Sample Space & Sample Points

Sample Space - set of all possible outcomes, represented by S. E.g. Tossing 2 6-faced dice, $S = \{(1,1),(1,2)...(2,1)...\}$ $(1,2) \neq (2,1)$. **Sample Points** - an outcome in the sample space. E.g $S = \{1,2,4,5,6\}$, Sample point : 1 or 2 or 3 ...

Events

Subset of sample space. $S = \{1,2,3,4,5,6\}$, an event that odd number occur $= \{1,3,5\}$.

- Simple Event - an event with 1 sample point
- Compound Event - more than 1 sample point
- Sure Event - sample space
- Null event - the empty set, $\emptyset$

Event Operations

- **Union** - $A \cup B$, contains all elements that belong to A or B or both.
- **Intersect** - $A \cap B$, contains all elements that belong to A and B.
- **Complement** - A' is all the elements in S that's not in A.
- **Mutual Exclusion** - $A \cap B = \emptyset$, A and B have no common elements

Note :

- $(A \cap B) \cup C \neq A \cap (B \cup C)$
- A and A' are mutually exclusive

Intersection of n events - $\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 ... A_n = \{x : x \in A_1 and ... x \in A_n\}$, contains all the elements common to events A_1, and ... A_n.

Union of n events - $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 ... A_n = \{x : x \in A_1 or ... x \in A_n\}$, contains all elements belong to **one or more** of events A_1, or ... A_n.

Operations

- $A \cap A' = \emptyset$
- $A \cap \emptyset = \emptyset$
- $A \cup A' = S$
- $(A')' = A$
- $(A \cap B)' = A' \cup B'$, applicable for more than 2 elements
- $(A \cup B)' = A' \cap B'$, applicable for more than 2 elements
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup B = A \cup (B \cap A')$
- $A = (A \cap B) \cup (A \cap B')$

Contained

If **all** elements in A are in B, then A is contained by B, $A \subset B = B \supset A$. If $A \subset B$ and $B \subset A$, then $A = B$

Multiplication Principle

If an operation can be performed n_1 ways, and if **for each of this ways** a second operation can be performed n_2 ways, then the total number of ways would be $n_1 n_2$ ways. Can be generalized to more than 2 $(n_1 n_2 ... n_k)$

Addition Principle

If an operation n_1 cannot be performed together with n_2, then $n_1 + n_2$ is the number of ways to perform n_1 or n_2. Can be generalized to more than 2, assuming no 2 procedures are performed together.

Permutation

Arrangement of r objects from a set of n objects where $r \leq n$. Order is taken into account. Number of ways to arrange n objects is $n!$.
Permutations of n distinct objects taken r at a time is denoted by : ${}_nP_r = n(n-1)...(n-(r-1)) = \frac{n!}{(n-r)!}$.
Idea : Putting n objects in r components. If $n = r$, then ${}_nP_n = n!$
Number of permutations of n distinct objects in **a circle** is $(n-1)!$. If not all n objects are distinct, then $n_1 + n_2 + ... n_k = n$, then the number of permutations will be ${}_nP_{n_1,n_2...n_k} = \frac{n!}{n_1! n_2! ... n_k!}$

Combination

Selecting r objects from n objects **without regard to order**. Creates a partition between 2 groups, 1 with r objects, the other with $n - r$ objects. Denoted by : $\binom{n}{r}$ or ${}_nC_r$

Frequency

$f_A = \frac{n_A}{n}$, relative frequency of A happening in n repetitions of an experiment. Properties :

- $0 \leq f_A \leq 1$
- $f_A = 1$ iff A occurs every time among n repetitions
- $f_A = 0$ iff A never occurs in n repetitions.
- if $A \cap B = \emptyset$ and $f_{A \cup B}$ is the frequency for event $A \cup B$, then $f_{A \cup B} = f_A + f_B$
- f_A stabilises near some definite value as the experiment is repeated more times.

Axiom of Probability

$Pr(A)$ is the chance that A occurs, satisfies :

- $0 \leq Pr(A) \leq 1$
- $Pr(S) = 1$
- if $A_1, A_2 ...$ are **mutually exclusive**, then $Pr(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Pr(A_i)$.
- If A and B are mutually exclusive, then $Pr(A \cup B) = Pr(A) + Pr(B)$

Properties

- $Pr(\emptyset) = 0$
- If $A_1, A_2 ... A_n$ are **mutually exclusive**, then $Pr(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} Pr(A_i)$.
- For any event A, $Pr(A') = 1 - Pr(A)$
- For any 2 events A and B, $Pr(A) = Pr(A \cap B) + Pr(A \cap B')$
- For any 2 events A and B, $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$
- For any 3 events A, B, C, $Pr(A \cup B \cup C) = Pr(A) + Pr(B) + Pr(C) - Pr(A \cap B) - Pr(A \cap C) - Pr(B \cap C) + Pr(A \cap B \cap C)$
- For any n events, $Pr(A_1 \cup A_2 \cup ... A_n) = \sum_{i=1}^{n} Pr(A_i) - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Pr(A_i \cap A_j)$
- If $A \subset B$, then $Pr(A) \leq Pr(B)$

Sample Space w Finite Outcomes

A sample space S with finite number of k outcomes, then $S = a_1, a_2, ... a_k$. Let $Pr(a_i) = p_i$ be the probability of a_i. Then let an event A consist of r possible outcomes, $1 \leq r \leq k$, where $A = a_{j_1}, a_{j_2}, ... a_{j_r}$, where $j_1, j_2 ... j_r$ represent any r indices from $1, 2, ... k$. Then $Pr(A) = p_{j_1} + p_{j_2} + ... + p_{j_r}$. **Probability of event A equals to sum of the various individual outcomes making up A**.
If all the outcomes are equally likely to occur, then $Pr(a_i) = \frac{1}{k}$. Then for any event A in that sample space, $Pr(A) = \frac{sample\ points\ in\ A}{sample\ points\ in\ S}$

Conditional Probability

$Pr(A \mid B)$ - **conditional probability** of A given that B occurred. Can just consider from the set of the event of that occurred. $Pr(A \mid B) = \frac{Pr(A \cap B)}{Pr(B)}$, $Pr(B \mid A) = \frac{Pr(A \cap B)}{Pr(A)}$.

Properties

- $0 \leq Pr(A \mid B) \leq 1$
- $Pr(S \mid B) = 1$
- If $B_1, B_2 ...$ are mutually exclusive, then $Pr(\bigcup_{i=1}^{\infty} B_i | A) = \sum_{i=1}^{\infty} Pr(B_i | A)$
- If B_1 and B_2 are disjoint, then $Pr(B_1 \cup B_2 | A) = Pr(B_1 | A) + Pr(B_2 | A)$

Multiplication

$Pr(A \cap B) = Pr(A) \times Pr(B|A)$ or
$Pr(A \cap B) = Pr(B) \times Pr(A|B)$, provided $Pr(A) > 0$ and $Pr(B) > 0$.
Extendable to more than 2 events,
$Pr(A_1 \cap ... A_n) = Pr(A_1) \times Pr(A_2|A_1) \times Pr(A_3|A_1 \cap A_2) ... Pr(A_n|A_1 \cap ... A_{n-1})$, provided $Pr(A_1 \cap ... A_{n-1} > 0)$

Total Probability

Partition - mutually exclusive $(A_i \cap A_j = \emptyset)$ and exhaustive $(\bigcup_{i=1}^{n} A_i = S)$, then for any event B, $Pr(B) = \sum_{i=1}^{n} Pr(B \cap A_i) = \sum_{i=1}^{n} Pr(A_i) Pr(B|A_i)$

Bayes Theorem

Let $A_1, A_2, ... A_n$ be a partition of sample space S. Then $Pr(A_k|B) = \frac{Pr(A_k) Pr(B|A_k)}{\sum_{i=1}^{n} Pr(A_i) Pr(B|A_i)}$, where $Pr(B)$ is just the denominator.

Independent Events

$Pr(A \cap B) = Pr(A) Pr(B)$. Independent events refer to events that don't influence other events.

Properties

$Pr(A) > 0$, $Pr(B) > 0$. Then

- $Pr(B|A) = Pr(B)$ and $Pr(A|B) = Pr(A)$
- A and B cannot be mutually exclusive.
- Sample space and empty set are independent of any event
- If $A \subset B$, then A and B are dependent unless $B = S$
- **Not represented on Venn diagram**, check independence using definition
- If A and B are independent, then so are A and B', A' and B, A' and B'

n Independent Events

Mutually independent implies pairwise independence, but not the other way

- Pairwise - $Pr(A_i \cap A_j) = Pr(A_i) Pr(A_j)$
- Any subset $\{A_{i_1}, A_{i_2}, ... A_{i_k}\}$ of $A_1, A_2, ... A_n$, $Pr(A_{i_1} \cap A_{i_2} ... A_{i_k}) = Pr(A_{i_1}) Pr(A_{i_2}) ... Pr(A_{i_k})$

Random Variable

Let S be a sample space associate with an experiment. A function X, which assigns a number to every element $s \in S$, is a random variable.

- X is a real-valued function
- The **range space** of X is the set of real numbers
- If $X(s) = s$, then $R_X = S$

Equivalent Events

Let R_X be the sample space of possible values of $X(s)$. Then let A be the a subset of the event space, and B be a subset of R_X. If $Pr(B) = Pr(A)$, then they are equivalent events.

$A_1 = \{HH\}$ is equivalent to $B_1 = \{2\}$
$A_2 = \{HT, TH\}$ is equivalent to $B_2 = \{1\}$
$A_3 = \{TT\}$ is equivalent to $B_3 = \{0\}$
$A_4 = \{HH, HT, TH\}$ is equivalent to $B_4 = \{2, 1\}$

Since probability of event B_i occuring equals probability of A_i, then they are equivalent. Basically $Pr(X = x)$ for all possible values of x

Discrete Probability Distributions

If a random variable is **finite or countable infinte**, then that variable is a discrete random variable.

Probability Function

Probability for each value of X to occur. Pairs of $(x_i, f(x_i))$ is the probability distribution of X. Must satisfy :

- $f(x_i) \geq 0$ for all x_i
- $\sum_{i=1}^{\infty} f(x_i) = 1$

$f(x) = Pr(X = x)$

Continuous Random Variable

If R_X is a range space of a random variable, then it is an **interval or collection of intervals**. Then X is a **continuous random variable**. **Probability Density Function** satisfies :

- $f(x_i) \geq 0$ for all $x \in R_X$
- $\int_{R_X} f(x)\,dx = 1$ or $\int_{\infty}^{-\infty} f(x)\,dx = 1$ since $f(x) = 0$ for x not in R_X Infinity values are replaced with the actual upper and lower bounds respectively.
- for any c and d such that $c < d$, $Pr(c \leq X \leq d) = \int_c^d f(x)\,dx$

Remarks :

- Area under the graph between $x = c$ and $x = d$.
- $Pr(X = x_0) = \int_{x_0}^{x_0} f(x)\,dx = 0$ then in the **continuous case**, $Pr(X) = 0$ and $Pr(c \leq X \leq d) = Pr(c \leq X < d) = Pr(c < X \leq d) = Pr(c < X < d)$ and $\leq$ and $<$ can be used **interchangably**.
- $Pr(A) = 0$ does not necessary imply $A = \emptyset$
- If X assumes value in some interval $[a, b]$, then $f(x) = 0$ for all X outside the interval.

Cumulative Distribution Function - $F(x) = Pr(X \leq x)$.

- If X is a discrete random variable then, $F(x) = \sum_{t \leq x} f(t) = \sum_{t \leq x} Pr(X = t)$

- $Pr(a \le X \le b) = Pr(X \le b) - Pr(X < a) = F(b) - F(a^-)$, where a^- is the largest possible value of X that is strictly less than a.
- If the only possible values are **integers**, and if a and b are integers, then $Pr(a \le X \le b) = Pr(X = a$ or $a+1$ or ...or $b) = F(b) - F(a-1)$. Taking $a = b$ gives $Pr(X = a) = F(a) - F(a-1)$
- If X is a continuous random variable then $F(x) = \int_{-\infty}^{x} f(t)$ dt, infinity value replaced with the actual lower bound
- $f(x) = \frac{dF(x)}{dx}$ if the derivative exists
- $Pr(a \le X \le b) = Pr(a < X \le b) = F(b) - F(a)$
- If $F(x)$ is a non-decreasing function, then $F(x_1) \le F(x_2)$ if $x_1 < x_2$
- $0 \le F(x) \le 1$

Mean and Variance

If X is discrete random variable with probability function $F_X(x)$, then **mean or expected value** of X, is $\mu_X = E(X) = \sum_i x_i f_X(x_i) = \sum_x x f_X(x)$. Expected value may not be a possible value of random variable.
If X is a **continuous** random variable with pdf $F_X(x)$, the mean of X is $\mu_X = E(X) = \int_{-\infty}^{\infty} x f_X(x)$ dx. Infinity values replace with actual lower and upper bounds.

- Provided the sum or integral exists
- In the discrete case, if $f_X(x) = \frac{1}{N}$ for each N values of x, then the mean is $E(X) = \sum_i x_i f(x_i) = \frac{1}{N}\sum_i x_i$ becomes the average of the N items.

Expection of a Function of a RV

For any function $g(X)$ of a random variable X with pf or pdf, $f_X(x)$.

- $E[g(X)] = \sum_x g(x) f_X(x)$, if X is a discrete random variable providing the sum exists, **and**
- $E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)$ dx, if X is a continuous random variable providing the integral exists. Infinity values replace with actual lower and upper bounds.

Special Cases

- $g(x) = (x - \mu_X)^2$
- Let X be random variable with pf or pdf $f(x)$, then variance of X is $\sigma_X^2 = V(X) = E[(X - \mu_X)^2]$
- If X is discrete, then $\sum_x (x - \mu_X)^2 f_X(x)$
- If X is continuous, then $\int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)$ dx
- $V(X) \ge 0$ and $V(X) = E(X^2) - [E(X)]^2$
- Positive squareroot of variance gives **standard deviation** - $\sigma_X = \sqrt{V(X)}$
- $g(x) = x^k$, then $E(X^k)$ is the **k-th moment of** X

Properties of Expectation

- $E(aX + b) = aE(X) + b$, where a and b are constants. Works for general number of inputs as long as the multipliers are constants.
- $V(aX + b) = a^2 V(X)$
- $(E((X - a)^2))$ **can be expanded into quadratic form to find** $E(X)$ **or** $E(X^2)$

Chebyshev's Inequality

Possible to get expected value and variance from random variable but not the other way. $E(X) = \mu$ and $V(X) = \sigma^2$, then for **any positive k**, $Pr(|X - \mu| \ge k\sigma) \le \frac{1}{k^2}$, the probability that a value of X lies at least k standard deviation from its mean is at most $\frac{1}{k^2}$, or $Pr(|X - \mu| < k\sigma) \ge 1 - \frac{1}{k^2}$ Notes :

- k can be any positive number
- Inequality is true for all distributions with finite mean and variance
- Lower bound on probability that $|X - \mu| < k\sigma$. No guarantee its close to the exact.

If the inequality gives undefined value for k, adjust the values until solution is found. Depending on the value adjusted, the result from Chebyshev's inequality will either be lower or upper bound. Symmetric about mean implies that a $Pr(X \le a - b) = Pr(X \ge a + b)$

2D Random Variable

If X and Y are functions that assign a real number to a sample sapace, then (X, Y) is a **2D random variable**. Can be extended to more than 2 functions.

- **Discrete** - finite or countable. Possible values are fixed
- **Continuous** - assume all values in some region

Joint Probability Function for Discrete RVs

If $f_{X,Y}(x_i, y_j)$ represents $Pr(X = x_i, Y = y_j)$ and satisfy the following

- $f_{X,Y}(x_i, y_j) \ge 0$ for all values tuples in $R_{X,Y}$
- $\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} Pr(X = x_i, Y = y_j) = 1$

If A is any set consisting of pairs of (x, y) values, then probability that $Pr((X, Y) \in A) = \sum\sum f_{X,Y}(x, y)$

Joint pdf for Continuous RVs

If (X, Y) is a 2D continuous random variable, then $f_{X,Y}(x, y)$ if it satisfies the following:

1. $f_{X,Y}(x,y) \ge 0$ for all $(x,y) \in R_{X,Y}$.
2. $$\iint_{(x,y)\in R_{X,Y}} f_{X,Y}(x,y)\,dx\,dy = 1$$

or

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx\,dy = 1.$$

Sequence to solving :

- Check if $f_{X,Y}(x, y)$ is a joint pdf, by seeing if it fulfills the conditions

Marginal Probability Distributions

Let (X, Y) be a 2D random variable with join pdf $f_{X,Y}(x, y)$. Then the marginal probability distribution is given by :

- For **discrete** case,

$$f_X(x) = \sum_y f_{X,Y}(x,y) \quad \text{and} \quad f_Y(y) = \sum_x f_{X,Y}(x,y)$$

- For **continuous** case,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

Conditional Distribution

- Then **the conditional distribution of** Y **given that** $X = x$ is given by

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad \text{if } f_X(x) > 0,$$

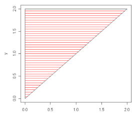
for each x within the range of X.

Satisfies requirements for all the 1D pf and pdfs. Uniform distribution over the triangle :

- Since $f_{X,Y}(x,y)$ is uniform over the triangle bounded by $0 \le x \le y, 0 \le y \le 2$, therefore

$$f_{X,Y}(x,y) = k \quad \text{for } 0 \le x \le y, 0 \le y \le 2.$$

- Note: The area bounded by $0 \le x \le y, 0 \le y \le 2$ is $(1/2)(2)(2) = 2$.

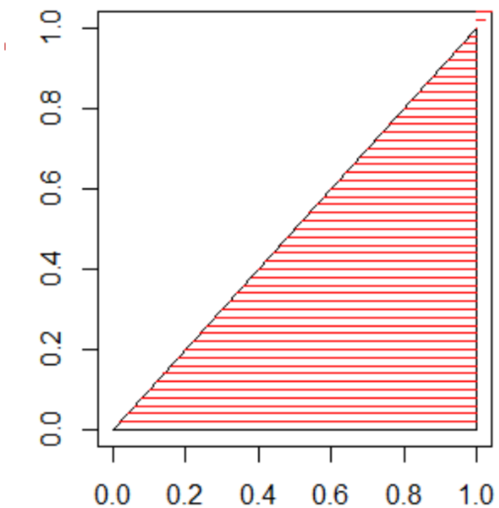


Independent Random Variable

Random variables are only independent if and only if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

Can be extended to multiple random variables. If the product is positive, then it is positive on a product space. If $f_{X,Y}(x, y) > 0$ is not a rectangle, it cannot be expressed as the product of 2 intervals, as seen in the image below.



If the region isn't a rectangle, just look for a point outside the region that satisfies both intervals. Then check if they are independent.

Method to determine independence

- Obtain $f_X(x)$ and $f_Y(y)$
- Multiply them together and see if they equal to $f_{X,Y}(x, y)$

Expectation

$$E[g(X,Y)] = \begin{cases} \sum_x\sum_y g(x,y) f_{X,Y}(x,y), & \text{for Discrete RV's,} \\ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dx\,dy, & \text{for Cont. RV's} \end{cases}$$

Covariance - variance of 2d rv.

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

- For **discrete** case

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sum_x\sum_y (x - \mu_X)(y - \mu_Y) f_{X,Y}(x,y)$$

- For **continuous** case

$$Cov(X,Y) = E[(X - \mu_X)(Y - \mu_Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f_{X,Y}(x,y)\,dx\,dy$$

Remarks :

- $Cov(X, Y) = E(XY) - E(X)E(Y)$
- If X and Y are independent, then $Cov(X, Y) = 0$, not necessarily true the other way.
- $Cov(aX + b, cY + d) = acCov(X, Y)$
- $V(aX + bY) = a^2V(X) + b^2V(Y) + 2abCov(X, Y)$

Correlation Coefficient

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$$

Remarks

- $-1 \le \rho_{X,Y} \le 1$
- Measure of degree of linear relationship
- If X and Y are independent, then coefficient is 0. Not necessarily true the other way.

Discrete Uniform Distribution

If random variable X has values with equal probability, then it has a discrete uniform distribution and the probability function is given by :

$$f_X(x) = \frac{1}{k}, \qquad x = x_1, x_2, \cdots, x_k,$$

and 0 otherwise.

Mean of the distribution would be :

$$\mu = E(X) = \sum_{\text{all } x} x f_X(x) = \sum_{i=1}^{k} x_i \frac{1}{k} = \frac{1}{k}\sum_{i=1}^{k} x_i,$$

Variance would be :

$$\sigma^2 = V(X) = \sum_{\text{all } x} (x-\mu)^2 f_X(x) = \frac{1}{k}\sum_{i=1}^{k}(x_i-\mu)^2$$

or

$$\sigma^2 = E(X^2) - \mu^2 = \frac{1}{k}\left(\sum_{i=1}^{k} x_i^2\right) - \mu^2$$

Bernoulli Distribution

Only has binary outcome. Satisfies probability function :

$$f_X(x) = p^x(1-p)^{1-x}, \quad x = 0, 1;$$

p is between 0 and 1 and $f_X(x) = 0$ for other values of X. $1-p$ is often denoted by q. The **mean** would be $\mu = p$ and the **variance** would be $\sigma^2 = pq$.

Binomial Distribution

Bernoulli distribution conducted n times with replacement and probability remains the same from trial to trial.

$X \sim B(n,p)$), if the probability function of X is given by

$$\Pr(X=x) = f_X(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} p^x q^{n-x},$$

for $x = 0, 1, \cdots, n$, where p satisfies $0 < p < 1$, $q = 1-p$, and n ranges over the positive integers.

X is number of successes that occur in n **independent** Bernoulli trials. When $n = 1$, its Bernoulli trial. **Mean** would be np. **Variance** is npq

Conditions :

- consists of n bernoulli trials
- only 2 outcomes in each trial
- probability of success constant for each trial
- trials are **independent**

Negative Binomial Distribution

Want to produce k successes where the last trial produces the kth success.

$$\Pr(X=x) = f_X(x) = \binom{x-1}{k-1} p^k q^{x-k},$$

for $x = k, k+1, k+2, \cdots$

Then **mean** would be $\frac{k}{p}$ and **variance** would be $\frac{(1-p)k}{p^2}$. Number of trials required to have first success is **geometric distribution**.

Poisson Distribution

Number of successes occurring during a given time interval or specific region. Properties :

- number of success in 1 region are **independent** of those occuring in any other disjoin region
- probability of single success during a small region is **proportional** to length of region.
- probability of more than 1 success occuring in a small region is neglible.

$$f_X(x) = \Pr(X=x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, 3, \cdots$$

where λ is the average number of successes occurring in the given time interval or specified region and $e \approx 2.1718281818\cdots$

Mean and **Variance** are λ. If want to find across multiple events, find p and then use binomial(provided independent events).

Approximation to Binomial - if $n \to \infty$ and $p \to 0$ in a way that $\lambda = np$ remains constant, then the binomial distribution will be approximately a Poisson one with $\lambda = np$.

$$\lim_{\substack{p\to 0\\ n\to\infty}} \Pr(X=x) = \frac{e^{-np}(np)^x}{x!}$$

Continuous Uniform Distribution

Uniform distribution over an interval if its pdf is

$$f_X(x) = \frac{1}{b-a}, \qquad \text{for } a \le x \le b,$$

Graph is shaped as a rectangle over the specified region. **Mean** is $\frac{a+b}{2}$ and **Variance** is $\frac{(b-a)^2}{12}$

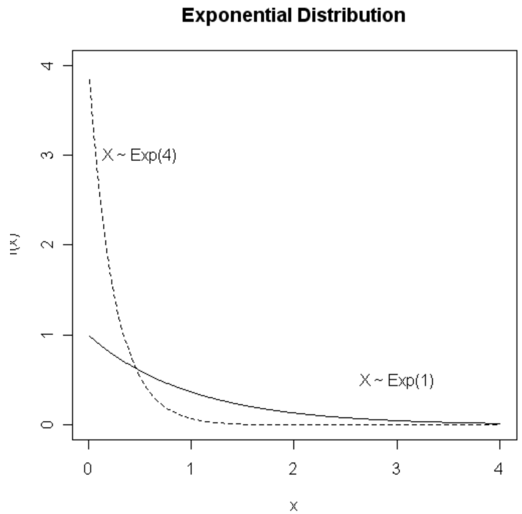
$$F_X(x) = \begin{cases} 0, & \text{for } x < a, \\ \dfrac{x-a}{b-a}, & \text{for } a \le x \le b, \\ 1, & \text{for } b < x. \end{cases}$$

Exponential Distribution

Assuming all nonnegative values and $\alpha > 0$

$$f_X(x) = \alpha e^{-\alpha x}, \qquad \text{for } x > 0.$$

Area under the graph is 1.



Mean is $\frac{1}{\alpha}$ and **Variance** is $\frac{1}{\alpha^2}$. Properties :

- for any 2 positive numbers s, t, $Pr(X > s+t | X > s) = Pr(X > t)$. No memory theorem analogy : lightbulb lasted for s, then probability itll last for $s+t$ is the same as probability it last for the first t as brand new.

CDF of exponential :

$$F_X(x) = \Pr(X \le x) = \int_0^x \alpha\, e^{-\alpha t}\, dt = [-e^{-\alpha t}]_0^x = 1 - e^{-\alpha x},$$

Normal Distribution

Assuming all real values,

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad -\infty < x < \infty,$$

where $-\infty < \mu < \infty$ and $\sigma > 0$.

- It is denoted by $N(\mu, \sigma^2)$.

Properties :

- This is the bellcurve and its symmetric about the vertical line $x = \mu$.
- maxmimum point at $x = \mu$, value is $\frac{1}{\sqrt{2\pi}\sigma}$
- Approaches the horizontal axis as we proceed away from mean
- Total area under the curve and above horizontal is 1.
- **Mean** is μ and **variance** is σ^2
- shape is determined by variance while position determined by mean.
- as σ increase, curve flattens, and vice versa.

If X has distribution $N(\mu, \sigma^2)$, and if

$$Z = \frac{(X-\mu)}{\sigma}$$

then Z has the $N(0,1)$ distribution.

That is, $E(Z) = 0$ and $V(Z) = 1$.

We say that Z has a standardized normal distribution.

That is, the p.d.f. of Z may be written as

$$f_Z(z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right).$$

Whenever X has distribution $N(\mu, \sigma^2)$, we can always simplify the process of evaluating the values of $\Pr(x_1 < X < x_2)$ by using the transformation $Z = (X-\mu)/\sigma$. Hence $x_1 < X < x_2$ is equivalent to

$$(x_1-\mu)/\sigma < Z < (x_2-\mu)/\sigma.$$

Let $z_1 = (x_1-\mu)/\sigma$ and $z_2 = (x_2-\mu)/\sigma$. Then

$$\Pr(x_1 < X < x_2) = \Pr(z_1 < Z < z_2).$$

Approximate binomial to normal distribution when $np > 5$ **and** $n(1-p) > 5$. Then $\mu = np$ and $\sigma^2 = np(1-p)$. As n becomes very big, $Z = \frac{X-np}{\sqrt{npq}}$ is approximately the standardised normal distribution.

Statistical Tables

Standardised normal distribution represented as $\Phi(z) = Pr(Z \le z)$. 100α percentage points. Since pdf of Z is symmetrical about 0, then :

$$\Pr(Z \ge z_\alpha) = \Pr(Z \le -z_\alpha) = \alpha.$$

$$\text{(i.e. } f_Z(-z) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(-z)^2}{2}\right) = f_Z(z)\text{)}$$

Continuity Correction

(a) $\Pr(X = k) \approx \Pr(k - ½ < X < k + ½)$.

(b) $\Pr(a \le X \le b) \approx \Pr(a - ½ < X < b + ½)$.
$\Pr(a < X \le b) \approx \Pr(a + ½ < X < b + ½)$.
$\Pr(a \le X < b) \approx \Pr(a - ½ < X < b - ½)$.
$\Pr(a < X < b) \approx \Pr(a + ½ < X < b - ½)$.

(c) $\Pr(X \le c) = \Pr(0 \le X \le c) \approx \Pr(-½ < X < c + ½)$.

(d) $\Pr(X > c) = \Pr(c < X \le n) \approx \Pr(c + ½ < X < n + ½)$.

Population & Sample

- Population : all possible outcomes or observations of a survey or experiment
- Finite Population : number of elements are finite (countable)
- Infinite Population : infinitely large number of elements
- Sample : subset of a population

Random Sampling

- Sample : members taken from a population
- Simple Random Sample : probability to take a sample is the same for every element

Sampling from a finite population

There are ${}_N C_n$ of sample size n that can be drawn from a population without replacement. Then each sample has equal chance of being chosen with probabiliy $\frac{1}{{}_N C_n}$.

Sampling with Replacement

Order matters in this case. Then number of samples that can be taken is N^n. Then probability of being chosen is $\frac{1}{N^n}$

Sampling from an infinite population

Equivalent to sampling from finite with replacement if

- In each draw, all elements have same probability of being selected
- Successive draws are independent

If X is a random variable with certain probability distribution $f_X(x)$. Let $X_1, \ldots X_n$ be n **independent** random variables having same distribution as X. Then the set with all the independent random variables is the **random sample of size n**.

- The joint p.f. (or p.d.f.) of $(X_1, X_2, \cdots, X_n)$ is given by $f_{X_1, X_2, \cdots, X_n}(x_1, x_2, \cdots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)$, where $f_X(x)$ is the p.f. (or p.d.f.) of the population.

Sample Mean

If $X_1, X_2, \ldots X_n$ is a random sample of size n, then sample mean would be

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

Similarly for values that are observed in a random sample.

- For random samples of size n taken from an **infinite population** or from a **finite population with replacement** having population mean μ and population standard deviation σ,
- the **sampling distribution of the sample mean** $\bar{X}$ has its mean and variance given by

$$\mu_{\bar{X}} = \mu_X \quad \text{and} \quad \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}.$$

Law of Large Number

As n gets larger, sample mean will approach population mean. In other words, probability that sample mean differs from population mean approaches 0.

Central Limit Theorem

If n is sufficiently large, then sampling distribution of sample mean is approximately normal.

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \text{ follows approximately } N(0, 1)$$

As long as some sample distribution is (or approximately is) a normal distribution, the sample mean will be approximately a normal distribution as well.

Sampling diff between 2 sample means

If 2 samples of size greater than or equal to 30 are drawn, then the difference between the 2 mean samples will be $\bar{X}_1 - \bar{X}_2$ and the resulting mean and sd will be :

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2 \quad \text{and} \quad \sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

This approximates the distribution to be standard normal.

Chi-Square Distribution

$$f_Y(y) = \frac{1}{2^{\frac{n}{2}} \Gamma(n/2)} y^{\frac{n}{2} - 1} e^{-\frac{y}{2}}$$

If $y > 0$ otherwise 0. n is the degree of freedom and must be **positive**. The gamma function is $\Gamma(n) = (n-1)!$ Properties :

- $E(Y) = n$ and $V(Y) = 2n$
- For large n, approximates to normal distribution with the mean and variance above
- Summing up chi-squared variables leads to a degree of freedom which is the sum of all of their degree of freedom
- If X follows a normal distribution, then X^2 chi-squared distribution with freedom of 1.
- α is area under the graph of distribution from point α to the end.
- $1 - \alpha$ is area under graph from point 0 to $1 - \alpha$
- If a probability statement has α, then that is the resulting probability.

$\chi^2(10; 0.9)$ means $\Pr(Y \geq \chi^2(10; 0.9)) = 0.9$ or $\Pr(Y \leq \chi^2(10; 0.9)) = 0.1$.

Sample variance is :

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{X}_i - \bar{X})^2 = \frac{1}{n-1} \sum_{i=1}^{n} \bar{X}_i^2 - n\bar{X}^2$$

If taken from a normal population with variance σ^2, then the random variable

$$\frac{(n-1)S^2}{\sigma^2}$$

has chi-square distribution with $n - 1$ degrees of freedom.

t-distribution

Suppose $Z \sim N(0, 1)$ and $U \sim \chi^2(n)$.

If Z and U are independent, and let

$$T = \frac{Z}{\sqrt{U/n}}$$

then the random variable T follows the t-distribution with n degrees of freedom. That is,

$$\frac{Z}{\sqrt{U/n}} \sim t(n)$$

$$f_T(t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\, \Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \qquad -\infty < t < \infty.$$

gamma function is the same as above. Properties :

- Symmetric about vertical axis, resembles graph of standard normal distribution
- As n approches ∞, pdf becomes $\frac{1}{\sqrt{2\pi} e^{\frac{-t^2}{2}}}$
- $Pr(T \geq t) = \int_t^{\infty} f_T(x)\, dx$
- Works with the alpha thing as well
- For $n > 2$, then $E(T) = 0$ and $V(T) = n/(n-2)$

If random sample was taken from a normal population, then

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{(\bar{X} - \mu)/(\sigma/\sqrt{n})}{\sqrt{\frac{(n-1)S^2}{\sigma^2} / (n-1)}} = \frac{Z}{\sqrt{U/(n-1)}} \sim t_{n-1},$$

- That is, T has a t-distribution with $n - 1$ d.f.

F-distribution

2 independent chi-squared distributions together creates a F distribution.

$$F = \frac{U/n_1}{V/n_2}$$

- The p.d.f. F is given by

$$f_F(x) = \frac{n_1^{n_1/2} n_2^{n_2/2} \Gamma\left(\frac{n_1 + n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} \frac{x^{n_1/2 - 1}}{(n_1 x + n_2)^{(n_1 + n_2)/2}},$$

for $x > 0$ and 0 otherwise.

- It can be shown that $E(X) = n_2/(n_2 - 2)$, with $n_2 > 2$ and $V(X) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}$, with $n_2 > 4$

$$F = \frac{U/(n_1 - 1)}{V/(n_2 - 1)} = \frac{\frac{(n_1 - 1)S_1^2/\sigma_1^2}{(n_1 - 1)}}{\frac{(n_2 - 1)S_2^2/\sigma_2^2}{(n_2 - 1)}}$$

$$= \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$$

Properties :

- $F \sim F(n, m)$ then $1/F \sim F(m, n)$
- $F(n_1, n_2; \alpha)$ such that $Pr(F > F(n_1, n_2; \alpha)) = \alpha$
- $F(n_1, n_2; 1 - \alpha) = 1/F(n_2, n_1; \alpha)$

Point Estimation

Assume that theres an unknown parameter in a function. Aim to estimate the value of this parameter, since we know the form.

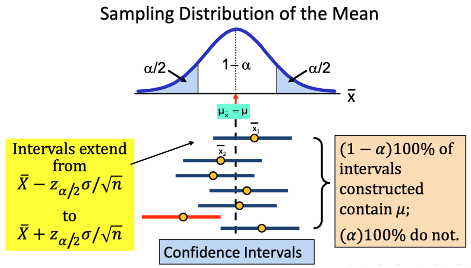
- Statistic - a function of random samples that don't depend on **any unknown parameters**. If any of the parameters are known, it can be a statistic.
- Statistic used to obtain point estimate is an estimator. Value is an estimate of paramter.
- If sample mean taken from population with a particular mean value is given, then the point estimate of that value is the given value.

Interval Estimation

Instead of using 1 statistic, use 2. The estimated value will lie between the 2 statistics. The upper and lower bounds depends on

- value of statistic for particular sample
- sampling distribution of statistic

Width of confidence interval is the 'distance' between the lower and upper limit. Any point in the boundary is a **point estimate**. Probability that the estimate is within the range is $1 - \alpha$. This is the degree of confidence.



Known Variance

When population is normal, we can expect $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ and we can normalise it as well. Then

If $\bar{X}$ is the mean of a random sample of size n from a population with known variance σ^2,

a $(1 - \alpha)100\%$ confidence interval for μ is given by

$$\left(\bar{X} - z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right) < \mu < \bar{X} + z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)\right)$$

Sample Size fo estimating mean

Most of the time, its an estimate, confirm have errors. Size of error would be $|\bar{X} - \mu|$. From above, we know that the probability of the error falling within the desired region is $1 - \alpha$. If e is the margin of error, and we don't want the error to be greater than e with probability greater than $1 - \alpha$, then $Pr(|\bar{X} - \mu| \leq e) \geq 1 - \alpha$.

- Since $\Pr\left(|\bar{X} - \mu| < z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$, therefore

$$e \geq z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

- Hence for a given margin of error e, the sample size is given by

$$n \geq \left(z_{\alpha/2} \frac{\sigma}{e}\right)^2.$$

Unbiased Estimator

Unbiased estimator means that the expected value of the estimator is that value. If X is an unbiased estimator of θ, then $E(X) = \theta$. If $E(X) \neq \theta$, then X is a biased estimator.

Unknown Variance

$$\Pr\left(-t_{n-1;\alpha/2} < T < t_{n-1;\alpha/2}\right) = 1-\alpha$$

$$\text{or}\quad \Pr\left(-t_{n-1;\alpha/2} < \frac{(\bar{X}-\mu)}{S/\sqrt{n}} < t_{n-1;\alpha/2}\right) = 1-\alpha$$

$$\text{or}\quad \Pr\left(-t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \bar{X}-\mu < t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right) = 1-\alpha$$

$$\text{or}\quad \Pr\left(\bar{X}-t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right) < \mu < \bar{X}+t_{n-1;\alpha/2}\left(\frac{S}{\sqrt{n}}\right)\right) = 1-\alpha$$

If sample size is smaller than 30, follow the equation above to estimate mean. If sample size is large, use standard normal distribution instead since t will be approximately standard normal distribution. Subtraction works the same as sampling between 2 means.

Known Variances

Possible scenarios for 2 variances that are known and not equal:

- both population are normal
- sample size greater than 30 for both

$$(\bar{X}_1-\bar{X}_2) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1-\mu_2$$
$$< (\bar{X}_1-\bar{X}_2) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}.$$

If population sd is unknown but the sample is sufficiently big, can estimate sample sd to be population sd. If the variances are equal, then the population variance can be estimated as such :

$$S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2},$$

If the population are normal with the same variance, can be converted into chi-squared distribution

$$\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\sigma^2} \sim \chi^2_{n_1+n_2-2}.$$

Confidence interval would then be given as :

$$(\bar{X}_1-\bar{X}_2) - t_{n_1+n_2-2;\alpha/2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} < \mu_1-\mu_2$$
$$< (\bar{X}_1-\bar{X}_2) + t_{n_1+n_2-2;\alpha/2}\, S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

If the sample size are large enough for both, use normal distribution instead of t.

Confidence Intervals for Variance

If mean is known, then

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;\,\alpha/2}} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{n;1-\alpha/2}}$$

Otherwise :

$$\frac{(n-1)S^2}{\chi^2_{n-1;\,\alpha/2}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{n-1;1-\alpha/2}}$$

where S^2 is the sample variance.

2 Variance Interval

If both means are unknown, use F distribution :

$$\frac{S_1^2}{S_2^2}\frac{1}{F_{n_1-1,n_2-1;\,\alpha/2}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2}F_{n_2-1,n_1-1;\,\alpha/2}$$

Null and Alternative Hypothesis

- Null hypothesis - formulate hypothesis with hope of rejecting it, H_0
- Alternative hypothesis - rejection of null hypothesis leads to acceptance of alternative hypothesis, H_1

Errors

	State of Nature	
Decision	H_0 is true	H_0 is false
Reject H_0	Type I error Pr(Reject H_0 given that H_0 is true) = α	Correct decision Pr(Reject H_0 given that H_0 is false) = 1 – β
Do not reject H_0	Correct decision Pr(Do not reject H_0 given that H_0 is true) = 1 – α	Type II error Pr(Do not reject H_0 given that H_0 is false) = β

- Type 1 - reject null hypothesis when its true, serious errors
- Type 2 - not reject null hypothesis when its false

Level of significance is the probability of rejecting null hypothesis when its true. Power of a test is 1 - probability of type 2 error.

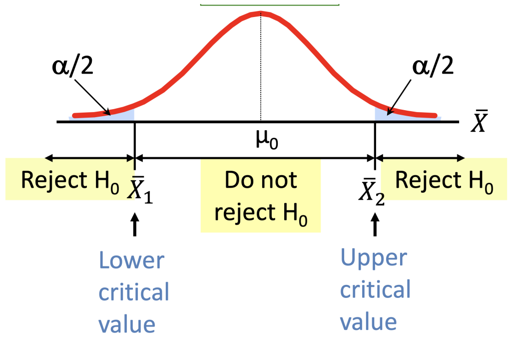
Acceptance & Rejection Regions

Value that separates rejection and acceptance regions is the **critical value**. In 2 tailed test, critical value is $\alpha/2$, otherwise its just α. If a value falls within the region, reject hypothesis.

Hypothesis Testing Concerning Mean

Figure below illustrates accepting or rejecting for 2 tailed test.

- 2 sided test - $\mu \neq \mu_0$
- 1 sided test - $\mu > \mu_0$ or $\mu < \mu_0$



Critical values can be found as :

Therefore

$$\Pr\left(-z_{\alpha/2} < \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$

or

$$\Pr\left(\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha.$$

Hence $\bar{x}_1 = \mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ and $\bar{x}_2 = \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$.

Process for hyp testing :

- Calculate sample mean
- If sample mean is in acceptance region, then accept that $\mu = \mu_0$. Otherwise reject
- Since $Z = (\bar{X}-\mu_0)/(\sigma/\sqrt{n})$, then $-z_{\alpha/2} < Z < z_{\alpha/2}$

With unknown variance, use sample variance. Instead of z-value, use t value instead.

p-value

Methodology :

- Convert to test statistic
- Obtain p-value
- Compare p-value with α
- If less than α, reject, else accept

Calculation :

$\bar{X}$ = 34.5 is translated to a Z score of Z = -2.33

Pr(Z < -2.33) = 0.0099
Pr(Z > -2.33) = 0.9901

***p*-value**
= 2 min{Pr(Z < -2.33), Pr(Z >-2.33) }
= 2(0.0099) = .0198

Double only if its double ended test

Testing

Observed level of significance - probability of obtaining test statistic more extreme than observed sample value **given** H_0 **is true**, also known as p-value. If α smaller than p-value, reject null hypothesis. Otherwise, accept it.

Testing for Variance

Assumption is that underlying distribution is normal. Use chi-squared value.

- H_0: $\sigma^2 = \sigma_0^2$ is rejected if the observed χ^2-value

H_1	Critical Region
$\sigma^2 > \sigma_0^2$	$\chi^2 > \chi^2_{n-1;\alpha}$
$\sigma^2 < \sigma_0^2$	$\chi^2 < \chi^2_{n-1;1-\alpha}$
$\sigma^2 \neq \sigma_0^2$	$\chi^2 < \chi^2_{n-1;1-\alpha/2}$ or $\chi^2 > \chi^2_{n-1;\alpha/2}$

where $\Pr(W > \chi^2_{n-1;\alpha}) = \alpha$ with $W \sim \chi^2(n-1)$

If testing for ratio of variance, use F-value instead.

- H_0: $\sigma_1^2 = \sigma_2^2$ is rejected if the observed F-value falls in the critical region

H_1	Critical Region
$\sigma_1^2 > \sigma_2^2$	$F > F_{(n_1-1,n_2-1;\alpha)}$
$\sigma_1^2 > \sigma_2^2$	$F < F_{(n_1-1,n_2-1;1-\alpha)}$
$\sigma_1^2 \neq \sigma_2^2$	$F < F_{(n_1-1,n_2-1;1-\alpha/2)}$ or $F > F_{(n_1-1,n_2-1;\alpha/2)}$

where $\Pr(W > F_{v_1,v_2;\alpha}) = \alpha$ with $W \sim F(v_1,v_2)$