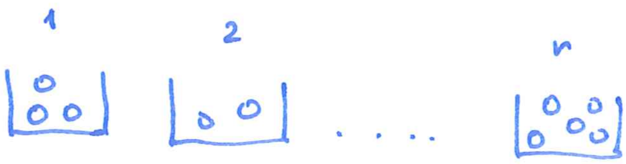
3. Multivariate distributions

3.1. Discrete multivariate distributions

Example: Suppose we have r boxes. We are dropping balls into these boxes at random. The probabilities that we hit box k is p_k for $k = 1, 2, \ldots, r$; Assume the subsequent drops are independent. There are n balls.

Figure:



We end up with random numbers of balls in boxes. Denote these random numbers by $X_1, X_2, \ldots, X_r$. These random numbers will „in the collective" equal to $k_1, k_2, \ldots, k_r$ where $k_i \geq 0$ and $\sum_{i=1}^{r} k_i = n$. All the random variables $X_1, X_2, \ldots, X_r$ simultaneously take a collection of values. The mathematical objects with several components are vectors.

By analogy we will say that

$\underline{X} = (X_1, X_2, \dots, X_r)$ is a random vector.

The possible values of this random vector are vectors $(k_1, k_2, \dots, k_r)$ with $k_i \geq 0$ and $\sum_{i=1}^{r} k_i = n$.

For discrete random variables we had that the distribution was given by $P(X = x)$ for all possible x. By analogy the distribution of the random vector $\underline{X}$ will be given by probabilities $P(\underline{X} = \underline{x})$ where $\underline{x}$ are possible collections/vectors of values. In the above example we need to compute

$$P(\underline{X} = (k_1, k_2, \dots, k_r)) = P(\underbrace{X_1 = k_1, X_2 = k_2, \dots, X_r = k_r}_{})$$

This notation means

$$\bigcap_{i=1}^{r} \{X_i = k_i\}$$

If we want to hit box 1, k_1 times, box 2 k_2 times, ... the possible disjoint ways for this to happen is to get a sequence of hits

$$n_1, n_2, \ldots\ldots, n_n$$

where k_1 of the $n_1, n_2, \ldots, n_n$ are equal to 1, k_2 are equal to k_2, ...
The probability of such a sequence of hits is by <u>independence</u>

$$p_1^{k_1} \cdot p_2^{k_2} \cdots p_r^{k_r}$$

How many sequences of this type are there? We have n positions

$$\overset{1}{\sqcup}\ \overset{2}{\sqcup}\ \sqcup \ \ldots\ldots\ldots \ \overset{n}{\sqcup} \quad \leftarrow \text{Positions}$$

We first choose k_1 positions for 1s. We can do this in $\binom{n}{k_1}$.

Among the $n-k_1$ positions left we choose k_2 positions for 2s. We can do this in $\binom{n-k_1}{k_2}$ wayse.

By the fundamental theorem of combinatorics the total number of possibilities is

$$\binom{n}{k_1}\binom{n-k_1}{k_2}\cdots\binom{n-k_1-\ldots-k_{r-1}}{k_r}$$

$$= \frac{n!}{k_1!\,(n-\cancel{k_1})!}\cdot\frac{(n-\cancel{k_1})!}{k_2!\,(n-\cancel{k_1-k_2})!}\cdots\frac{(n-\cancel{k_1}-\ldots-k_r)!}{k_r!\cdot 0!}$$

$$= \frac{n!}{k_1!\cdot k_2!\cdots k_r!}$$

All the sequences are disjoint events with the same probabilities. It follows

$$P(X_1=k_1,\ldots,X_r=k_r) = \frac{n!}{k_1!\cdots k_r!}\,p_1^{k_1}\cdots p_r^{k_r}$$

for $k_i \geq 0$ for $i=1,2,\ldots,r$ and $\sum_{i=1}^{r} k_i = n$.

Definition: For a vector with the above distribution we say that it has the multinomial distribution with parameters n and $\underline{p} = (p_1, p_2, \ldots, p_r)$.

Shorthand: $\underline{X} \sim \text{Multinom}(n, \underline{p})$.

Definition: A discrete random vector $\underline{X} = (X_1, X_2, \ldots, X_r)$ is a function $\underline{X} : \Omega \to \{\underline{x}_1, \underline{x}_2, \ldots, \ldots\}$ where $\{\underline{x}_1, \underline{x}_2, \ldots\}$ is a finite or countable set of possible values, and such that all components $X_1, X_2, \ldots, X_r$ are random variables.

Definition: The distribution of a random vector $\underline{X}$ with values in $\{\underline{x}_1, \underline{x}_2 \ldots\}$ is given by probabilities $P(\underline{X} = \underline{x}_k)$ for all $k = 1, 2, \ldots$.

Remark: Typically we will write

$P(X_1 = x_1, \dots, X_r = x_r)$. When the number of components is small we often write $P(X = x, Y = y)$ or $P(X = x, Y = y, Z = z)$.

Example: Let $N \geq 3$. Choose three numers at random ~~fr~~ from $\{1, 2, \dots, N\}$ without replacement so that all subsets of three numbers are equally likely. Let X be the smallest of the three numbers, Z the largest and Y the remaining one.

Example: If we choose $5, 3, 7$ ~~wthe~~ we have $X = 3$, $Y = 5$, $Z = 7$.

What is the distribution of (X, Y, Z)?

The possible values are triplets (i, j, k) with $1 \leq i < j < k \leq N$.

We have

$$P(X=i, Y=j, Z=k)$$

$$= P(\text{we select the subset } \{i,j,k\})$$

$$= \frac{1}{\binom{N}{3}}$$

What is the distribution of X?

It has possible values $1, 2, \dots, N-2$.

We notice $\{X=i\} = \bigcup_{i<j<k\leq N} \{X=i, Y=j, Z=k\}$

$$P(X=i) = \sum_{i<j<k\leq N} P(X=i, Y=j, Z=k)$$

$$= \frac{\binom{N-i}{2}}{\binom{N}{3}}$$

$$= \frac{\frac{(N-i)(N-i-1)}{2}}{\frac{N(N-1)(N-2)}{6}}$$

$$= \frac{3(N-i)(N-i-1)}{N(N-1)(N-2)}$$

Definitions:

(i) The distributions of components of a random vector $\underline{X} = (X_1, X_2, \dots, X_r)$ are called univariate marginal distributions.

(ii) The distributions of subvectors like $(X_1, X_2, \dots, X_s)$ for $s < r$ are called multivariate marginal distributions.

Example (continuation): What is the distribution of (X, Y)? We write

$$\{X = i, Y = j\} = \underbrace{\bigcup_{k=j+1}^{N} \{X = i, Y = j, Z = k\}}_{\text{disjoint events}}$$

We have

$$P(X = i, Y = j) = \sum_{k=j+1}^{N} P(X = i, Y = j, Z = k)$$

$$= \frac{N - j}{\binom{N}{3}}$$

for $1 \leq i < j < N$.

If $\underline{X} = (X_1, \dots, X_r)$ is a random vector let us write

$$\underline{X}^1 = (X_1, \dots, X_s) \quad \text{and} \quad \underline{X}^2 = (X_{s+1}, \dots, X_r).$$

<u>Theorem 3.1</u> : Let $R = \{\underline{x}_1, \underline{x}_2 \dots\}$ be the set of possible values of $\underline{X}$.

The marginal distribution of $\underline{X}^1$ is given by

$$P(\underline{X} = \underline{x}^1) = \sum_{(\underline{x}^1, \underline{x}^2) \in R} P(\underline{X} = (\underline{x}^1, \underline{x}^2))$$

$$= \sum_{(\underline{x}^1, \underline{x}^2) \in R} P(\underline{X}^1 = \underline{x}^1, \underline{X}^2 = \underline{x}^2)$$

<u>Proof</u> ! We write

$$\{\underline{X}^1 = \underline{x}^1\} = \underbrace{\bigcup_{(\underline{x}^1, \underline{x}^2) \in R} \{\underline{X}^1 = \underline{x}^1, \underline{X}^2 = \underline{x}^2\}}_{\text{disjoint union.}}.$$

It follows that

$$P(\underline{X}^1 = \underline{x}^1) = \sum_{(\underline{x}^1, \underline{x}^2) \in R} P(\underline{X}^1 = \underline{x}^1, \underline{X}^2 = \underline{x}^2).$$

Independence

For two events A and B we say that they are independent if $P(A \cap B) = P(A) \cdot P(B)$. We would like to define independence for random variables. If X and Y are to be independent we expect the events $\{X = x\}$ and $\{Y = y\}$ to be independent. So we need $P(X = x, Y = y) = P(X = x) P(Y = y)$. This is the right intuition. For the formal definition we generalize to

$$
\begin{aligned}
P(X \in A, Y \in B) &= \sum_{(x,y) \in A \times B} P(X = x, Y = y) \\
&= \sum_{(x,y) \in A \times B} P(X = x) P(Y = y) \\
&= \Big(\sum_{x \in A} P(X = x) \Big) \Big(\sum_{y \in B} P(Y = y) \Big) \\
&= P(X \in A) \cdot P(Y \in B).
\end{aligned}
$$

Definitions:

(i) Discrete random variables X and Y are independent if

$$P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$$

for any two sets A and B.

(ii) Random variables $X_1, X_2, \dots, X_r$ are independent if

$$P(X_1 \in A_1, \dots, X_r \in A_r) = P(X_1 \in A_1) \cdots P(X_r = A_r)$$

for any sets $A_1, A_2, \dots, A_r$.

Remark: The second definition is equivalent to saying that

$$P(X_1 = x_1, X_2 = x_2, \dots, X_r = x_r) =$$

$$= P(X_1 = x_1)\, P(X_2 = x_2) \cdots P(X_r = x_r)$$

for all possible values $(x_1, \dots, x_r)$ of $\underline{X} = (X_1, \dots, X_r)$.

Example: Let $\underline{X} \sim \text{Multinom}(n, \underline{p})$.
We can easily guess that

$$X_k \sim \text{Bin}(n, p_k) \quad \text{for } k = 1, 2, \ldots, r.$$

So

$$P(X_1 = k_1) \cdots P(X_r = k_r)$$

$$= \binom{n}{k_1} p_1^{k_1} (1-p_1)^{n-k_1} \cdots \binom{n}{k_r} p_r^{k_r} (1-p_r)^{n-k_r}$$

and

$$P(X_1 = k_1, \ldots, X_r = k_r) = \frac{n!}{k_1! \cdots k_r!} p_1^{k_1} \cdots p_r^{k_r}.$$

Since $P(X_1 = k_1, \ldots, X_r = k_r) \neq P(X_1 = k_1) \cdots P(X_r = k_r)$
there is no independente.

Example: Suppose the number ~~to~~ of children in a family is Poisson with parameter $\lambda > 0$. Suppose all children are boys or girls with equal probability independently of each other. Let X be the number of boys and Y the number of girls.

We compute with $N = X + Y$

$$P(X=k, Y=\ell) = P(X=k, Y=\ell, N=k+\ell)$$

$$= P(X=k, Y=\ell \mid N=k+\ell)\, P(N=k+\ell)$$

$$= \binom{k+\ell}{k} \left(\tfrac{1}{2}\right)^{k+\ell} \cdot e^{-\lambda} \cdot \frac{\lambda^{k+\ell}}{(k+\ell)!}$$

$$= e^{-\lambda/2} \cdot \frac{(\lambda/2)^k}{k!} \cdot e^{-\lambda/2} \cdot \frac{(\lambda/2)^\ell}{\ell!}$$

On the other hand

$$P(X=k) = \sum_{n=k}^{\infty} P(X=k, N=n)$$

$$= \sum_{n=k}^{\infty} P(X=k \mid N=n)\, P(N=n)$$

$$= \sum_{n=k}^{\infty} \binom{n}{k} \left(\tfrac{1}{2}\right)^n \cdot e^{-\lambda} \cdot \frac{\lambda^n}{n!}$$

$$= \sum_{n=k}^{\infty} \frac{n!}{k!\,(n-k)!} \left(\tfrac{1}{2}\right)^n \cdot e^{-\lambda} \frac{\lambda^n}{n!}$$

$$= \frac{e^{-\lambda} \lambda^k (1/2)^k}{k!} \underbrace{\sum_{n=k}^{\infty} \frac{(\lambda/2)^{n-k}}{(n-k)!}}_{e^{\lambda/2}}$$

$$= \frac{e^{-\lambda/2} (\lambda/2)^k}{k!}$$

We have (the same calculation is valid for girls)

$$P(X=k, Y=e) = P(X=k) P(Y=e)$$

so X, Y are independent.

Theorem 3.2 : Suppose X, Y are discrete random variables with values in $\{x_1, x_2, \dots\}$ and $\{y_1, y_2, \dots\}$. Suppose we have

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

for all pairs $(x,y) \in \{x_1, x_2 \dots\} \times \{y_1, y_2 \dots\}$ for some functions $f : \{x_1, x_2 \dots\} \to \mathbb{R}$ and $g : \{y_1, y_2, \dots\} \to \mathbb{R}$. Then X and Y are independent.

Proof: By Theorem 3.1 the marginal distribution are

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

$$= \sum_y f(x) g(y)$$

$$= f(x) \cdot \underbrace{\sum_y g(y)}_{= c_1}$$

Similarly

$$P(Y=y) = c_2 \cdot g(y).$$

It follows that

$$P(X=x, Y=y) = \cancel{P(X=x)}$$

$$= f(x) g(y)$$

$$= \frac{P(X=x)}{c_1} \cdot \frac{P(Y=y)}{c_2}$$

To finish the proof we need $c_1 c_2 = 1$.

But

$$\sum_{x,y} P(X=x, Y=y) = 1 \quad \text{and}$$

$$\sum_{x,y} P(X=x)P(Y=y)$$

$$= \left(\sum_x P(X=x)\right)\left(\sum_y P(Y=y)\right)$$

$$= 1 \cdot 1 .$$

Summing up we get

$$\sum_{x,y} P(X=x, Y=y) = \frac{1}{c_1 c_2} \sum_{x,y} P(X=x)P(Y=y)$$

or $1 = \frac{1}{c_1 \cdot c_2} \cdot 1 \Rightarrow c_1 \cdot c_2 = 1 .$

<u>Definition</u>: Random vectors $\underline{X}$ and $\underline{Y}$ are independent if

$$P(\underline{X} \in A, \underline{Y} \in B) = P(\underline{X} \in A) \cdot P(\underline{Y} \in B).$$

for all sets A, B.

Remark: The definition is equivalent to

$$P(\underline{X}=x, \underline{Y}=y) = P(\underline{X}=x)\,P(\underline{Y}=y)$$

for all pairs of possible values.

Theorem 3.2 is valid in the following form:

If $P(\underline{X}=x, \underline{Y}=y) = f(x)\,g(y)$ for some functions f, g then $\underline{X}, \underline{Y}$ are independent.

3.2. Expected value

Example: In one of on-line games you have 12 tickets

$$\boxed{1}\ \boxed{1}\ \boxed{1}\ \boxed{1}\ \boxed{2}\ \boxed{2}\ \boxed{3}\ \boxed{D}\ \boxed{S}\ \boxed{S}\ \boxed{S}\ \boxed{S}$$

The tickets are turned around and randomly permuted. The player sees

The player then turns tickets from left to right until the ticket

$\boxed{S}$ = STOP appears. Example:

$\boxed{1}$ $\boxed{2}$ $\boxed{D}$ $\boxed{1}$ $\boxed{S}$

The payout is the sum of all numbers, multiplied by 2 if $\boxed{D}$ = double appears among the tickets. In the above example the payout is 8.

What is the fair price for this game?

Suppose we played this game many times. We can interpret the payout as a random variable, X say. Possible values of X are $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 14, 16, 18, 20, 22\}$.

We have denoting possible values of X by $\{x_1, x_2, \ldots\ldots, x_{17}\}$:

$$\frac{v_1 + \cdots + v_n}{n}$$

$$= \sum_{k=1}^{17} x_k \cdot \underbrace{\frac{\#\text{ of occurrencies of } x_k}{n}}_{\approx P(X = x_k)}$$

So the „long term" average will be

$$\sum_{k=1}^{17} x_k P(X = x_k)$$

We will call this average the expected value of a random variable.

<u>Definition</u>: Let X be a discrete random variable with values $\{x_1, x_2, \ldots\}$. The <u>expected value</u> $E(X)$ is defined as

$$\boxed{E(X) = \sum_{x_k} x_k P(X = x_k)}$$

<u>Technical note</u>: We say that X exist if the sum

$$\sum_{x_k} |x_k| \cdot P(X = x_k)$$ converges.

If f is a function then $Y = f(X)$ is again a discrete random variable. If we „repeat" X we also „repeat" Y. The expectation $E(Y)$ will be approximately

$$\frac{f(v_1) + \dots + f(v_n)}{n} \approx \sum_{x_k} f(x_k) P(X = x_k)$$

by exactly the same argument as before. Formally, we state:

<u>Theorem 3.3</u>: If X is a discrete random variable with values in $\{x_1, x_2, \dots\}$. Let $f : \{x_1, \dots, \dots\} \to \mathbb{R}$.

We have

$$E[f(X)] = \sum_{x_k} f(x_k) P(X = x_k)$$

Proof: Denote $Y = f(X)$. Possible values are $\{y_1, y_2, \ldots\}$. By definition

$$
\begin{aligned}
E[f(X)] &= E(Y) \\
&= \sum_{y_\ell} y_\ell \, P(Y = y_\ell) \\
&= \sum_{y_\ell} y_\ell \sum_{\{x_k : f(x_k) = y_\ell\}} P(X = x_k) \\
&= \sum_{y_\ell} \sum_{\{x_k : f(x_k) = y_\ell\}} f(x_k) \, P(X = x_k) \\
&= \sum_{x_k} f(x_k) \, P(X = x_k)
\end{aligned}
$$

Technical note: We say that $E(f(X))$ exists if the sum

$$\sum_{x_k} |f(x_k)| \, P(X = x_k)$$

exists.

Examples:

(i) Let $X \sim \text{Bin}(n, p)$. We compute

$$E(X) = \sum_{k=0}^{n} k \cdot P(X = k)$$

$$= \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} \qquad q := 1-p$$

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

$$= n \cdot p \underbrace{\sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)}}_{= (p+q)^{n-1} = 1}$$

$$= n \cdot p$$

Similarly

$$E(X^2) = \sum_{k=0}^{n} k^2 \cdot P(X = k)$$

$$= \sum_{k=1}^{n} [k(k-1) + k] \binom{n}{k} p^k q^{n-k}$$

$$= \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k q^{n-k}$$

$$+ \underbrace{\sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k}}_{= n \cdot p}$$

$$= \sum_{k=2}^{n} n(n-1)\binom{n-2}{k-2} \cdot p^2\, p^{k-2}\, q^{(n-2)-(k-2)}$$

$$+ np$$

$$= n(n-1)p^2 \underbrace{\sum_{k=2}^{n} \binom{n-2}{k-2} p^{k-2} q^{(n-2)-(k-2)}}_{=(p+q)^{n-2} = 1}$$

$$+ np$$

$$= n(n-1)p^2 + np$$

$$= n^2p^2 + npq$$

(ii) Let $X \sim NegBin(m, p)$.

We have

$$P(X=k) = \binom{k-1}{m-1} p^m q^{k-m}$$

for $k = m, m+1, \ldots$. We compute

$$E(X) = \sum_{k=m}^{\infty} k \cdot \binom{k-1}{m-1} p^m \cdot q^{k-m}$$

$$= \sum_{k=m}^{\infty} \binom{(k+1)-1}{(m+1)-1} \cdot m \cdot p^m q^{k-m}$$

$$= \sum_{k=m}^{\infty} \frac{m}{p} \cdot \binom{(k+1)-1}{(m+1)-1} p^{m+1} q^{(k+1)-(m+1)}$$

$$= \frac{m}{p} \cdot \underbrace{\sum_{k=m}^{\infty} \binom{(k+1)-1}{(m+1)-1} p^{m+1} q^{(k+1)-(m+1)}}$$

= 1, because this is the sum of all probabilities in the NegBin(m+1, p) distribution

$$= \frac{m}{p}.$$

In a similar way we find that

$$E(X^2) = \sum_{k=m}^{\infty} k^2 \binom{k-1}{m-1} p^m \cdot q^{k-m}$$

$$= \sum_{k=m}^{\infty} [k(k+1) - k] \binom{k-1}{m-1} p^m q^{k-m}$$

$$= \sum_{k=m}^{\infty} \binom{(k+2)-1}{(m+2)-1} \frac{m(m+1)}{p^2} p^{m+2} q^{k-m}$$

$$- \frac{m}{p}$$

$$= \frac{m(m+1)}{p^2} - \frac{m}{p}$$

$$= \frac{m^2}{p^2} + \frac{m}{p}\left(\frac{1}{p} - 1\right)$$

$$= \frac{m^2}{p^2} + \frac{m \cdot q}{p^2}$$

(iii) Let $X \sim$ HiperGeom(n, B, N).

Let us agree that $\binom{a}{b} = 0$ if $b > a$ or $b < 0$. We compute

$$E(X) = \sum_k k \cdot \frac{\binom{B}{k}\binom{R}{n-k}}{\binom{N}{n}}$$

$$= \sum_k \frac{B\binom{B-1}{k-1} \cdot \binom{R}{(n-1)-(k-1)} \cdot n}{\binom{N-1}{n-1} \cdot N}$$

$$= n \cdot \frac{B}{N} \cdot \underbrace{\sum_k \frac{\binom{B-1}{k-1}\binom{R}{(n-1)-(k-1)}}{\binom{N-1}{n-1}}}_{\substack{= 1, \text{ because} \\ \text{this is the sum of} \\ \text{all probs. in} \\ \text{HiperGeom}(n-1, B-1, N-1) \\ \text{distribution.}}}$$

$$= n \cdot \frac{B}{N}$$

The most important theoretical property of expectation is linearity.

<u>Theorem 3.4</u> : Let X, Y be discrete random variables. We have

$$\boxed{E(aX+bY) = aE(X) + bE(Y)}$$

<u>Proof</u> : Denote $Z = aX + bY$. Z is a discrete random variable with values $\{z_1, z_2, \ldots\}$. We have

$$E(Z) = \sum_{z_m} z_m \cdot P(Z = z_m)$$

$$= \sum_{z_m} z_m \cdot \sum_{\{(x_k, y_l) : ax_k + by_l = z_m\}} P(X = x_k, Y = y_l)$$

$$= \sum_{z_m} \sum_{-\text{''}-} (ax_k + by_l) \, P(-\text{''}-)$$

$$= \sum_{x_k, y_\ell} (a x_k + b y_\ell) P(X = x_k, Y = y_\ell)$$

$$= a \cdot \sum_{x_k, y_\ell} x_k P(X = x_k, Y = y_\ell)$$

$$+ b \sum_{x_k, y_\ell} y_\ell P(X = x_k, Y = y_\ell)$$

$$= a \cdot \sum_{x_k} x_k P(X = x_k)$$

$$+ b \cdot \sum_{y_\ell} y_\ell P(Y = y_\ell)$$

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

Technical note : We assume that $E(X)$ and $E(Y)$ exist. In this case $E(aX + bY)$ exist as well.

Remark : We have derived that

$$E(X) = \sum_{x_k, y_\ell} x_k P(X = x_k, Y = y_\ell)$$

A consequence of Theorem 3.4 is that linearity is valid for more general linear combinations.

If $X_1, X_2, \ldots, X_r$ are random variables such that $E(X_k)$ exists then

$$\boxed{E\left[\sum_{k=1}^{r} a_k X_k\right] = \sum_{k=1}^{r} a_k E(X_k).}$$

Finally, we state

Theorem 3.5: Let $\underline{X}$ be a discrete random vector in $\mathbb{R}^n$ and let $f: \mathbb{R}^n \to \mathbb{R}$ be a function. We have

$$E[f(\underline{X})] = \sum_{\underline{x}_k} f(\underline{x}_k) P(\underline{X} = \underline{x}_k)$$

Proof: The proof is identical to the proof of Theorem 3.3.

Example: Let $\underline{X} \sim \text{Multinomial}(n, \underline{p})$. What is $E(X_k \cdot X_e)$? We know that

$X_k + X_e \sim \text{Bin}(n, p_k + p_e)$ so

$$E[(X_k + X_e)^2] = n(p_k + p_e)(1 - p_k - p_e) + n^2 (p_k + p_e)^2$$

$$\parallel$$

$$E[X_k^2 + 2X_k X_e + X_e^2]$$

$$\parallel$$

$$E(X_k^2) + 2E(X_k X_e) + E(X_e^2)$$

$$= n p_k (1 - p_k) + n^2 p_k^2$$

$$+ 2 E(X_k X_\ell)$$

$$+ n p_\ell (1 - p_\ell) + n^2 p_\ell^2$$

This is an equation for $E(X_k X_\ell)$ from which we compute

$$E(X_k X_\ell) = - n p_k p_\ell + n^2 p_k p_\ell$$

<u>Definition</u>: A random variable X with values in $\{0,1\}$ is called an indicator or a Bernoulli random variable. We denote $p = P(X=1)$

<u>Shorthand</u>: $X \sim \text{Bernoulli}(p)$.

By definition

$$E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) = p$$

<u>Remark</u>: Since $X : \Omega \to \{0,1\}$ we can denote $A = \{X = 1\}$ which is an event.

Every indicator is associated with an event A. We will write I_A or 1_A for the indicator of A i.e. the random variable X, for which $X(\omega) = 1$ if $\omega \in A$ and 0 else.

In many cases complicated random variables can be written as linear combinations of ~~more~~ ~~Complicate~~ simpler random variables. Expectations can then be computed in simpler ways using linearity.

Example: Let us return to the first example.

$$\overset{1}{\boxed{1}}\ \overset{2}{\boxed{1}}\ \overset{3}{\boxed{1}}\ \overset{4}{\boxed{1}}\ \overset{1}{\boxed{2}}\ \overset{2}{\boxed{2}}\ \boxed{3}\ \boxed{0}\ \boxed{5}\ \boxed{5}\ \boxed{5}\ \boxed{5}$$

Label the tickets with 1 from 1 to 4, and the tickets with 2 from 1 to 2.

If we know whether a ticket has contributed to the final payout and whether $\boxed{D}$ ~~appeared~~ appeared we can reconstruct the payout.

<u>Example</u>:

	1	2	3	4	1	2						
	$\boxed{1}$	$\boxed{1}$	$\boxed{1}$	$\boxed{1}$	$\boxed{2}$	$\boxed{2}$	$\boxed{3}$	$\boxed{D}$	$\boxed{S}$	$\boxed{S}$	$\boxed{S}$	$\boxed{S}$
	↑	↑	↑	↑	↑	↑	↑	↑				
C[illegible] :	Y	N	N	Y	N	Y	N	Y				

Payout $= 4 \times 2 = 8$

Define events

$A_{1,i} = \{$ ticket $\boxed{1}^{i}$ contributed, but $\boxed{D}$ did not $\}$

$A_{2,i} = \{$ — " — , and $\boxed{D}$ did $\}$

$B_{1,i} = \{$ ticket $\boxed{2}^{i}$ contributed, $\boxed{D}$ did not $\}$

$B_{2,i} = \{$ — " — , $\boxed{D}$ did $\}$

$C_1 = \{$ $\boxed{3}$ contributed, $\boxed{D}$ did not $\}$

$C_2 = \{$ — " — , $\boxed{D}$ did $\}$.

We have

$$X = \sum_{i=1}^{4} 1_{A_{1,i}} + 2 \sum_{i=1}^{4} 1_{A_{2,i}}$$

$$+ 2 \sum_{i=1}^{2} 1_{B_{1,i}} + 4 \sum_{i=1}^{2} 1_{B_{2,i}}$$

$$+ 3\, 1_{C_1} + 6\, 1_{C_2}$$

By symmetry

$$P(A_{1,i}) = P(B_{1,i}) = P(C_1)$$

and

$$P(A_{2,i}) = P(B_{2,i}) = P(C_2).$$

This means that

$$E(X) = 11 \cdot P(A_{1,1}) + 22 \cdot P(A_{2,1})$$

We compute $P(A_{1,1})$ and $P(A_{2,1})$ by noticing that if we only look at tickets

$\boxed{1}$ $\boxed{D}$ $\boxed{S}$ $\boxed{S}$ $\boxed{S}$ $\boxed{S}$ among the 12 permuted tickets they too are randomly permuted. We say that the induced permutation is random. It follows that $A_{1,1}$ happens if we see

$\boxed{1}\ \boxed{3}\ *\ *\ *\ *$

The probability is

$$\frac{1}{6} \times \frac{4}{5} = \frac{2}{15}$$

The event $A_{2,1}$ happens if we see

$\boxed{1}\ \boxed{D}\ *\ *\ *\ *$ or $\boxed{D}\ \boxed{1}\ *\ *\ *\ *$.

The probability is

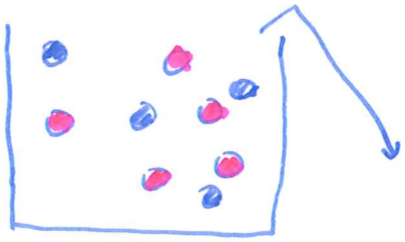
$$2 \cdot \frac{1}{6} \cdot \frac{1}{5} = \frac{1}{15}$$

Finally,

$$E(X) = 11 \cdot \frac{2}{15} + 22 \cdot \frac{1}{15} = \frac{44}{15}$$

$$\doteq 2.93$$

Example : The hyper-geometric distribution is created by selecting balls out of a box.



Select n balls at random

$X = \#$ of black balls

We can imagine that balls are selected one by one at random until we have n balls. Define

$$I_k = \begin{cases} 1, & \text{if the } k\text{-th ball is black.} \\ 0, & \text{else,} \end{cases}$$

for $k = 1, 2, \dots, n$. We have

$$E(X) = E(\underbrace{I_1 + \dots + I_n}_{= X})$$

$$= E(I_1) + \dots + E(I_n)$$

$$= P(I_1=1) + P(I_2=1) + \dots + P(I_n=1)$$

But the k-th ball is equally likely to be any of the N balls. We are asking this question before the selection process begins. This means that

$$P(I_1=1) = P(I_2=1) = \dots = P(I_n=1).$$

But

$$P(I_1=1) = P(\text{first ball selected black})$$

$$= \frac{B}{N}.$$

It follows that

$$E(X) = n \cdot \frac{B}{N}.$$

<u>Comment</u>: The idea to write X as a linear combination of indicators is called the method of indicators.

3.3. Joint continuous distributions

For a continuous random variable X with density f_X we have

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx .$$

More generally, for a set A we can say

$$P(X \in A) = \int_A f_X(x)\,dx$$

$$= \int_{-\infty}^{\infty} f_X(x)\,\chi_A(x)\,dx ,$$

where χ_A is the characteristic function of the set A. This last form has an easy extension to $\mathbb{R}^2$, $\mathbb{R}^3$ or $\mathbb{R}^4$. For $\mathbb{R}^2$ we can say that

$$P(\underline{X} \in A) = \iint_A f_{\underline{X}}(x,y)\,dx\,dy$$

for an appropriate non-negative function. In $\mathbb{R}^3$

we have for $\underline{X} = (X_1, X_2, X_3)$

$$P(\underline{X} \in A) = \iiint_A f_{\underline{X}}(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3 .$$

In probability we will write single integrals even in higher dimensions.
If $A \subseteq \mathbb{R}^n$ we will write

$$\iint_A \cdots \int f(x_1, x_2 \ldots x_n)\, dx_1\, dx_2 \ldots dx_n$$

$$= \int_A f(\underline{x})\, d\underline{x}$$

<u>Definition</u>: If for a random vector $\underline{X}$ we have

$$P(\underline{X} \in A) = \int_A f_{\underline{X}}(\underline{x})\, d\underline{x}$$

for a non-negative function $f_{\underline{X}} : \mathbb{R}^n \to \mathbb{R}$ and all (reasonable) sets A we say that $\underline{X}$ has continuous distribution with density $f_{\underline{X}}$.

<u>Technical note</u>: In more dimensions in general we say that the distribution of $\underline{X}$ is described by probabilities $P(\underline{X} \in A)$ for all reasonable sets $A \subseteq \mathbb{R}^n$.

„Reasonable" means all sets that are formed from open sets by complements, countable unions and countable intersections. Such sets are called Borel sets.

<u>Example</u>: Let (X,Y) be a random vector with density $f_{X,Y}(x,y)$ given by

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}$$

for $\rho \in (-1,1)$. Let us check that $f_{X,Y}$ is a density. This means that it is non-negative and integrates to 1. We know that

$$\frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx = 1,$$

because the latter is the integral of the normal density. We integrate

$$\int_{\mathbb{R}^2} f_{X,Y}(x,y)\,dx\,dy =$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}}\int_{\mathbb{R}^2} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}\,dx\,dy$$

$$= \underbrace{\frac{1}{2\pi\sqrt{1-\rho^2}}\int_{-\infty}^{\infty} dx \cdot \int_{-\infty}^{\infty} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}\,dy}_{\text{This is called Fubini's theorem.}}$$

$$= (*)$$

We note

$$(x^2 - 2\rho xy + y^2)/(1-\rho^2)$$

$$= [(y-\rho x)^2 + (1-\rho^2)x^2]/(1-\rho^2)$$

$$= \frac{(y-\rho x)^2}{1-\rho^2} + x^2$$

$$(*) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx \cdot \underbrace{\int_{-\infty}^{\infty} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}} dy}_{=\sqrt{2\pi}\cdot\sqrt{1-\rho^2}}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx$$

$$= 1.$$

Let us continue with this example and compute $P(X \geq 0, Y \geq 0)$.

In other words

$$P(X \geq 0, Y \geq 0) = P((X,Y) \in [0,\infty)^2).$$

By definition

$$P((X,Y) \in [0,\infty)^2)$$

$$= \int_{[0,\infty)^2} f_{X,Y}(x,y)\, dx\, dy$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{[0,\infty)^2} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}} dx\, dy$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_0^\infty e^{-x^2/2} dx \cdot \underbrace{\int_0^\infty e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}} dy}$$

New variable:

$$\frac{y-\rho x}{\sqrt{1-\rho^2}} = u$$

$$= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_0^\infty e^{-x^2/2} dx \cdot$$

$$\int_{-\frac{\rho x}{\sqrt{1-\rho^2}}}^{\infty} e^{-u^2/2} du \cdot \sqrt{1-\rho^2}$$

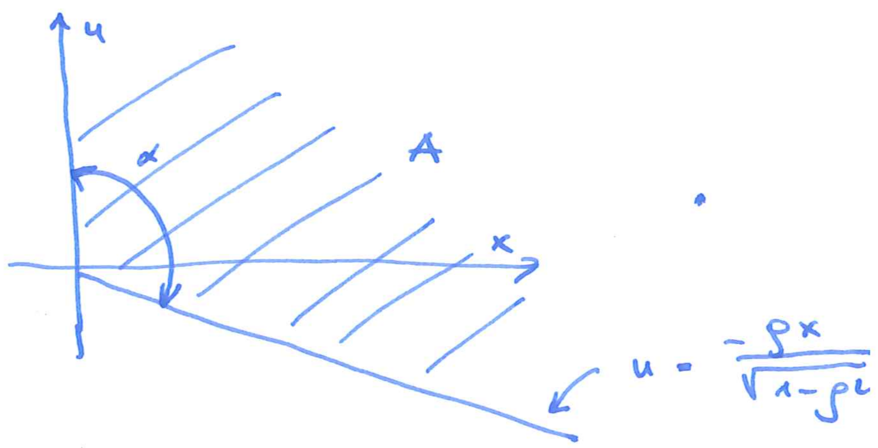
This last integral is the integral of the function

$$f(x,u) = \frac{1}{2\pi} e^{-\frac{x^2+u^2}{2}}$$ over the set

$$A = \{(x,u) : x \geq 0,\ u \geq -\frac{\rho x}{\sqrt{1-\rho^2}}\}$$

by Fubini.

<u>Figure</u> :



We observe:

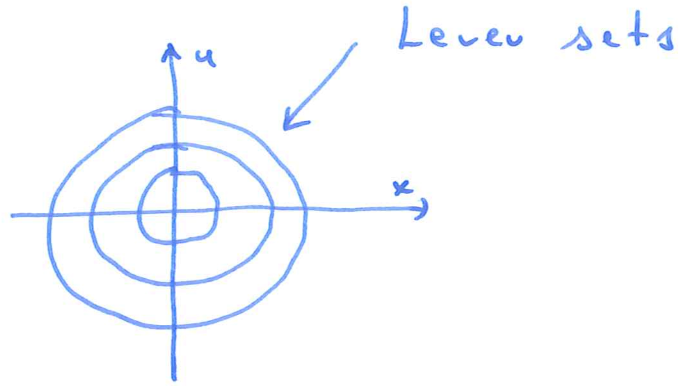
(i) $f(x,u)$ integrates to 1.
We get this by taking $\rho = 0$ in the previous example.

(ii) The function

$f(x,u)$ is rotationally

symmetric.

Figure :

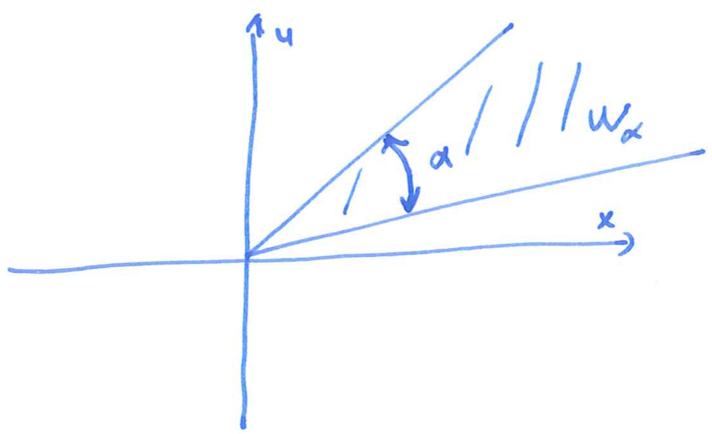


The integral over a wedge

of angle α is proportional

to the angle

Figure :



$$\int_{W_\alpha} f(x,u)\,dx\,du = \frac{\alpha}{2\pi}$$

In our case the angle α equals

$$\alpha = \frac{\pi}{2} + \operatorname{arctg}\left(\frac{\rho}{\sqrt{1-\rho^2}}\right).$$

Finally we have

$$P(X \geq 0, Y \geq 0) =$$

$$= \frac{1}{4} + \frac{1}{2\pi} \operatorname{arctg}\left(\frac{\rho}{\sqrt{1-\rho^2}}\right)$$

Marginal distributions

Suppose (X,Y) has density $f_{X,Y}(x,y)$. What is the density of X?

We know from the ~~1st~~ 2nd Chapter that if for any $a<b$

$$P(a \leq X \leq b) = \int_a^b g(x)\,dx$$ this

implies that $g(x) = f_X(x)$.

we compute

$$P(a \le X \le b) = P(a \le X \le b,\ Y \in \mathbb{R})$$

$$= P((X,Y) \in [a,b] \times \mathbb{R})$$

$$= \int_{[a,b]\times\mathbb{R}} f_{X,Y}(x,y)\, dx\, dy$$

$$= \int_a^b dx \cdot \underbrace{\int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy}_{\text{This is a function of } x \text{, say } g(x)}$$

$$= \int_a^b g(x)\, dx .$$

Theorem 3.6: Let (X,Y) have the density $f_{X,Y}(x,y)$. We have

$$\boxed{\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx \end{aligned}}$$

Proof: Done already.

Comments

(i) The two formulae in Theorem 3.6 are called formulae for marginal densities.

(ii) A rigorous statement must include the assumptions that $x \mapsto f_{X,Y}(x,y)$ and $y \mapsto f_{X,Y}(x,y)$ are Riemann integrable, and that f_X and f_Y are Riemann integrable.

Example :

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2-2xy+y^2}{2(1-\rho^2)}}$$

We have

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

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-x^2/2} \cdot e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}} dy$$

$$= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}} dy}_{}$$

$= 1$, because it is the integral of the $N(\rho x, 1-\rho^2)$ dist

$$= \frac{1}{\sqrt{2\pi}} \cdot e^{-x^2/2}.$$

Conclusion : $X \sim N(0,1)$.

Theorem 3.6 has a more general version.

<u>Theorem 3.6a</u>: Let $(\underline{X}, \underline{Y})$ be a random vector with $\underline{X} \in \mathbb{R}^p$ and $Y \in \mathbb{R}^q$ with density $f_{\underline{X},\underline{Y}}(\underline{x}, \underline{y})$. Then

$$f_{\underline{X}}(\underline{x}) = \int_{\mathbb{R}^q} f_{\underline{X},\underline{Y}}(\underline{x}, \underline{y})\, d\underline{y}$$

$$f_{\underline{Y}}(\underline{y}) = \int_{\mathbb{R}^p} f_{\underline{X},\underline{Y}}(\underline{x}, \underline{y})\, d\underline{x}$$

<u>Proof</u>: Same as before.

<u>Example</u>: If (X, Y, Z) has density $f_{X,Y,Z}(x,y,z)$ then

$$f_{X,Y}(x,y) = \int_{-\infty}^{\infty} f_{X,Y,Z}(x,y,z)\, dz$$

and

$$f_X(x) = \int_{\mathbb{R}^2} f_{X,Y,Z}(x,y,z)\, dy\, dz.$$

Independence

In general we say that X, Y are independent if

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B).$$

If (X, Y) has density $f_{X,Y}(x,y)$ this means that for $A = [a,b]$ and $B = [c,d]$ we have

$$\underbrace{\int_{[a,b]\times[c,d]} f_{X,Y}(x,y)\, dx\, dy}$$

$$= P(X \in [a,b], Y \in [c,d])$$

$$= \underbrace{\left(\int_a^b f_X(x)\, dx \right)}_{P(a \le X \le b)} \cdot \underbrace{\left(\int_c^d f_Y(y)\, dy \right)}_{P(c \le Y \le d)}$$

Fubini.

$$= \int_{[a,b]\times[c,d]} f_X(x) \cdot f_Y(y)\, dx\, dy$$

We borrow a statement from Analysis 2: if for functions $f(x,y)$ and $g(x,y)$ we have:

(i) For all rectangles $Q = [a,b] \times [c,d]$ the functions are Riemann integrable.

(ii)
$$\int_Q f(x,y)\,dx\,dy = \int_Q g(x,y)\,dx\,dy$$
for all Q

then $f(x,y) = g(x,y)$.

Technical note: in fact f, g can differ but only on a set of measure 0.

If X, Y are independent we have

$$\int_Q f_{X,Y}(x,y)\,dx\,dy = \int_Q f_X(x) f_Y(y)\,dx\,dy$$

Theorem 3.7: Let (X,Y) have density $f_{X,Y}(x,y)$. The random variables X and Y are independent if and only if $f_{X,Y}(x,y) = f_X(x) f_Y(y)$.

Proof: If $f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y)$

then
$$\underbrace{P((X,Y)\in A\times B)}_{\int_{A\times B} f_{X,Y}(x,y)\,dx\,dy} = P(X\in A, Y\in B)$$

$$\overset{\text{Fubini}}{=} \underbrace{\int_A f_X(x)\,dx}_{P(X\in A)} \cdot \underbrace{\int_B f_Y(y)\,dy}_{P(Y\in B)}$$

Independence follows.

If X, Y are independent we proved above that $f_{X,Y}(x,y) =$
$= f_X(x)\cdot f_Y(y)$.

Theorem 3.7 has a more general version.

<u>Theorem 3.7a</u> : Let $\underline{X}, \underline{Y}$ be continuous random vectors with density $f_{\underline{X},\underline{Y}}(\underline{x},\underline{y})$. The vectors $\underline{X}$ and $\underline{Y}$ are independent if and only if $f_{\underline{X},\underline{Y}}(\underline{x},\underline{y}) = f_{\underline{X}}(\underline{x}) \cdot f_{\underline{Y}}(\underline{y})$.

<u>Proof</u> : Same as above.

<u>Theorem 3.8</u> : Let (X,Y) have density $f_{X,Y}(x,y)$. If

$$f_{X,Y}(x,y) = g(x)\,h(y)$$

for nonnegative functions g and h then X, Y are independent.

Proof: By the formulae for marginal density we have

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

$$= \int_{-\infty}^{\infty} g(x)\,h(y)\,dy$$

$$= g(x) \cdot \underbrace{\int_{-\infty}^{\infty} h(y)\,dy}_{=c_1}$$

Similarly

$$f_Y(y) = h(y) \cdot \underbrace{\int_{-\infty}^{\infty} g(x)\,dx}_{c_2}$$

It follows

$$f_{X,Y}(x,y) = \frac{f_X(x)}{c_1} \cdot \frac{f_Y(y)}{c_2}$$

We need to prove that $c_1 \cdot c_2 = 1$.

Integrate both sides over $\mathbb{R}^2$.

We get

$$1 = \int_{\mathbb{R}^2} f_{X,Y}(x,y)\, dx\, dy$$

$$= \frac{1}{c_1 c_2} \int_{\mathbb{R}^2} f_X(x) f_Y(y)\, dx\, dy.$$

$$\overset{\text{Fubini}}{=} \frac{1}{c_1 c_2} \int_{-\infty}^{\infty} f_X(x)\, dx \cdot \int_{-\infty}^{\infty} f_Y(y)\, dy$$

$$= \frac{1}{c_1 \cdot c_2} \cdot 1 \cdot 1$$

It follows $c_1 \cdot c_2 = 1$.

Example: Let

$$f_{X,Y}(x,y) = \frac{1}{2\pi \sqrt{1-\rho^2}}\, e^{-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}}$$

We computed

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$ and

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

We see that

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

only when $\rho = 0$.

3.4. Functions of random vectors

Discrete case

In the discrete case we will only look at integrer valued random variables. If X, Y are two such variables then $Z = X + Y$ is an integer valued random variable. We have

$$\{Z = n\} = \underbrace{\bigcup_{k \in \mathbb{Z}} \{X = k, Y = n-k\}}_{\text{disjoint union}}$$

We have

$$\boxed{P(Z = n) = \sum_{k \in \mathbb{Z}} P(X = k, Y = n-k)}$$

Special cases:

(i) if X, Y are non-negative we have

$$P(Z = n) = \sum_{k=0}^{n} P(X = k, Y = n-k)$$

(ii) if X, Y are independent then

$$P(Z = n) = \sum_{k \in \mathbb{Z}} P(X = k) P(Y = n-k)$$

<u>Examples</u>: (i) Let X, Y be independent and $X \sim Po(\mu)$, $Y \sim Po(\lambda)$. Let $Z = X + Y$. By the above formula

$$P(Z = n) = \sum_{k=0}^{n} P(X = k) \cdot P(Y = n-k)$$

$$= \sum_{k=0}^{n} \frac{e^{-\mu} \mu^k}{k!} \cdot \frac{e^{-\lambda} \cdot \lambda^{n-k}}{(n-k)!}$$

$$= \frac{e^{-(\lambda+\mu)}}{n!} \sum_{k=0}^{n} \underbrace{\frac{n!}{k!\,(n-k)!}}_{= \binom{n}{k}} \mu^k \lambda^{n-k}$$

$$= \frac{e^{-(\lambda+\mu)}}{n!} (\lambda+\mu)^n .$$

<u>Conclusion</u>: $Z = X + Y \sim Po(\lambda + \mu)$.

(ii) Let X, Y be independent and have the Pólya distribution.

This means that

$$P(X=k) = \frac{\beta^a (a)_k}{k!\,(1+\beta)^{a+k}} \qquad k=0,1,\dots$$

$$P(Y=\ell) = \frac{\beta^b (b)_\ell}{\ell!\,(1+\beta)^{b+\ell}} \qquad \ell=0,1,\dots$$

Here $(a)_0 = 1$ and

$$(a)_k = a(a+1)\dots(a+k-1)$$

is the Pochhammer symbol.

Let $Z = X+Y$. We are looking for the distribution of Z. By the formula we have

$$P(Z=n) = \sum_{k=0}^{n} P(X=k)P(Y=n-k)$$

$$= \frac{\beta^{a+b}}{(1+\beta)^{a+b+n}} \sum_{k=0}^{n} \frac{(a)_k (b)_{n-k}}{k!\,(n-k)!}$$

$$= \frac{\beta^{a+b}}{(1+\beta)^{a+b+n}\cdot n!} \sum_{k=0}^{n} \binom{n}{k} (a)_k (b)_{n-k}.$$

The last formula is similar to the binomial formula.

To prove it we will use a few facts from Analysis:

(i) The gamma function is defined as

$$\Gamma(x) = \int_0^\infty u^{x-1} \cdot e^{-u} \, du, \qquad x > 0$$

Integration by parts gives

$$\Gamma(x+1) = x\,\Gamma(x) \quad \text{and}$$

as a consequence

$$\Gamma(a+n) = (a+n-1)(a+n-2) \cdot \ldots \cdot a \cdot \Gamma(a)$$

We can write

$$(a)_n = \frac{\Gamma(a+n)}{\Gamma(a)}$$

(ii) The Beta function is defined as

$$B(p,q) = \int_0^1 u^{p-1}(1-u)^{q-1}\,du \quad , \; p,q>0$$

The connection between Γ and B functions is given by Euler:

$$\boxed{B(p,q) = \frac{\Gamma(p)\cdot\Gamma(q)}{\Gamma(p+q)}}$$

We compute

$$\sum_{k=0}^{n} \binom{n}{k} (a)_k (b)_{n-k}$$

$$= \sum_{k=0}^{n} \binom{n}{k} \frac{\Gamma(a+k)}{\Gamma(a)} \cdot \frac{\Gamma(b+n-k)}{\Gamma(b)}$$

$$= \frac{\Gamma(a+b+n)}{\Gamma(a)\Gamma(b)} \sum_{k=0}^{n} \binom{n}{k} \frac{\Gamma(a+k)\,\Gamma(b+n-k)}{\Gamma(a+b+n)}$$

$$= \frac{\Gamma(a+b+n)}{\Gamma(a)\Gamma(b)} \sum_{k=0}^{n} \binom{n}{k} B(a+k, b+n-k)$$

$$= \frac{\Gamma(a+b+n)}{\Gamma(a)\,\Gamma(b)} \sum_{k=0}^{n} \binom{n}{k} \int_0^1 u^{a+k-1} (1-u)^{b+n-k-1} \, du$$

$$= \frac{\Gamma(a+b+n)}{\Gamma(a)\,\Gamma(b)} \int_0^1 u^{a-1} (1-u)^{b-1} \cdot \underbrace{\sum_{k=0}^{n} \binom{n}{k} u^{k} (1-u)^{n-k}}_{=1} \, du$$

$$\overset{\text{def.}}{=} \frac{\Gamma(a+b+n)}{\Gamma(a)\,\Gamma(b)} \, B(a,b)$$

$$\overset{\text{Euler}}{=} \frac{\Gamma(a+b+n)}{\cancel{\Gamma(a)}\,\cancel{\Gamma(b)}} \cdot \frac{\cancel{\Gamma(a)}\,\cancel{\Gamma(b)}}{\Gamma(a+b)}$$

$$= \frac{\Gamma(a+b+n)}{\Gamma(a+b)}$$

$$= (a+b)_n$$

Finally we have

$$P(Z=n) = \frac{\beta^{a+b}\,(a+b)_n}{n!\,(1+\beta)^{a+b+n}}, \quad n = 0, 1, \ldots$$

Example: Suppose X, Y are independent and $X \sim \mathrm{Bin}(m, p)$, $Y \sim \mathrm{Bin}(n, p)$. We expect

$Z = X + Y \sim \mathrm{Bin}(m+n, p)$. The formal proof:

$$P(Z = \ell) = \sum_{k=0}^{\ell} P(X = k, Y = \ell - k)$$

$$= \sum_{k = \max(0, n-\ell)}^{\min(\ell, m)} \binom{m}{k} p^{k} q^{m-k} \binom{n}{\ell - k} p^{\ell - k} q^{n - \ell + k}$$

$$= \sum_{k = \max(0, n-\ell)}^{\min(\ell, m)} \binom{m}{k}\binom{n}{\ell - k} \underbrace{p^{\ell} q^{m+n-\ell}}$$

does not depend on k.

The sum is computed by the following combinatorial argument:

Suppose we need to choose l elements from the union of sets with m and n elements. This can be done in $\binom{m+n}{l}$ ways. We can count in another way: we choose k elements from the first set and $l-k$ from the other. This is possible for $k \geq \max(0, n-l)$ and $l \leq \min(l, m)$. This splits all the choices in disjoint subsets so

$$\binom{m+n}{l} = \sum_{k=\max(0,n-l)}^{\min(l,m)} \binom{m}{k}\binom{n}{l-k}.$$

Finally

$$P(Z = l) = \binom{m+n}{l} p^{l} q^{m+n-l}$$

Continuous case

The most important formula is the transformation formula.

Suppose the vector (X,Y) has density $f_{X,Y}(x,y)$. We form a new vector (U,V) by

$$\Phi(X,Y) = (\Phi_1(X,Y), \Phi_2(X,Y)).$$

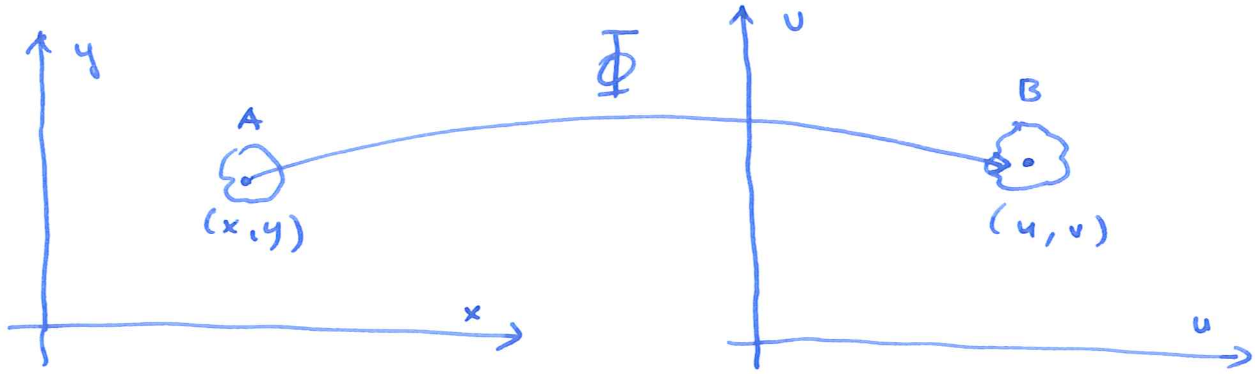
Example:

$$\Phi(x,y) = \left(\frac{x}{x+y},\ x+y\right)$$

$$(U,V) = \left(\frac{X}{X+Y},\ X+Y\right).$$

Question: What is the density $f_{U,V}(u,v)$ of (U,V)?

Idea :



By definition for „small" A and B

$$P((x,y) \in A) \approx f_{X,Y}(x,y)\,|A|$$

$$P((U,V) \in B) \approx f_{U,V}(u,v)\cdot|B|$$

If Φ is bijective then

$$P((X,Y) \in A) = P((U,V) \in B)$$

if $B = \Phi(A)$. So

$$f_{X,Y}(x,y)\,|A| \approx f_{U,V}(u,v)\cdot|B|$$

or

$$f_{U,V}(u,v) \approx f_{X,Y}(x,y)\cdot\frac{|A|}{|B|}$$

But from Analysis 2 we know

$$\frac{|A|}{|B|} \approx |J_{\Phi^{-1}}(u,v)|.$$

<u>Theorem 3.9</u> (transformation formula). Let (X,Y) be a vector with density $f_{X,Y}(x,y)$. Suppose $P((X,Y) \in \mathcal{U}) = 1$ for an open set $\mathcal{U}$. Let $\Phi : \mathcal{U} \to \mathcal{S}$ be a bijective map which is continuously partially differentiable. Let

$$(U,V) = \Phi(X,Y).$$

The ~~the~~ density $f_{U,V}(u,v)$ is

$$f_{U,V}(u,v) = f_{X,Y}(\Phi^{-1}(u,v)) \cdot |J_{\Phi^{-1}}(u,v)|$$

where $J_{\Phi^{-1}}$ is the Jacobian determinant of Φ^{-1}.

Proof: Let $B \subseteq S$. We compute

$$P((U,V) \in B)$$

$$= P((X,Y) \in \Phi^{-1}(B))$$

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

$$= (*)$$

New variable: $(x,y) = \Phi^{-1}(u,v)$

$dx\,dy = |J_{\Phi^{-1}}(u,v)|\,du\,dv$

$$(*) = \int_B f_{X,Y}(\Phi^{-1}(u,v))\,|J_{\Phi^{-1}}(u,v)|\,du\,dv.$$

Comment: We used the formula for a new variable in double integrals.

Example: Let X, Y be independent with $X \sim \Gamma(a, \lambda)$ and $Y \sim \Gamma(b, \lambda)$. This means

$$f_X(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} \cdot e^{-\lambda x}, \quad x > 0$$

$$f_Y(y) = \frac{\lambda^b}{\Gamma(b)} y^{b-1} e^{-\lambda y}, \quad y > 0$$

By independence

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

Let

$$\Phi(x,y) = \left(\frac{x}{x+y}, x+y \right)$$

for $x, y > 0$. We can take $\mathcal{U} = (0,\infty)^2$ and $\mathcal{V} = (0,1) \times (0,\infty)$. Φ is bijective and continuously differentiable. To find Φ^{-1} we need to solve equations

$$\frac{x}{x+y} = u, \qquad x + y = v.$$

We get

$$\begin{aligned} x &= u \cdot v \\ y &= v - x = v - u \cdot v \\ &= v(1-u) \end{aligned}$$

This means

$$\Phi^{-1}(u,v) = (uv, v(1-u)).$$

We compute

$$J\Phi^{-1}(u,v) = \det\begin{pmatrix} v & u \\ -v & 1-u \end{pmatrix}$$

$$= v$$

The density $f_{U,V}(u,v)$ is given by

$$f_{U,V}(u,v) = f_{X,Y}(uv, v(1-u)) \cdot |J\Phi^{-1}(u,v)| =$$

$$= f_X(uv) \cdot f_Y(v(1-u)) \cdot v$$

$$= \frac{\lambda^a}{\Gamma(a)} (uv)^{a-1} \cdot e^{-\lambda uv} \cdot \frac{\lambda^b}{\Gamma(b)} [v(1-u)]^{b-1} \cdot e^{-\lambda v(1-u)} \cdot v$$

$$= \frac{\lambda^{a+b}}{\Gamma(a)\,\Gamma(b)}\, u^{a-1}(1-u)^{b-1}$$

$$\cdot\, v^{a+b-1} \cdot e^{-\lambda v}$$

for $(u,v) \in (0,1) \times (0,\infty)$.

We note:

(i) U, V are independent

(ii) $f_U(u) = \text{const.}\; u^{a-1}(1-u)^{b-1}$

$f_V(v) = \text{const.}\; v^{a+b-1} \cdot e^{-\lambda v}$.

It follows $U = \frac{X}{X+Y} \sim \text{Beta}(a,b)$

and $V = X+Y \sim \Gamma(a+b, \lambda)$.

Example : Suppose (X,Y) has density $f_{X,Y}(x,y)$. Let

$\Phi(x,y) = (x, x+y) = (x,z)$

What is the density $f_{X,Z}(x,z)$?

By the transformation formula

$$f_{X,Z}(x,z) = f_{X,Y}(x, z-x) \cdot |J_{\Phi^{-1}}(x,z)|.$$

But $\Phi^{-1}(x,z) = (x, z-x) \Rightarrow$

$$J_{\Phi^{-1}}(x,z) = 1.$$

We have

$$\boxed{f_{X,Z}(x,z) = f_{X,Y}(x, z-x)}$$

The density of Z is the marginal density of $f_{X,Z}(x,z)$.

We have

$$\boxed{f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,dx}$$

If X, Y are independent we get

$$\boxed{f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\,dx}$$

<u>Comment</u>: The above formula is known as <u>convolution</u> in Analysis.

Example : Let X, Y be independent with $X \sim N(\mu, \sigma^2)$ and $Y \sim N(\nu, \tau^2)$. What is the density of $Z = X + Y$. Assume first $\mu = \nu = 0$ and $\sigma^2 + \tau^2 = 1$. In this case

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx$$

$$= \frac{1}{2\pi \cdot \sigma\tau} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \cdot e^{-\frac{(z-x)^2}{2\tau^2}}\, dx$$

$$= \frac{1}{2\pi\sigma \cdot \tau} \int_{-\infty}^{\infty} e^{-x^2\left[\frac{1}{2\sigma^2} + \frac{1}{2\tau^2}\right]} \cdot e^{\frac{xz}{\tau^2}} \cdot e^{-\frac{z^2}{2\tau^2}}\, dx$$

$$= \frac{1}{2\pi\sigma \cdot \tau} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2\tau^2} + \frac{xz}{\tau^2}} \cdot e^{-\frac{z^2}{2\tau^2}}\, dx$$

$$= \frac{1}{2\pi \sigma \cdot \tau} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2\tau^2}(x - z\sigma^2)^2} \cdot e^{\frac{z^2\sigma^2}{2\tau^2}} \cdot e^{-\frac{z^2}{2\tau^2}} \, dx$$

$$= \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma\tau} \int_{-\infty}^{\infty} e^{-\frac{(x-z\sigma^2)^2}{2\sigma^2\cdot\tau^2}} \, dx}_{=1} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}\underbrace{\left[\frac{1}{\tau^2} - \frac{\sigma^2}{\tau^2}\right]}_{=1}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

<u>Conclusion</u> : $Z \sim N(0,1)$.

We know : if $X \sim N(\mu, \sigma)$ then

$aX + b \sim N(a\mu + b, a^2\sigma^2)$.

In general : $X \sim N(\mu, \sigma^2)$, $Y \sim N(\nu, \tau^2)$

$$X + Y = \sqrt{\sigma^2 + \tau^2} \times \left(\underbrace{\frac{X-\mu}{\sqrt{\sigma^2+\tau^2}}}_{X'} + \underbrace{\frac{Y-\nu}{\sqrt{\sigma^2+\tau^2}}}_{Y'} \right) + \mu + \nu$$

We have $X' \sim N(0, \frac{\sigma^2}{\sigma^2+\tau^2})$ in $Y' \sim N(0, \frac{\tau^2}{\sigma^2+\tau^2})$. The expression $X'+Y' \sim N(0,1)$.

It follows that

$$\boxed{X+Y \sim N(\mu+\nu, \sigma^2+\tau^2)}$$

<u>Example</u>: Let X, Y be independent standar normal.

Let $Z = \frac{Y}{X}$. Density of Z ?

Define

$$\Phi(x,y) = (x, y/x)$$

$$\Phi^{-1}(x,z) = (x, xz) \Rightarrow$$

$$J\Phi^{-1}(x,z) = \det\begin{pmatrix} 1 & 0 \\ z & x \end{pmatrix} = x$$

The density of (X, Z) is

$$f_{X,Z}(x,z) = f_X(x)\, f_Y(xz) \cdot |x|$$

$$= \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}}\, e^{-\frac{(xz)^2}{2}} \cdot |x|$$

We get the density of Z as the marginal density

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{2\pi} \cdot e^{-\frac{x^2(1+z^2)}{2}} \cdot |x|\, dx$$

$$= \frac{1}{\pi} \int_0^{\infty} e^{-\frac{x^2(1+z^2)}{2}}\, x \cdot dx$$

$$= \frac{1}{\pi(1+z^2)} \left(- e^{-\frac{x^2(1+z^2)}{2}} \right) \Big|_0^{\infty}$$

$$= \frac{1}{\pi(1+z^2)}$$

Example: Let X, Y be independent with $X \sim \Gamma(a, \lambda)$ and $Y \sim \Gamma(b, \lambda)$. Let $Z = X + Y$. We established that $Z \sim \Gamma(a+b, \lambda)$ but will ~~do~~ it again using convolution.

$$f_Z(z) = \int_0^z f_X(x) \, f_Y(z-x) \, dx$$

$$= \int_0^z \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x} \cdot \frac{\lambda^b}{\Gamma(b)} (z-x)^{b-1} e^{-\lambda(z-x)} \, dx$$

$$= \frac{\lambda^{a+b}}{\Gamma(a)\,\Gamma(b)} \cdot e^{-\lambda z} \cdot \int_0^z x^{a-1} (z-x)^{b-1} \, dx$$

New variable: $x = z \cdot u$

$dx = z \cdot du$

$$= \frac{\lambda^{a+b}}{\Gamma(a)\,\Gamma(b)} \cdot e^{-z} \cdot \int_0^1 u^{a-1} \cdot (1-u)^{b-1} \cdot z^{a+b-1} \, du$$

$$= \frac{\lambda^{a+b}}{\Gamma(a)\,\Gamma(b)}\, e^{-z} \cdot z^{a+b-1} \quad B(a,b)$$

The result is a density which means that it integrates to 1.

But we know that

$$\frac{\lambda^{a+b}}{\Gamma(a+b)} \int_0^\infty z^{a+b-1} \cdot e^{-\lambda z}\, dz = 1.$$

This means that

$$\frac{\Gamma(a+b)}{\Gamma(a)\,\Gamma(b)} \cdot B(a,b) = 1$$

which is Euler identity!

We have used probability to derive Euler's identity.

Theorem 3.6 has a more general version.

<u>Theorem 3.6 a</u>: Let $\underline{X}$ be a random vector with density $f_{\underline{X}}(\underline{x})$. Assume $P(\underline{X} \in \mathcal{U}) = 1$ for some open set $\mathcal{U} \subseteq \mathbb{R}^n$ and let $\Phi : \mathcal{U} \to \mathcal{S} \subseteq \mathbb{R}^n$ be a bijective map between $\mathcal{U}$ and $\mathcal{S}$ such that Φ and Φ^{-1} are continuously partially differentiable. Let $\underline{Y} = \Phi(\underline{X})$. Then $\underline{Y}$ has the density

$$\boxed{f_{\underline{Y}}(y) = f_{\underline{X}}(\Phi^{-1}(y)) \cdot |J_{\Phi^{-1}}(y)|.}$$

<u>Proof</u>: Same as before.

<u>Example</u>: Let $\underline{X} = (X_1, X_2, \dots, X_r)$ such that $X_1, X_2, \dots, X_r$ are independent and $X_k \sim N(0,1)$ for all $k = 1, 2, \dots, r$; Let $\underline{A}$ be an invertible matrix. Define

$$\Phi(\underline{x}) = \underline{A}\,\underline{x} + \underline{\mu} \quad \text{for } \underline{\mu} \in \mathbb{R}^n.$$

We have $\Phi^{-1}(y) = \underline{A}^{-1}(\underline{x} - \underline{\mu})$

and $J\Phi^{-1}(y) = \det(\underline{A}^{-1}) = \dfrac{1}{\det(A)}$

The transformation formula gives for $\underline{Y} = \underline{A}\,\underline{X} + \underline{\mu}$

$$f_{\underline{Y}}(y) = f_{\underline{X}}(\Phi^{-1}(y)) \cdot |J\Phi^{-1}(y)|.$$

We have

$$f_{\underline{X}}(\underline{x}) = \prod_{k=1}^{r} f_{X_k}(x_k)$$

$$= \frac{1}{(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}\sum_{k=1}^{n} x_k^2}$$

$$= \frac{1}{(2\pi)^{n/2}} \cdot e^{-\frac{1}{2}\underline{x}^T \cdot \underline{x}}$$

It follows

$$f_{\underline{Y}}(y) = \frac{1}{(2\pi)^{n/2} |\det(\underline{A})|}$$

$$\times \; e^{-1/2 [\underline{A}^{-1}(\underline{x}-\underline{\mu})]^T [\underline{A}^{-1}(\underline{x}-\underline{\mu})]}$$

$$= \frac{1}{(2\pi)^{n/2} |\det(\underline{A})|}$$

$$\times \; e^{-1/2 (\underline{x}-\underline{\mu})^T (\underline{A}^{-1})^T \underline{A}^{-1} (\underline{x}-\underline{\mu})}$$

Denote $\underline{\Sigma} = \underline{A} \cdot \underline{A}^T$. We have

$$\underline{\Sigma}^{-1} = (\underline{A}^T)^{-1} \cdot A^{-1} = (\underline{A}^{-1})^T (\underline{A}^{-1})$$

We have

$$\frac{1}{|\det(\underline{A})|} = \frac{1}{\sqrt{\det(\underline{\Sigma})}}$$

and

$$f_{\underline{Y}}(y) = \frac{1}{(2\pi)^{r/2}\sqrt{\det(\underline{\Sigma})}}$$

$$\times \quad e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \underline{\Sigma}^{-1}(\underline{x}-\underline{\mu})}.$$

<u>Comment</u>: The above density is called the multivariate normal density with parameters $\underline{\mu} \in \mathbb{R}^r$ and $\underline{\Sigma}$ $(r \times r)$.

Example: Let $\underline{X}$ have density

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \underline{\Sigma}}}$$

$$\times \; e^{-\frac{1}{2}(\underline{x}-\underline{\mu})^T \underline{\Sigma}^{-1}(\underline{x}-\underline{\mu})}$$

If $\underline{X} = (X_1, X_2, \dots, X_n)$ what is the distribution of $\underline{X}^{(1)} = (X_1, X_2, \dots, X_p)$, $p < n$? Denote $\underline{X} = (\underline{X}^{(1)}, \underline{X}^{(2)})$,

$$\underline{\mu} = \begin{pmatrix} \underline{\mu}^{(1)} \\ \underline{\mu}^{(2)} \end{pmatrix} \quad \text{and} \quad \underline{\Sigma} = \begin{pmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{pmatrix}.$$

$\underline{\mu}^{(1)}$ is a p-dimensional vector, $\underline{\Sigma}_{11}$ $(p \times p)$, $\underline{\Sigma}_{21}$ $(p \times q)$, $\underline{\Sigma}_{21}$ $(q \times p)$, $\underline{\Sigma}_{22}$ $(q \times q)$. Define $\underline{\Phi} : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\underline{\Phi}(\underline{x}) = \begin{pmatrix} \underline{x}^{(1)} \\ \underline{x}^{(2)} - \underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1}\underline{x}^{(1)} \end{pmatrix}.$$

$\underline{\Phi}(\underline{x})$ is a linear map

$$\underline{\Phi}(\underline{x}) = \underbrace{\begin{pmatrix} \underline{I}_r & 0 \\ -\underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1} & I_\ell \end{pmatrix}}_{\underline{A}} \underline{x} .$$

Since the matrix is lower triangular we have

$$D\underline{\Phi} = A \Rightarrow J_{\underline{\Phi}}(\underline{x}) = 1$$

$$\Rightarrow J_{\underline{\Phi}^{-1}}(\underline{y}) = 1 .$$

We have

$$\underline{\Phi}^{-1}(\underline{y}) = \begin{pmatrix} y^{(1)} \\ \underline{y}^{(2)} + \underline{\Sigma}_{21}\Sigma_{11}^{-1}\underline{y}^{(1)} \end{pmatrix}$$

It follows that

$$f_{\underline{Y}}(\underline{y}) = f_{\underline{X}}(\underline{\Phi}^{-1}(\underline{y})) \cdot 1$$

We need some linear algebra.

Suppose $\underline{A}$, $\underline{B}$ are invertible matrices. Write

$$\underline{A} = \begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} \quad \text{and} \quad \underline{B} = \begin{pmatrix} \underline{B}_{11} & \underline{B}_{12} \\ \underline{B}_{21} & \underline{B}_{22} \end{pmatrix}$$

where $\underline{A}_{ij}$ and $\underline{B}_{ij}$ are of the same dimension. If $\underline{A} \cdot \underline{B} = \underline{I}$ we have

$$\underline{A}_{11} \underline{B}_{11} + \underline{A}_{12} \underline{B}_{21} = \underline{I}$$
$$\underline{A}_{11} \underline{B}_{12} + \underline{A}_{12} \cdot \underline{B}_{22} = 0$$

For simplicity we assume $\underline{\mu} = 0$.

We need to compute

$$[\underline{\Phi}^{-1}(y)]^T \cdot \underline{\Sigma}^{-1} [\underline{\Phi}^{-1}(y)].$$

In matrix form this means

$$\underline{y}^T \begin{pmatrix} \underline{I}_p & \underline{\Sigma}_{11}^{-1} \underline{\Sigma}_{21} \\ \underline{0} & \underline{I}_q \end{pmatrix} \underline{\Sigma}^{-1} \begin{pmatrix} \underline{I}_p & \underline{0} \\ \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} & \underline{I}_q \end{pmatrix}$$

Denote $\underline{A} = \underline{\Sigma}^{-1}$. From

$$\begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} \begin{pmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{pmatrix} = \underline{I}_n$$

we have

$$\underline{A}_{11} \underline{\Sigma}_{11} + \underline{A}_{21} \underline{\Sigma}_{21} = \underline{I}_p$$
$$\underline{A}_{11} \underline{\Sigma}_{12} + \underline{A}_{12} \underline{\Sigma}_{22} = \underline{0}$$
$$\underline{A}_{21} \underline{\Sigma}_{11} + \underline{A}_{22} \underline{\Sigma}_{21} = \underline{0}$$

We compute

$$\begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} \begin{pmatrix} \underline{I}_p & \underline{0} \\ \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} & \underline{I}_q \end{pmatrix}$$

$$= \begin{pmatrix} \underline{A}_{11} + \underline{A}_{12} \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1}, & \underline{A}_{12} \\ \underline{A}_{21} + \underline{A}_{22} \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1}, & \underline{A}_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \underline{\Sigma}_{11}^{-1} & \underline{A}_{12} \\ \underline{0} & \underline{A}_{22} \end{pmatrix}$$

Continue to get

$$\begin{pmatrix} \underline{I}_p & \underline{\Sigma}_{11}^{-1}\underline{\Sigma}_{12} \\ 0 & \underline{I}_q \end{pmatrix} \begin{pmatrix} \underline{\Sigma}_{11}^{-1} & \underline{A}_{12} \\ \underline{0} & \underline{A}_{22} \end{pmatrix}$$

$$= \begin{pmatrix} \underline{\Sigma}_{11}^{-1} , & \underline{A}_{12} + \underline{\Sigma}_{11}^{-1}\underline{\Sigma}_{12}\underline{A}_{22} \\ -''- , & \underline{A}_{22} \end{pmatrix}$$

But

$$\begin{pmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} = \underline{I}_n$$

gives

$$\underline{\Sigma}_{11}\underline{A}_{12} + \underline{\Sigma}_{12}\underline{\Sigma}_{22} = 0 \quad , \text{ so}$$

$$\underline{\Sigma}_{11}\left(\underline{A}_{12} + \underline{\Sigma}_{11}^{-1}\underline{\Sigma}_{12}\underline{A}_{22}\right) = 0$$

The linear equations give

$$\underline{A}_{22} = \left(\underline{\Sigma}_{22} - \underline{\Sigma}_{12}\underline{\Sigma}_{11}^{-1}\underline{\Sigma}_{21}\right)$$

(see Appendix)

So we have

$$[\underline{\Phi}^{-1}(y)]^T \underline{\Sigma}^{-1} [\underline{\Phi}^{-1}(y)]$$

$$= y^T \begin{pmatrix} \underline{\Sigma}_{11}^{-1} & \underline{0} \\ \underline{0} & (\underline{\Sigma}_{22} - \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} \underline{\Sigma}_{12})^{-1} \end{pmatrix} y$$

$$= [y^{(1)}]^T \underline{\Sigma}_{11}^{-1} y^{(1)}$$

$$+ [y^{(2)}]^T (\underline{\Sigma}_{22} - \underline{\Sigma}_{21} \underline{\Sigma}_{11} \underline{\Sigma}_{12})^{-1} y^{(2)}$$

Comment: in general replace y by $y - \underline{\mu}$. So we have

$$f_{\underline{Y}}(y) = f(\underline{y}^{(1)}) \cdot g(y^{(2)}).$$

This means that

$$\underline{Y}^{(1)} = (X_1, \dots, X_r)$$

$$\underline{Y}^{(2)} = \underline{X}^{(2)} - \underline{\Sigma}_{21} \cdot \underline{\Sigma}_{11}^{-1} X^{(1)}$$

are independent vectors.

Appendix: if we have

$$\begin{pmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{12} \\ \underline{\Sigma}_{21} & \underline{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \underline{A}_{11} & \underline{A}_{12} \\ \underline{A}_{21} & \underline{A}_{22} \end{pmatrix} = \underline{I}_n$$

then

$$\underline{\Sigma}_{11} \underline{A}_{11} + \underline{\Sigma}_{12} \underline{A}_{22} = \underline{I}_p$$

$$\underline{\Sigma}_{11} \underline{A}_{12} + \underline{\Sigma}_{12} \underline{A}_{22} = \underline{0}$$

$$\underline{\Sigma}_{21} \underline{A}_{11} + \underline{\Sigma}_{22} \underline{A}_{21} = \underline{0}$$

$$\underline{\Sigma}_{21} \underline{A}_{12} + \underline{\Sigma}_{22} \underline{A}_{22} = \underline{I}_q$$

we have a system of 4 linear equations with 4 unknowns.

Multiply the second equation with $\underline{\Sigma}_{11}^{-1}$ from the left to get

$$\underline{A}_{12} + \underline{\Sigma}_{11}^{-1} \underline{\Sigma}_{12} \underline{A}_{22} = \underline{0}$$

Insert this into the last equation to get

$$- \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} \underline{\Sigma}_{12} \underline{A}_{22} + \underline{\Sigma}_{22} \underline{A}_{22} = \underline{I}_{\ell}$$

we have

$$\underline{A}_{22} = (\underline{\Sigma}_{22} - \underline{\Sigma}_{12} \underline{\Sigma}_{11}^{-1} \underline{\Sigma}_{21})^{-1}$$

This result is known as the inversion lemma.

<u>Remark</u>: Invertibility follows from the fact that the product is $\underline{I}_{\ell}$.

3.5. Conditional distributions

In elementary probability we have that $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$.

If X is a discrete random variable with values $\{x_1, x_2, \ldots\}$ the distribution is given by the probabilities $P(X = x_k)$.

If we have additional information in the sense that the event B has happened our opinion about the probabilities of event $\{X = x_k\}$ change to the conditional probabilities

$$P(\{X = x_k\} \mid B) = \frac{P(\{X = x_k\} \cap B)}{P(B)}.$$

We can verify easily that

$$\sum_{x_k} P(X = x_k \mid B) = 1.$$

This observation motivates the definition of conditional probabilities and distributions.

<u>Definition</u>: Let X be a discrete random variable with values in $\{x_1, x_2, \dots\}$. The <u>conditional</u> <u>distribution</u> of X given B with $P(B) > 0$ is given by

$$P(X = x_k \mid B) = \frac{P(\{X = x_k\} \cap B)}{P(B)}.$$

<u>Comment</u>: In most cases the event B is of the form $B = \{Y = y_\ell\}$ for some random variable Y.

Example: Let X, Y be independent with $X \sim \text{Bin}(m, p)$ and $Y \sim \text{Bin}(n, p)$. Let $Z = X + Y$. We know that $Z \sim \text{Bin}(m+n, p)$.

The conditional distribution of X given $\{Z = r\}$ is given by

$$P(X = k \mid Z = r) = \frac{P(X = k, Z = r)}{P(Z = r)}$$

$$= \frac{P(X = k, Y = r-k)}{P(Z = r)}$$

$$= \frac{P(X = k)\, P(Y = r-k)}{P(Z = r)} \qquad \underline{\text{indep}}$$

$$= \frac{\binom{m}{k} p^k q^{m-k} \cdot \binom{n}{r-k} p^{r-k} \cdot q^{n-r+k}}{\binom{m+n}{r} p^r q^{m+n-r}}$$

$$= \frac{\binom{m}{k}\binom{n}{r-k}}{\binom{m+n}{r}}$$

for $k \leq \min(m, r)$ and

$k \geq \max(0, r-n)$.

We recognize the hypergeometric distribution. We write

$$X \mid_{Z=r} \sim \text{Hipergeom}(r, m, m+n).$$

<u>Definition</u>: Let $\underline{X}$ be a discrete random vector with values $\{\underline{x}_1, \underline{x}_2, \ldots\}$. Let B be an event. The conditional distribution of $\underline{X}$ given B with $P(B) > 0$ is given by conditional probabilities

$$P(\underline{X} = \underline{x}_k \mid B) = \frac{P(\{X = x_k\} \cap B)}{P(B)}.$$

As before in most cases B is of the form $B = \{\underline{Y} = y\}$ for some random vector $\underline{Y}$.

<u>Example</u>: Let $\underline{X} = (X_1, \dots, X_r)$ be multinomial with parameters n and $\underline{p} = (p_1, p_2, \dots, p_r)$. Let $s < r$. What is the conditional distribution of $(X_1, X_2, \dots, X_s)$ given $Y = X_1 + X_2 + \dots + X_s = m$.

Denote $Z = X_{s+1} + \dots + X_r$. We know that $Y \sim \text{Bin}(n, p_1 + \dots + p_s)$. We compute for $k_1 + \dots + k_s = m$

$$P(X_1 = k_1, \dots, X_s = k_s \mid Y = m)$$

$$= \frac{P(X_1 = k_1, \dots, X_s = k_s, Z = n-m)}{P(Z = n-m)}$$

$$= \frac{n!}{k_1! \cdots k_s! \, (n-m)!}$$

$$\times \; p_1^{k_1} \cdots p_s^{k_s} \, (1 - p_1 - \cdots - p_s)^{n-m} \Big/$$

$$\Big/ \binom{n}{m} (p_1 + \cdots + p_s)^m \, (1 - p_1 - \cdots - p_s)^{n-m}$$

$$= \frac{m!}{k_1! \cdots k_s!}$$

$$\times \; \frac{p_1^{k_1} \cdots p_s^{k_s}}{(p_1 + \cdots + p_s)^m}$$

$$= \quad (*)$$

We denote: $\hat{p}_k = \dfrac{p_k}{(p_1 + \cdots + p_s)}$

for $k = 1, 2, \ldots, s$. We have:

$$* = \frac{m!}{k_1! \cdots k_s!} \, \hat{p}_1^{k_1} \cdots \hat{p}_s^{k_s}.$$

Conclusion: $(X_1, X_2, \ldots, X_s)$

has the multinomial distribution.

with parameters $n-m$ and $\hat{p} = (\hat{p}_1, \dots, \hat{p}_s)$. We write

$$\underline{X}' = (X_1 \dots, X_s) \quad \text{and}$$

$$\underline{X}' \,|_{X_1 + \dots + X_s = m} \sim \text{Multinom}(m, \hat{p}).$$

For the continuous case the intuitive idea is that we will define conditional densities. If (X, Y) has density $f_{X,Y}(x,y)$ then the conditional density of Y given $\{X = x\}$ should be proportional to the function

$$y \longmapsto f_{X,Y}(x,y)$$

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

Example : Let

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \times$$

$$\times e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}}$$

for $|\rho| < 1$. We know that

$X \sim N(0,1)$ i.e.

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

We write

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-x^2/2} \times$$

$$\times e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}$$

It follows that

$$f_{Y|X=x}(y)$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} e^{-\frac{(y-\rho x)^2}{2(1-\rho^2)}}$$

We notice

$$Y \mid X=x \sim N(\rho x, 1-\rho^2)$$

The definition has a vector version.

<u>Definition</u>: Let $(\underline{X}, \underline{Y})$ have density $f_{\underline{X},\underline{Y}}(\underline{x}, \underline{y})$. Assume $f_{\underline{X}}(\underline{x}) > 0$. The conditional density of $\underline{Y}$ given $\{\underline{X} = \underline{x}\}$ is given by

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

Example: Let $\underline{X} = (\underline{X}^{(1)}, \underline{X}^{(2)}) \sim N(\underline{\mu}, \underline{\Sigma})$.

What is $f_{\underline{X}^{(2)} \mid \underline{X}^{(1)} = \underline{x}^{(1)}}(\underline{x}^{(2)})$?

Direct calculation is difficult but we found out that $\underline{X}^{(1)}$ and $\underline{X}^{(2)} - \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} \underline{X}^{(1)}$ are independent vectors. If we write $\underline{Y} = \underline{X}^{(2)} - \underline{\Sigma}_{21} \underline{\Sigma}_{11} \underline{X}^{(1)}$ we know that

$$f_{\underline{Y}}(\underline{y}) = \frac{1}{(2\pi)^{q/2} \sqrt{\det\left(\underline{\Sigma}_{22} - \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} \underline{\Sigma}_{12}\right)}}$$

$$\times\ e^{-\frac{1}{2}\left(\underline{y} - \underline{\mu}^{(2)} + \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} \underline{\mu}^{(1)}\right)^T \left(\underline{\Sigma}_{22} - \underline{\Sigma}_{21} \underline{\Sigma}_{11}^{-1} \underline{\Sigma}_{12}\right)^{-1} \left(- \text{''} - \right)}.$$

But

$$\begin{pmatrix} \underline{X}^{(1)} \\ \underline{X}^{(2)} \end{pmatrix} = \begin{pmatrix} \underline{I}_p & 0 \\ \underline{\Sigma}_{21} \underline{\Sigma}_{11} & \underline{I}_q \end{pmatrix} \begin{pmatrix} \underline{X}^{(1)} \\ \underline{Y} \end{pmatrix}$$

The Jacobian of this is 1

so we can write

$$f_{\underline{X}}(\underline{x}) = f_{\underline{X}^{(1)}}(\underline{x}^{(1)})$$

$$\times \; f_{\underline{Y}}\left(\underline{x}^{(2)} - \underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1}\underline{x}^{(1)}\right)$$

Now it is easy to divide by $f_{\underline{X}^{(1)}}(\underline{x}^{(1)})$. We get

$$f_{\underline{X}^{(2)} \mid \underline{X}^{(1)} = \underline{x}^{(1)}}(\underline{x}^{(2)})$$

$$= f_{\underline{Y}}\left(\underline{x}^{(2)} - \underline{\Sigma}_{12}\underline{\Sigma}_{11}^{-1}\underline{x}^{(1)}\right).$$

Using the form of $f_{\underline{Y}}$ we find:

$$\underline{X}^{(2)} \mid \underline{X}^{(1)} = \underline{x}^{(1)} \sim N\left(\underline{\mu}^{(2)} + \underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1}(\underline{x}^{(1)} - \underline{\mu}^{(1)}), \; \underline{\Sigma}_{22} - \underline{\Sigma}_{21}\underline{\Sigma}_{11}^{-1}\underline{\Sigma}_{12}\right)$$