

Suppose we have a sequence of random variables $Y_1, \ldots, Y_n, Y_{n+1}, \ldots$.

Definition: Convergence in Distribution
Y_n converges in distribution to Y ($Y_n \xrightarrow{d} Y$) if for every y such that $P(Y = y) = 0$, $\lim_{n\to\infty} F_n(y) = F(y)$, where $F_n(y) = P(Y_n \le y)$ and $F(y) = P(Y \le y)$.

Note
$Y_n \xrightarrow{d} Y \Leftrightarrow E[g(Y_n)] \to E[g(Y)]$ for all bounded continuous $g \Leftrightarrow c_n(t) \to c(t) \Leftarrow m_n(t) \to m(t)$, $G_n(s) \to G(s) \Leftarrow f_n(y) \to f(y)$.

Definition: Convergence in Mean Square
Y_n converges in mean square to Y ($Y_n \xrightarrow{ms} Y$) if $E\left[(Y_n - Y)^2\right] \to 0$.

Definition: Convergence in Probability
Y_n converges in probability to Y ($Y_n \xrightarrow{p} Y$) if for every $\varepsilon > 0$, $P\left[|Y_n - Y| \le \varepsilon\right] \to 0$.

Definition: Convergence With Probability One
Y_n converges with probability one to Y ($Y_n \xrightarrow{wp1} Y$) if $P(Y_n \to Y) \to 1$.

Weak Law of Large Numbers
Let $X_1, X_2, \ldots$ be iid with mean μ and variance σ^2. Then $\overline{X} = \dfrac{X_1 + \cdots + X_n}{n} \xrightarrow{p} \mu$.

Proof: Let $\varepsilon > 0$. Then $P\left(\left|\overline{X} - \mu\right| > \varepsilon\right) = P\left(\left(\overline{X} - \mu\right)^2 > \varepsilon^2\right) \le \dfrac{E\left[\left(\overline{X} - \mu\right)^2\right]}{\varepsilon^2}$. Now,
$E\left[\overline{X}\right] = \dfrac{1}{n}\left(E[X_1] + \cdots + E[X_n]\right) = \dfrac{n\mu}{n} = \mu$ and $\operatorname{var}\left(\overline{X}\right) = \dfrac{1}{n^2}\left(\operatorname{var}(X_1) + \cdots + \operatorname{var}(X_n)\right) = \dfrac{n\sigma^2}{n^2} = \dfrac{\sigma^2}{n}$. So $\dfrac{E\left[\left(\overline{X} - \mu\right)^2\right]}{\varepsilon^2} = \dfrac{\operatorname{var}\left(\overline{X}\right)}{\varepsilon^2} = \dfrac{\sigma^2}{n\varepsilon^2} \to 0$. Therefore $P\left(\left|\overline{X} - \mu\right| > \varepsilon\right) \to 0$ and thus $\overline{X} \xrightarrow{p} \mu$.

Convex Functions

Definition
A function is convex if for all x_0 there exists a c such that $g(x) \ge g(x_0) + c(x - x_0), \forall x$.

Jenson's Inequality
Let g be convex. Then $E[g(X)] \ge g(E[X])$.

MORE EXAMPLES

Example

Let $Y_1,\dots,Y_n$ be iid uniform$([0,1])$ and consider the order statistics $Y_{(1)} \le \cdots \le Y_{(n)}$. Set $W_n = n(1-Y_n)$.

Then $Y_{(n)} \xrightarrow{p} 1$. Let $\varepsilon > 0$. Then $P(|Y_{(n)}-1| \le \varepsilon) = P(1-Y_{(n)} \le \varepsilon) = P(Y_{(n)} \ge 1-\varepsilon) = 1-P(Y_{(n)} \le 1-\varepsilon) = 1-P(Y_{(1)} \le 1-\varepsilon,\dots Y_{(n)} \le 1-\varepsilon)$; since $Y_1,\dots,Y_n$ are iid uniform$([0,1])$, $1-P(Y_{(1)} \le 1-\varepsilon,\dots Y_{(n)} \le 1-\varepsilon) = 1-[P(Y_{(1)} \le 1-\varepsilon)]^n = 1-(1-\varepsilon)^n \to 1$. So $Y_{(n)} \xrightarrow{p} 1$.

Set $W_n = n(1-Y_n)$. Then $W_{(n)} \xrightarrow{d}$ exponential(1). Look at $F_n(w) = P(W_n \le w)$. If $W \le 0$, then $F(w) = 0$. If $W > 0$, then $F_n(w) = P(n(1-Y_{(n)}) \le w) = P\left(Y_{(n)} \ge 1-\frac{w}{n}\right) = 1-P\left(Y_{(n)} \le 1-\frac{w}{n}\right) = 1-\left(1-\frac{w}{n}\right)^n \to 1-e^{-w}$.

Therefore $\lim_{n\to\infty} F_n(w) = \begin{cases} 1-e^{-w}, w>0 \\ 0, \text{otherwise} \end{cases}$, which is the df of an exponentail(1). So $W_{(n)} \xrightarrow{d}$ exponential(1).

CENTRAL LIMIT THEOREM

Let $X_1, X_2, \dots$ be iid with mean μ and variance σ^2. Set $\overline{X} = \frac{X_1+\cdots+X_n}{n}$. Then $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{d} \mathrm{N}(0,1)$.

BIVARIATE RANDOM VECTORS

Bivariate Normal

Let Z_1, Z_2 be iid $\mathrm{N}(0,1)$, and take $\rho^2 < 1$, $\mu_1,\mu_2 \in \mathbf{R}$, $\sigma_1,\sigma_2 > 0$. Set $\begin{cases} X = \mu_1 + \sigma_1 Z_1 \\ Y = \mu_2 + \sigma_2\left(\rho Z_1 + \sqrt{1-\rho^2}\, Z_2\right) \end{cases}$.

$\begin{pmatrix} X \\ Y \end{pmatrix}$ is the bivariate normal. $\begin{cases} X \sim \mathrm{N}(\mu_1, \sigma_1{}^2) \\ Y \sim \mathrm{N}(\mu_1, \sigma_1{}^2) \end{cases}$.

Note that $\begin{cases} E(X) = \mu_1 \\ \mathrm{var}(X) = \sigma_1{}^2 \end{cases}$ and $\begin{cases} E(/Y) = \mu_2 \\ \mathrm{var}(Y) = \sigma_2{}^2 \end{cases}$; $\mathrm{cov}(X,Y) = \sigma_1\sigma_2\rho$ and $\mathrm{corr}(X,Y) = \rho$.

Also, $r(x) = E(Y \mid X = x) = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho(x-\mu_1)$ and $E(Y \mid X) = r(X) = \mu_2 + \frac{\sigma_2}{\sigma_1}\rho(X-\mu_1)$

Bivariate Poisson

Let $U \sim$ Poisson(λ_1), $V \sim$ Poisson(λ_2), $W \sim$ Poisson(λ_3), and set $\begin{cases} X = U + V \\ Y = V + W \end{cases}$. Then $\begin{pmatrix} X \\ Y \end{pmatrix}$ is the bivariate Poisson.

Note that $E(Y \mid X = n) = np + \lambda_3$.