Also, the joint cdf of X and Y is defined as

$$F_{X,Y}(x,y) = P[X \le x, Y \le y]$$

$$= \sum_{u \le x} \sum_{y \le v} p_{X,Y}(u,v) .$$

Thm: Let $\underset{\sim}{Z} = (X, Y)$ be a discrete random vector defined on a prob space $(\Omega, \mathcal{P}(\Omega), P)$ with Ω ctble. Let $\cancel{\phi : \mathbb{R}^2 \to \mathbb{R}}$ $\phi : \text{Range}(\underset{\sim}{Z}) \to \mathbb{R}$ be any map. Then $\phi(X,Y)$ is a r.v. defined on $(\Omega, \mathcal{P}(\Omega), P)$.

(a) If $\displaystyle\sum_{(x,y) \in \mathbb{R}^2} |\phi(x,y)| \, p_{X,Y}(x,y) < \infty$, then $\phi(X,Y)$ has finite mean given by

$$E[\phi(X,Y)] = \sum_{(x,y) \in \mathbb{R}^2} \phi(x,y) \, p_{X,Y}(x,y).$$

(b) If ϕ takes only nonnegative values, then

$$E[\phi(X,Y)] = \sum_{(x,y) \in \mathbb{R}^2} \phi(x,y) \, p_{X,Y}(x,y).$$

Pf: Exactly same proof as in the r.v. case.

Remark: One can again talk about ϕ^+, ϕ^- and evaluate $E[\phi(X,Y)]$ by $E(\phi^+(X,Y)) - E(\phi^-(X,Y))$ when it exists.

Cor: If X and Y are two r.v.s defined on the same prob space, ~~the~~ and both of them have finite mean, then $X+Y$ also has finite mean given by $E(X+Y) = E(X) + E(Y)$.

Proof: $\underset{\sim}{Z} : \Omega \to \mathbb{R}^2$ defined by $\underset{\sim}{Z}(\omega) = (X(\omega), Y(\omega))$ defines a random vector on $(\Omega, \mathcal{P}(\Omega), P)$, the prob space on which X, Y are defined.

Take $\phi : \mathbb{R}^2 \to \mathbb{R}$ defined by $\phi(x,y) = x+y$.

Then $\sum_x \sum_y |\phi(x,y)|\, p_{X,Y}(x,y)$

$$= \sum_x \sum_y |x+y|\, p_{X,Y}(x,y)$$

$$\leq \sum_x \sum_y |x|\, p_{X,Y}(x,y) + \sum_x \sum_y |y|\, p_{X,Y}(x,y)$$

$$= \sum_x |x| \sum_y p_{X,Y}(x,y) + \sum_y |y| \sum_x p_{X,Y}(x,y) \quad [\because \text{terms are nonneg}]$$

$$= \sum_x |x|\, p_X(x) + \sum_y |y|\, p_Y(y) < \infty \quad \text{since } X, Y \text{ have finite mean}$$

Therefore $E(X+Y) = \sum_x \sum_y (x+y)\, p_{X,Y}(x,y)$

$$= \sum_x x \sum_y p_{X,Y}(x,y) + \sum_y y \sum_x p_{X,Y}(x,y) \quad [\text{By absolute summability}]$$

$$= \sum_x x\, p_X(x) + \sum_y y\, p_Y(y)$$

$$= E(X) + E(Y)$$

We have used the following fact from analysis.

Fact: Let I, J be two countable sets and $(a_{ij})_{i\in I, j\in J}$ be a double-sequence of real numbers.

(a) If $a_{ij} \geq 0 \quad \forall (i,j) \in I\times J$, then

$$\sum_{i\in I}\sum_{j\in J} a_{ij} = \sum_{j\in J}\sum_{i\in I} a_{ij}.$$

(b) If $\sum_{i\in I}\sum_{j\in J} |a_{ij}| < \infty$, then

$$\sum_{i\in I}\sum_{j\in J} a_{ij} = \sum_{j\in J}\sum_{i\in I} a_{ij} \in (-\infty, \infty).$$

Back to prob:

Cor: If $X_1, X_2, \dots, X_k$ are r.v.s defined on same prob space and each X_i has finite mean, then so does $X_1 + X_2 + \cdots + X_k$ and $E\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} E(X_i)$.

Pf: Use induction on k.

Remark: If $X_1, X_2, \dots, X_k$ are nonnegative r.v.s defined on same prob space, then $E\left(\sum_{i=1}^{k} X_i\right) = \sum_{i=1}^{k} E(X_i)$.

Cor: If $X_1, X_2, \ldots, X_k$ are r.v.s defined on the same prob space each with finite mean, then for all $\alpha_1, \alpha_2, \ldots, \alpha_k \in \mathbb{R}$, $\sum_{i=1}^{k} \alpha_i X_i$ has finite mean and

$$E\left(\sum_{i=1}^{k} \alpha_i X_i\right) = \sum_{i=1}^{k} \alpha_i E(X_i).$$

Pf: By the previous corollary, it is enough to show the following: if X is a r.v. with finite mean, ~~then for all~~ and $\alpha \in \mathbb{R}$, then αX also has finite mean, ~~This is~~ and $E(\alpha X) = \alpha E(X)$.

This follows because

$$\sum_{x} |\alpha x| \, p_X(x) = |\alpha| \sum_{x} |x| \, p_X(x) < \infty$$

and, hence $E(\alpha X) = \sum_{x} \alpha x \, p_X(x) = \alpha \sum_{x} x \, p_X(x) = \alpha E(X)$.

In particular, $\boxed{E(\alpha X + \beta) = \alpha E(X) + \beta.}$ (Expectation behaves very nicely with change of unit)

Remark: Fix a prob space $(\Omega, \mathcal{P}(\Omega), P)$.
ctble

Then $V \doteq$ tsoa r.v.s on this prob space forms a vector space. Define $L^1(\Omega, \mathcal{P}(\Omega), P)$ to be $\{X \in V : X \text{ has finite mean}\}$. The above result says that $L^1(\Omega, \mathcal{P}(\Omega), P)$ forms a subspace of V. Also the expectation map $E : L^1 \to \mathbb{R}$, $X \longmapsto E(X)$ ~~defined~~ is a linear map.

Application:

1. $X \sim Bin(n, p) \Rightarrow E(X) = np.$

$$X \sim Bin(n,p) \Rightarrow X = \sum_{i=1}^{n} X_i \text{ , where each } X_i \sim Ber(p).$$

$$\Rightarrow E(X) = \sum_{i=1}^{n} E(X_i) = np.$$

2. $X \sim NB(r, p) \Rightarrow E(X) = \frac{r}{p}.$

$$X \sim NB(r,p) \Rightarrow X = \sum_{i=1}^{r} X_i \text{ , where each } X_i \sim Geo(p)$$

$$\Rightarrow E(X) = \sum_{i=1}^{r} E(X_i) = \frac{r}{p}.$$

3. $X \sim Hyp(N, M, n) \Rightarrow E(X) = n\frac{M}{N}.$

Recall: N items $\equiv$ M Type I $+$ $(N-M)$ Type II

n distinct items are selected together.

$X =$ No. of Type I items selected.

Instead of selecting n distinct items together, choose one item at a time without replacement. Of course, this will ~~enforce~~ mean ~~an~~ ordering ~~among~~ of the selected items will become important, and hence the sample space and the prob will both change. However since X is the no. of items of Type I, the distn. of X will remain unchanged.

170

Define $X_i = I_{(i\text{th selected item is of Type I})} \quad \forall\ i = 1, 2, \ldots, n$

$$= \begin{cases} 1 & \text{if } i\text{th selected item is of Type I}, \\ 0 & \text{otherwise.} \end{cases}$$

Then $X = \sum_{i=1}^{n} X_i$.

$$\Rightarrow \quad E(X) = \sum_{i=1}^{n} E(X_i)$$

$$= \sum_{i=1}^{n} P(i\text{th selected item is of Type I})$$

Let us name the items $1, 2, \ldots, N$. Without loss of generality, assume that items $1, 2, \ldots, M$ are of Type I and $M+1, M+2, \ldots, N$ are of Type II.

Then $P(1^{st}$ selected item is of Type I$)$

$$= \sum_{j=1}^{M} P(1^{st} \text{ selected item is item } j) = \frac{M}{N}.$$

$P(2^{nd}$ selected item is of Type I$)$

$$= \sum_{j=1}^{M} P(2^{nd} \text{ selected item is item } j) = \frac{M}{N}.$$

In general, for all $i = 1, 2, \ldots, n$, $E(X_i) = \frac{M}{N}$.

Therefore, $E(X) = \sum_{i=1}^{n} E(X_i) = n \frac{M}{N}$.

4. Suppose we have a population of ~~size~~ N distinct items and we sample randomly from this population with replacement. Let S_r = ~~No of samples~~ Sample size necessary to obtain r distinct items in the sample. $(2 \leq r \leq N)$

$E(S_r) = ?$

Let X_1 = No. of draws necessary (after the first draw) to get a new item.

X_2 = No. of draws necessary after the $(X_1+1)^{th}$ draw to obtain a new item,

$\vdots$

X_{r-1} = No. of draws necessary after the $(X_1+X_2+\cdots+X_{r-2}+1)^{th}$ draw to obtain a new item.

Clearly, $S_r = 1 + X_1 + X_2 + \cdots + X_{r-1}$

and each $X_i \sim Geo\left(\frac{N-i}{N}\right)$.

Therefore $E(S_r) = 1 + E(X_1) + E(X_2) + \cdots + E(X_{r-1})$

$$= 1 + \left(\frac{N-1}{N}\right)^{-1} + \left(\frac{N-2}{N}\right)^{-1} + \cdots \left(\frac{N-r+1}{N}\right)^{-1}$$

$$= N\left(\frac{1}{N} + \frac{1}{N-1} + \cdots + \frac{1}{N-r+1}\right).$$

Ref: See also Coupon Collector's problem from Feller (Vol I).
Exc: Top-to-random shuffle.

Variance: Let X be any r.v. with finite mean μ. Then variance of X is defined by

$$\sigma_X^2 = V(X) = Var(X) = E\left[(X-\mu)^2\right].$$

$$Var(X) = E\left[(X-\mu)^2\right] = E\left[X^2 - 2\mu X + \mu^2\right]$$

$$= E(X^2) - 2\mu E(X) + \mu^2$$

$$= E(X^2) - \mu^2$$

Therefore $Var(X) < \infty$ iff $E(X^2) < \infty$.

When $E(X^2) = \infty$, sometimes it is said that variance of X does not exist.

Examples: 0. $X \sim Ber(p) \Rightarrow Var(X) = p(1-p)$

1. $X \sim Bin(n,p)$: $E(X^2) = n(n-1)p^2 + np$

$$\Rightarrow Var(X) = E(X^2) - (E(X))^2 = n^2p^2 - np^2 + np - n^2p^2$$
$$= np(1-p).$$

2. $X \sim \text{Poi}(\lambda)$: Calculate $\text{Var}(X)$. (Exc)

3. $X \sim \text{Geo}(p)$:

$$E(X^2) = \frac{2-p}{p^2}$$

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

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

4. $X \sim \text{Unif}\{x_1, x_2, \dots, x_n\}$

$$\text{Var}(X) = E(X^2) - (E(X))^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\bar{x})^2 \quad , \quad \text{where } \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

$$= \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

In particular $X \sim \text{Unif}\{1, 2, \dots, n\}$

$$\Rightarrow \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n} i^2 - \left(\frac{n+1}{2}\right)^2$$

$$= \frac{(n+1)(2n+1)}{6} - \frac{(n+1)(n+1)}{4}$$

$$= \frac{(n+1)(4n+2-3n-3)}{12} = \frac{n^2-1}{12}.$$

Propn: $\underline{Var(aX+b) = a^2 Var(X)}$

Pf: $E(aX+b) = a E(X) + b$

$$Var(aX+b) = E\left[\left((aX+b) - (aE(X)+b)\right)^2\right]$$

$$= a^2 E\left[(X - E(X))^2\right]$$

$$= a^2 Var(X).$$

Cor: If X has finite mean μ and finite variance σ^2 $(\sigma > 0)$, then $X^* = \dfrac{X-\mu}{\sigma}$ has mean 0 and variance 1.

X^* is called the "standardized value" of X.

$+\sqrt{Var(X)} = \sigma_X$ is called the standard deviation of X.

Now we shall derive a formula that can be used as a shortcut to calculate the variance ~~of~~ in many cases.

Then $E(X)$ and $E(X^2)$ exist finitely but

$E(X^3) = \infty$.

In this case, define $Y = X^2$. Then

$E(XY)$ is infinite although both $E(X)$ and $E(Y)$ are finite.

Now, we are in a position to define covariance of two random variables.

<u>Defn:</u> Let X, Y be two discrete r.v.s defined on the same prob space. ~~Then~~ and they both have finite ~~mean~~ mean. Then the covariance between X and Y is defined by $Cov(X,Y) \doteq E\left[(X-\mu_X)(Y-\mu_Y)\right]$ provided this expectation exists finitely.

<u>Note:</u> $$(X-\mu_X)(Y-\mu_Y) = XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y$$

$\Rightarrow$ $Cov(X,Y)$ is defined $\Leftrightarrow$ $E(XY)$ exists finitely

and in this case, $Cov(X,Y) = E(XY) - E(X)E(Y)$.

Fact: X, Y are ind r.v.s with finite means $\Rightarrow Cov(X,Y)=0$

~~Defn:~~ One sufficient condition for $Cov(X,Y)$ to be defined is that both X and Y have finite 2nd moments.

This is because $\sum_y \sum_x |xy| \, p_{X,Y}(x,y)$

$$< \sum_y \sum_x \frac{x^2+y^2}{2} \, p_{X,Y}(x,y)$$

$$= \frac{1}{2}\left(E(X^2) + E(Y^2)\right) < \infty .$$

Defn: X, Y are said to be

(i) positively correlated if $Cov(X,Y) > 0$;

(ii) negatively correlated if $Cov(X,Y) < 0$;

(iii) uncorrelated if $Cov(X,Y) = 0$.

Fact: X, Y ind with finite means $\Rightarrow$ X, Y are uncorrelated.

Properties of Covariance: Assume that r.v.s $X, Y, X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n$ are all defined on the same prob space and $a_0, a_1, \ldots, a_m, b_0, b_1, \ldots, b_n$ are real numbers. Also assume that all the quantities written below are defined and finite.

1. $\underline{Cov(X,X) = Var(X)}$

2. $\underline{Cov(X,Y) = Cov(Y,X)}$

3. $\underline{Cov(aX+b, cY+d) = ac\, Cov(X,Y)}$

$$(aX+b) - E(aX+b) = a(X-\mu_x)$$
$$(cY+d) - E(cY+d) = c(Y-\mu_y)$$

$$\Rightarrow Cov(aX+b, cY+d) = ac\, Cov(X,Y)$$

(Note: If we take $Y = X$, $a=c$, $b=d$, then we get ~~back~~ $Var(aX+b) = a^2\, Var(X)$, which we have proved already.)

This result tells us how covariance changes with ~~changes~~ ~~with~~ the change of unit.

4. $$\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i + a_0,\ \sum_{j=1}^{n} b_j Y_j + b_0\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \,\mathrm{Cov}(X_i, Y_j)$$

Pf: First we ~~show~~ claim that $\mathrm{Cov}(X_1+X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y)$.

$$\cancel{(X_1+X_2)Y - E((X_1+X_2)Y) =}$$

This is true because

$$\begin{aligned}
\mathrm{Cov}(X_1+X_2, Y) &= E\left[(X_1+X_2)Y\right] - E(X_1+X_2)E(Y) \\
&= \left(E(X_1Y) - E(X_1)E(Y)\right) \\
&\quad + \left(E(X_2Y) - E(X_2)E(Y)\right) \\
&= \mathrm{Cov}(X_1, Y) - \mathrm{Cov}(X_2, Y).
\end{aligned}$$

Now use Property 2 to observe that

$$\mathrm{Cov}(X, Y_1+Y_2) = \mathrm{Cov}(X, Y_1) + \mathrm{Cov}(X, Y_2).$$

Therefore

$$\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i + a_0,\ \sum_{j=1}^{n} b_j Y_j + b_0\right)$$

$$\overset{3}{=} \mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i,\ \sum_{j=1}^{n} b_j Y_j\right) \underset{m}{\overset{\text{ind on}}{=}} \sum_{i=1}^{m} \mathrm{Cov}\left(a_i X_i,\ \sum_{j=1}^{n} b_j Y_j\right)$$

$$\overset{\text{ind}}{\underset{\text{cons}}{=}} \sum_{i=1}^{m}\sum_{j=1}^{n} Cov(a_i X_i, b_j Y_j) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \, Cov(X_i, Y_j).$$

4.1. $$Var\left(\sum_{i=1}^{m} a_i X_i + a_0\right) = \sum_{i=1}^{m} a_i^2 \, Var(X_i) + 2\sum_{i=1}^{m}\sum_{\substack{j=1 \\ i<j}}^{m} a_i a_j \, Cov(X_i, X_j)$$

Pf: Take $m = n$, $a_i = b_i \ \forall \ i = 0, 1, \dots, n$ and $X_i = Y_i \ \forall \ i = 1, 2, \dots, n$ in 4 to get

$$Var\left(\sum_{i=1}^{m} a_i X_i + a_0\right) = \sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j \, Cov(X_i, X_j)$$

$$= \sum_{i=1}^{m} a_i^2 \, Var(X_i) + 2\sum\sum_{i<j} a_i a_j \, Cov(X_i, X_j)$$

4.2. If $X_1, X_2, \dots, X_m$ are (pairwise) uncorrelated, (i.e., $Cov(X_i, X_j) = 0 \ \forall \ i \neq j$), then

$$Var\left(\sum_{i=1}^{m} a_i X_i + a_0\right) = \sum_{i=1}^{m} a_i^2 \, Var(X_i). \quad \dots (4.2)$$

In particular if $X_1, X_2, \dots, X_m$ are ind, then (4.2) holds.

4.3. $$\mathrm{Var}\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} \mathrm{Var}(X_i) + 2\sum_{i=1}^{m}\sum_{\substack{j=1 \\ i<j}}^{m} \mathrm{Cov}(X_i, X_j).$$

4.4. If $X_1, X_2, \ldots, X_m$ are pairwise uncorrelated, then $$\mathrm{Var}\left(\sum_{i=1}^{m} X_i\right) = \sum_{i=1}^{m} \mathrm{Var}(X_i). \quad \ldots (4.4)$$

In part, if $X_1, X_2, \ldots, X_m$ are ind then (4.4) holds.

Consequence:

1. $X \sim \mathrm{Bin}(n, p) \Rightarrow X = \sum_{i=1}^{n} X_i$, where

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathrm{Ber}(p)$

$$\Rightarrow \mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = np(1-p).$$

2. $X \sim NB(r, p) \Rightarrow X = \sum_{i=1}^{r} X_i$,

where $X_1, X_2, \ldots, X_r \overset{iid}{\sim} \mathrm{Geo}(p)$.

$$\Rightarrow \mathrm{Var}(X) = \sum_{i=1}^{r} \mathrm{Var}(X_i) = \frac{rq}{p^2}.$$

Qn: What do variance and covariance measure?

Variance measures the average squared deviation of a r.v. from its mean. ~~in th~~ If X is concentrated around in mean, then it has a small variance. On the other hand, if X is very spread out from its mean then $V(X)$ is large.

For covariance, observe that

$E\left((X-\mu_x)(Y-\mu_y)\right)$ will be +ve ~~for~~ if for most cases, $X > \mu_x \Rightarrow Y > \mu_y$ and $X < \mu_x \Rightarrow Y < \mu_y$

$E\left((X-\mu_x)(Y-\mu_y)\right)$ will be -ve if for most cases, $X > \mu_x \Rightarrow Y < \mu_y$ and $X < \mu_x \Rightarrow Y > \mu_y$.

~~Therefore~~ Covariance measures the amount of linear ~~relati~~ relationship between X and Y. The sign gives the direction of the relationship. This explains why $Cov(X,Y) < 0$ in the RW on triangle with 3 steps.

Here is an example to show

$Cov(X,Y) = 0$ may not imply X, Y are ind

Let $X \sim \text{Unif}\{-2,-1,0,1,2\}$.

$Y = X^2$, Then $Cov(X,Y) = 0$ (Check this!)

but X, Y are not ind (check this!)

Note: Here X, Y have a perfect quadratic relationship but no linear relationship.

We ~~show~~ shall now use 4.3 to compute the variance of hypergeometric distribution. Before doing it, we need study another concept which will be useful in this context and also other contexts.

Defn: n random variables $X_1, X_2, \ldots, X_n$ defined on the same prob space are called exchangeable if for all $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$, and for all permutation π,

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(X_{\pi(1)} = x_1, \ldots, X_{\pi(n)} = x_n).$$

Slightly different way of giving this defn is as follows.

Defn: Suppose $\underset{\sim}{X} = (X_1, X_2, \ldots, X_n)$ and $\underset{\sim}{Y} = (Y_1, Y_2, \ldots, Y_n)$ are two random vectors defined on prob spaces $(\Omega, \mathcal{P}(\Omega), P)$ and $(\Omega', \mathcal{P}(\Omega'), P')$, respectively, where Ω, Ω' are ctble. We say that $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ are equal in distribution (denoted by $(X_1, X_2, \ldots, X_n) \overset{d}{=} (Y_1, Y_2, \ldots, Y_n)$) if they have the same pmf, ~~In other words~~, i.e., for all $\underset{\sim}{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$,

$$P[X_1 = u_1, X_2 = u_2, \ldots, X_n = u_n] = P'[Y_1 = u_1, Y_2 = u_2, \ldots, Y_n = u_n].$$

Fact: $(X_1, X_2, \ldots, X_n) \overset{d}{=} (Y_1, Y_2, \ldots, Y_n)$

$\Rightarrow$ For all $A \subseteq \mathbb{R}^n$,

$$P[(X_1, X_2, \ldots, X_n) \in A] = P'[(Y_1, Y_2, \ldots, Y_n) \in A].$$

Pf: Take $A \subseteq \mathbb{R}^n$. Then

$$P[\underset{\sim}{X} \in A] = \sum_{\underset{\sim}{u} \in A} p_{\underset{\sim}{X}}(\underset{\sim}{u}) = \sum_{\underset{\sim}{u} \in A} p_{\underset{\sim}{Y}}(\underset{\sim}{u}) = P'[\underset{\sim}{Y} \in A]$$

Fact: $\underline{(X_1, X_2, \dots, X_n) \stackrel{d}{=} (Y_1, Y_2, \dots, Y_n)}$

$\Rightarrow$ For all function $\phi : \mathbb{R}^n \to \mathbb{R}^m$,

$$\phi(X_1, X_2, \dots, X_n) \stackrel{d}{=} \phi(Y_1, Y_2, \dots, Y_n)$$

Pf: Take $\underset{\sim}{v} = (v_1, v_2, \dots, v_m) \in \mathbb{R}^m$. Then

$$P[\phi(\underset{\sim}{X}) = \underset{\sim}{v}] = P[\underset{\sim}{X} \in \phi^{-1}(\{\underset{\sim}{v}\})]$$

$$= P'[\underset{\sim}{Y} \in \phi^{-1}(\{\underset{\sim}{v}\})]$$

$$= P'[\phi(\underset{\sim}{Y}) = \underset{\sim}{v}].$$

Hence $\phi(\underset{\sim}{X}) \stackrel{d}{=} \phi(\underset{\sim}{Y})$.

Cor: $\underline{(X_1, X_2, \dots, X_n) \stackrel{d}{=} (Y_1, Y_2, \dots, Y_n)}$

$\Rightarrow$ For all $\cancel{k \geq 1}$ $1 \leq k \leq n$ and for all $\{i_1, i_2, \dots, i_k\} \subseteq \{1, 2, \dots, n\}$,

$$(X_{i_1}, X_{i_2}, \dots, X_{i_k}) \stackrel{d}{=} (Y_{i_1}, Y_{i_2}, \dots, Y_{i_k}).$$

Pf: Apply Fact with $\phi : \mathbb{R}^n \to \mathbb{R}^k$

$(u_1, u_2, \dots, u_n) \mapsto (u_{i_1}, u_{i_2}, \dots, u_{i_k})$.

Cor: $(X_1, X_2, \ldots, X_n) \stackrel{d}{=} (Y_1, Y_2, \ldots, Y_n)$

$\Rightarrow$ For all function $h: \mathbb{R}^n \to \mathbb{R}$,

$$E[h(X_1, X_2, \ldots, X_n)] = E[h(Y_1, Y_2, \ldots, Y_n)]$$

as long as at least one of the expectations is defined.

Pf: Exc.

Now we can define exchangeable r.v.s in this language.

Defn: n random variables $X_1, X_2, \ldots, X_n$ defined on the same discrete prob space are called exchangeable if $(X_1, X_2, \ldots, X_n) \stackrel{d}{=} (X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)})$ for all permutation π of $\{1, 2, \ldots, n\}$.

Examples:

1. If $X_1, X_2, \ldots, X_n$ are iid r.v.s defined on the same prob space, then they are exchangeable. (Exc.)

2. Suppose we have a population $\{1, 2, \ldots, N\}$ of N items and we choose a random sample $(X_1, X_2, \ldots, X_n)$ of size n from this population without replacement. Then $X_1, X_2, \ldots, X_n$ are exchangeable (Exc).

Fact: If $X_1, X_2, \ldots, X_n$ are exchangeable r.v.s, then for all $1 \le k \le n$ and for all k-element subset $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$,

$$(X_{i_1}, X_{i_2}, \ldots, X_{i_k}) \stackrel{d}{=} (X_1, X_2, \ldots, X_k).$$

Pf: Choose a permutation π on $\{1, 2, \ldots, n\}$ that sends j to i_j for all $j = 1, 2, \ldots, k$. Then by exchangeability,

$$(X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)}) \stackrel{d}{=} (X_1, X_2, \ldots, X_n)$$

$$\Rightarrow (X_{i_1}, X_{i_2}, \ldots, X_{i_k}, X_{\pi(k+1)}, \ldots, X_{\pi(n)}) \stackrel{d}{=} (X_1, X_2, \ldots, X_k, X_{k+1}, \ldots, X_n)$$

$$\Rightarrow (X_{i_1}, X_{i_2}, \ldots, X_{i_k}) \stackrel{d}{=} (X_1, X_2, \ldots, X_k).$$

Cor: If $(X_1, X_2, \ldots, X_n)$ is a SRSWOR drawn from $\{1, 2, \ldots, N\}$ then we have:–

(a) For all $1 \le i \le n$ and for all $1 \le l \le N$,

$$P[X_i = l] = \frac{1}{N}.$$

(b) For all $1 \le i < j \le n$ and for all $1 \le l_1 \ne l_2 \le N$,

$$P[X_i = l_1, X_j = l_2] = \frac{1}{N(N-1)}.$$

A Quick Application of Exchangeability

If $(X_1, X_2, \dots, X_n)$ is a simple random sample drawn from the population $\{1, 2, \dots, N\}$ without replacement, then

$$E\left(\frac{X_1}{\sum_{j=1}^{n} X_j}\right) = \frac{1}{n}.$$

Pf: Take $2 \le i \le n$ and let π be a permutation that sends 1 to i.

Then by exchangeability,

$$(X_1, X_2, \dots, X_n) \stackrel{d}{=} (X_i, X_{\pi(2)}, \dots, X_{\pi(n)}).$$

Now apply the function $h: \mathbb{R}^n \to \mathbb{R}$

$$h(x_1, x_2, \dots, x_n) = \begin{cases} \dfrac{x_1}{\sum_{i=1}^{n} x_i} & \text{if } \sum_{i=1}^{n} x_i \neq 0, \\ 0 & \text{if } \sum_{i=1}^{n} x_i = 0. \end{cases}$$

on both sides to conclude

$$E\left(\frac{X_1}{\sum_{i=1}^{n} X_j}\right) = E\left(\frac{X_i}{\sum_{i=1}^{n} X_j}\right) \quad \text{for all } 2 \le i \le n.$$

$$= \frac{1}{n}.$$

Two More Applications

<u>Ex</u>: Let S_n denote the total number of black balls drawn in the first n drawings of a Polya's urn scheme. $E(S_n) = ?$ $Var(S_n) = ?$

<u>Ans</u>: Define $X_i = I_{(i^{th} \text{ selected ball is black})}$, $i = 1, 2, \dots$.

Then $\forall\, n \geq 1$, $(X_1, X_2, \dots, X_n)$ is exchangeable.

<u>Qn</u>: <u>How do we show this?</u>

<u>Ans</u>: To show $\forall\ i_1, i_2, \dots, i_n \in \{0, 1\}$,

$$P(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n) = P(X_{\pi(1)} = i_1, \dots, X_{\pi(n)} = i_n)$$
$$= P(X_1 = i_{\pi^{-1}(1)}, \dots, X_n = i_{\pi^{-1}(n)})$$

Take $n = 3$; $i_1 = 1$, $i_2 = 0$, $i_3 = 0$.

To show

$$P(X_1 = 1, X_2 = 0, X_3 = 0) = P(X_1 = 0, X_2 = 1, X_3 = 0)$$
$$= P(X_1 = 0, X_2 = 0, X_3 = 1)$$

i.e., to show $P(R_1^c \cap R_2 \cap R_3) = P(R_1 \cap R_2^c \cap R_3)$
$$= P(R_1 \cap R_2 \cap R_3^c),$$

Where $R_i \equiv$ ith ball is red. (This has ~~already~~ already been shown)

197.5

Note that $\forall \ n \geq 1, \quad S_n = X_1 + X_2 + \cdots + X_n.$

$$\Rightarrow \ E(S_n) = \sum_{i=1}^{n} E(X_i) = n E(X_1) \quad \text{(by exchangeability)}$$

$$= n \frac{b}{b+r}.$$

$$Var(S_n) = \sum_{i=1}^{n} Var(X_i) + 2 \sum\sum_{1 \leq i < j \leq n} Cov(X_i, X_j)$$

$$\overset{\text{exch}}{=} n \ Var(X_1) + 2 \binom{n}{2} Cov(X_1, X_2)$$

$$X_1 \sim Ber\left(\frac{b}{b+r}\right) \quad \Rightarrow \quad Var(X_1) = \frac{br}{(b+r)^2}.$$

$$Cov(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$$

$$= P(R_1^c \cap R_2^c) - \frac{b^2}{(b+r)^2}$$

$$= \frac{b}{b+r} \cdot \frac{b+c}{b+r+c} - \frac{b^2}{(b+r)^2}$$

$$\Rightarrow \quad Var(S_n) = \frac{nbr}{(b+r)^2} + n(n-1)\left[\frac{b(b+c)}{(b+r)(b+r+c)} - \frac{b^2}{(b+r)^2}\right].$$

⟨2⟩ Ex: Suppose r distinguishable balls are arranged at random in n distinguishable boxes. Let N denote the number of empty boxes. Compute $E(N)$ and $Var(N)$.

Ans: Let $I_j = I_{(j^{th} \text{ box is empty})}$, $j = 1, 2, \ldots, n$.

Then $N = \sum_{j=1}^{n} I_j$.

Note: Since balls are arranged at random, $(I_1, I_2, \ldots, I_n)$ is exchangeable.

$$\Rightarrow \quad E(N) = \sum_{j=1}^{n} E(I_j) = n E(I_1)$$

$$= n \cdot \frac{(n-1)^r}{n^r} = \frac{(n-1)^r}{n^{r-1}}$$

$$Var(N) = \sum_{j=1}^{n} Var(I_j) + 2 \sum_{1 \le i < j \le n} \sum Cov(I_i, I_j)$$

$$= n \, Var(I_j) + 2 \binom{n}{2} Cov(I_1, I_2)$$

$$= n \frac{(n-1)^r}{n^r} \left(1 - \frac{(n-1)^r}{n^r}\right) + n(n-1) \, Cov(I_1, I_2)$$

$$\text{Cov}(I_1, I_2) = E(I_1 I_2) - E(I_1)E(I_2)$$

$$= \frac{(n-2)^r}{n^r} - \frac{(n-1)^r}{n^r} \cdot \frac{(n-1)^r}{n^r}$$

$$\Rightarrow \quad \text{Var}(N) = \frac{(n-1)^r}{n^{r-1}} \left(\frac{n^r - (n-1)^r}{n^r} \right) \cancel{+ n(n}$$

$$+ n(n-1)\left[\left(\frac{n-2}{n}\right)^r - \left(\frac{n-1}{n}\right)^{2r} \right]$$

Exc: Suppose n distinguishable flags are arranged at random on r distinguishable poles. Let N denote the number of empty poles. Calculate $E(N)$ and $\text{Var}(N)$.

Another Application: $X \sim Hyp(N, M, n) \Rightarrow Var(X) = ?$

N items $\equiv$ M of Type I + (N − M) of Type II

~~n distinct items~~ Choose a random sample of size n WOR.

X = No. of items of Type I in the sample.

Define $X_i = I_{(\text{ith } \cancel{\text{sample}}\ \cancel{\text{point}} \text{ selected item is of Type I})}$ — A_i

$= \begin{cases} 1 & \text{if ith selected item is of Type I,} \\ 0 & \text{otherwise.} \end{cases} = I_{A_i}.$

Then $X = \sum_{i=1}^{n} X_i$.

$\Rightarrow \quad Var(X) = \sum_{i=1}^{n} V(X_i) + 2 \sum\sum_{i<j} Cov(X_i, X_j)$

$I_{A_i} \sim Ber(P(A_i))$

$\Rightarrow \quad Var(X_i) = P(A_i)(1 - P(A_i))$

$= \frac{M}{N}\left(1 - \frac{M}{N}\right) = \frac{M(N-M)}{N^2}$

$Cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j)$

$= E(I_{A_i} I_{A_j}) - E(I_{A_i})E(I_{A_j})$

$= P(A_i \cap A_j) - P(A_i)P(A_j)$

$= P(A_i \cap A_j) - \frac{M^2}{N^2}.$

$$P(A_i \cap A_j) = P(i^{th} \text{ and } j^{th} \text{ seclected item are both of Type I})$$

$$= \sum_{\substack{l_1=1 \\ l_1 \neq l_2}}^{M} \sum_{l_2=1}^{M} P(i^{th} \text{ selected item is } l_1 \text{ and } j^{th} \text{ selected item is } l_2)$$

$$= \frac{M(M-1)}{N(N-1)} \quad [\text{By exchangeability}]$$

Therefore, $$Var(X) = n\frac{M(N-M)}{N^2} + n(n-1)\left[\frac{M(M-1)}{N(N-1)} - \frac{M^2}{N^2}\right]$$

$$= \frac{nM(N-M)}{N^2}\left(1 - \frac{n-1}{N-1}\right) = \frac{n(N-n)M(N-M)}{N^2(N-1)}.$$

__Exc:__ Recall Polya's Urn scheme. Let $X_i = I_{(i^{th} \text{ selected ball is black})}, i=1,2,\dots,n$. Then $(X_1, X_2, \dots, X_n)$ are exchangeable.

__Exc:__ Let S_n denote the total number of black balls drawn in the first n drawings of a Polya's urn scheme. Calculate $E(S_n)$ and $Var(S_n)$.

Covariance is a useful quantity because (a) it measures the linear asscoation between 2 r.vs and (b) it helps in the computation of variance. However, its value depends on the unit. In order to overcome this problem we define a new quantity as follows:

__Def^n:__ Let X, Y be two nondegenerate r.v.s defined on the same prob space and having finite variance. Then the correlation coefficient of X and Y is defined by

$$\text{Corr}(X,Y) = \rho(X,Y) = \rho_{X,Y} := \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}} = \frac{\text{Cov}(X,Y)}{\sigma_X\,\sigma_Y}.$$

Note: $\text{Var}(X),\ \text{Var}(Y) < \infty \Rightarrow E(X^2), E(Y^2) < \infty$,

$\Rightarrow \text{Cov}(X,Y)$ exists

Also X, Y nondegenerate $\Rightarrow \text{Var}(X),\ \text{Var}(Y) > 0$

Therefore $\rho(X,Y)$ is well-defined.

Exc: $\rho(aX+b, cY+d) = \text{sgn}(ac)\,\rho(X,Y)$. In part, it does not depend on unit.

Thm: (i) For two nondegenerate r.v.s X and Y defined on the same prob space and having finite variance,

$-1 \leq \rho_{X,Y} \leq 1$.

(ii) $\rho_{X,Y} = 1$ if and only if $\exists\ a, b \in \mathbb{R},\ a>0$ such that $P[Y = aX+b] = 1$.

(iii) $\rho_{X,Y} = -1$ if and only if $\exists\ a, b \in \mathbb{R},\ a<0$ such that $P[Y = aX+b] = 1$.