MATH 3338 Probability

Wenjiang Fu
wfu2@central.uh.edu

Lecture 4 - Chapter 4. Conditional Probability

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Outline

Conditional Probability

- In this chapter, we study how the outcome of one event affect another event.
  Examples: We know the general mortality rate is not high, but if the mortality rate is much higher in the hospital ER room or the late stage cancer treatment group, or a senior home.
  In these examples, the group of special subjects with specific conditions, like ER room, senior home, etc potentially make a big difference in the occurrence of the event we are interested in studying.
- Conditional probability of event F given event E, assuming the probability $P(E) > 0$, is

$$P(F|E) = \frac{P(F \cap E)}{P(E)}$$

Conditional Probability

- Example 4.2 (Life Table). In a population of females, 89.835% can expect to live to age 60, while 57.062% can expect to live to age 80. Given that a woman is 60, what is the probability that she lives to age 80?

  Solution Let E denote the event of living up to age 60, and F the event of living up to age 80. Since the woman is age 60, she has lived up to age 60, ie. E occurs. So the probability of her living up to age 80 is

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{P(F)}{P(E)} = \frac{0.57062}{0.89835} = 0.6352$$

Conditional Probability

- Example 4.5 Two urns, I and II. Urn I has 2 black balls and 3 white balls. Urn II has 1 black ball and 1 white ball. An urn is drawn at random and a ball is chosen at random from it.
  We can ask once an urn is chosen, what is the probability of drawing a black ball?
  This can be done using conditional probability following the path of choosing an urn first, and then draw a ball from it.
  Or we can also ask if a black ball is drawn, what is the probability that this black ball is from Urn 1?
  Solution Following the conditional prob

$$P(I|B) = \frac{P(I \cap B)}{P(B)} = \frac{P(I \cap B)}{P(B \cap I) + P(B \cap II)}$$

$$= \frac{P(B|I)P(I)}{P(B|I)P(I) + P(B|II)P(II)} = \frac{(2/5)(1/2)}{(2/5)(1/2) + (1/2)(1/2)} = \frac{4}{9}$$

Bayes Probability

- **Tree diagram** Urn $\rightarrow$ *ball*:

$$
\begin{array}{cccccc}
 & \nearrow & Urn1 & \begin{array}{c}\rightarrow \\ \searrow\end{array} & \begin{array}{c}Black \\ White\end{array} \\
Urn & & & & \\
 & \searrow & Urn2 & \begin{array}{c}\rightarrow \\ \searrow\end{array} & \begin{array}{c}Black \\ White\end{array}
\end{array}
$$

- **Reverse tree diagram** Ball $\rightarrow$ *Urn*:

$$
\begin{array}{cccccc}
 & \nearrow & Black & \begin{array}{c}\rightarrow \\ \searrow\end{array} & \begin{array}{c}Urn1 \\ Urn2\end{array} \\
Ball & & & & \\
 & \searrow & White & \begin{array}{c}\rightarrow \\ \searrow\end{array} & \begin{array}{c}Urn1 \\ Urn2\end{array}
\end{array}
$$

Bayes Probability

- **Two conditional Probabilities:**
  $P(Ball\ color|specific\ Urn)$ is a natural conditional probability.
  Its inverse prob is $P(specific\ Urn|ball\ color)$ is called Bayes prob.
  We now know how to calculate the Bayes probability.
- **Monty Door game**
  3 doors, one has a car behind it, and the others are empty.
  Contestants are required to pick one first but not to open it. Then the Monty will open one door with no car. Contests are then asked if they want to switch the door before they open the door they choose.
  How many want to switch the door? how many want to stay with their original choice.

- **Independent Events** Two events may affect the each other in the way that the outcome of one event may affect the outcome of the other.
- **Definition 4.1** Let E and F be two events. We say that they are independent if either 1) both events have positive probability and

$$P(E|F) = P(E) \text{ and } P(F|E) = P(F),$$

or 2) at least one of the events has probability 0.
Note: often, we do not consider the case where one event has 0 prob to occur.

Independent Events

- **Theorem 4.1** Two events E and F are independent iff (if and only if)

$$P(E \cap F) = P(E)P(F)$$

Proof We need to prove both ways.
1) Assume independence of E and F. Then by the multiplication law and the definition of independence,

$$P(E \cap F) = P(E|F)P(F) = P(E)P(F)$$

2) Assume the above equation holds true,

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{P(E)P(F)}{P(F)} = P(E)$$

Similarly, $P(F|E) = P(F)$.

Independent Events

- **Example of Independent Events** (Example 4.8)
  Some independent events seem to be obvious, while others may not.
  In tossing a **fair** coin twice, we define two events.
  A is "The first toss is a head".
  B is "The two outcomes are the same".

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(\{HH\})}{P(\{HH, HT\})} = \frac{1/4}{1/2} = 1/2 = P(B)$$

  Note that the textbook does not emphasize fair coin. If $P(H) \neq .5$, are the two events A and B still independent?

- Example of dependent events.
  $A =${HH}, B is the "first is head". Then A and B are not independent.
  Why?

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

- **Definition 4.2** (Mutually independent)
  A set of events $\{A_1, A_2, ..., A_n\}$ is said to be mutually independent if for any subsets $\{A_i, A_j, ..., A_m\}$ of these events,
  $$P(A_i \cap A_j \cap ... \cap A_m) = P(A_i)P(A_j)...P(A_m),$$
  or equivalently, their complement events $\overline{A_i}$, ..., satisfy the above equation.

  Q: If all pairs of events are independent, are they mutually independent?
  A: not necessary.

- It is also important to know that the statement
  $$P(A_1 \cap A_2 \cap ... \cap A_n) = P(A_1)P(A_2)...P(A_n),$$
  does not imply that the events $A_1, A_2, ..., A_n$ are mutually independent.

Joint Distribution

- **Joint Distribution Functions**
  We consider multiple random variables $X_1, X_2, ..., X_n$, corresponding to a given experiment. We may consider the joint random variable $X = (X_1, X_2, .., X_n)$, called $n-$tuple, the corresponding n outcomes of RV $X_1, X_2, ..., X_n$.
  Note, $X_1, X_2, ..., X_n$ may or may not be the same kind of experiments.

- **Definition 4.3** Let $X_1, X_2, ..., X_n$ be RVs. Suppose the sample space of X_i is set R_i. Then the joint RV $X = (X_1, ..., X_n)$ is defined to be the RV whose outcomes consist of ordered n-tuples of outcomes, with the ith coordinate lying in the set R_i. The sample space Ω of X is the Cartesian product of R_i's

$$\Omega = R_1 \times R_2 \times ... \times R_n.$$

  The joint distribution function of X is the function which gives the probability of each of the outcomes of X.

- **Joint Distribution**

  Example 4.13 A group of 60 have smokers, non-smokers, cancer patients and no cancer subjects, summarized in a table.

	Non-smoker	Smoker	Total
No cancer	40	10	50
Cancer	7	3	10
Total	47	13	60

  Joint distribution of Cancer (Ca) and Smoking (Smk): (Ca, Smk).

Ca $\backslash Smk$	0	1	Total
0	40/60	10/60	50/60
1	7/60	3/60	10/60
Total	47/60	13/60	1

- We study the joint distribution of two random variables because their outcomes may affect each other.

Joint Distribution

- **Definition 4.4** The random variables $X_1, X_2, ..., X_n$ are mutually independent if

$$P(X_1 = r_1, X_2 = r_2, ..., X_n = r_n)$$

$$= P(X_1 = r_1)P(X_2 = r_2)...P(X_n = r_n)$$

for any choice of $r_1, r_2, ..., r_n$. Thus if $X_1, X_2, ..., X_n$ are mutually independent, then the joint distribution function of the random variable $X = (X_1, , X_2, ..., X_n)$ is just the product of individual distribution functions.
For two random variables mutually independent, we just say they are independent.
Note If two or more random variables are not independent, their intersection probability (joint prob, not joint events "$\cup$") is not equal to the product of their individual probs.

Joint Distribution

- **Marginal distributions** The probabilities in the subtotal column or subtotal row are the marginal probability. Marginal probs of Ca are $P(Ca = 0) = 50/60 = 5/6$, $P(Ca = 1) = 10/60 = 1/6$.
  and marginal probs of Smk are $P(Smk = 0) = 47/60$, $P(Smk = 1) = 13/60$.

- How to examine if two random variables are independent?
  In the above example, how do we examine if two RVs Ca and Smk are indep?
  Following the definition of indep. RVs.

  $$P(Ca = i, Smk = j) = P(Ca = i)P(Smk = j), \text{ for } any\ i, j = 0, 1$$

  Note $P(Ca = 0, Smk = 0) = 40/60 = 2/3$, and $P(Ca = 0) = 5/6$, $P(Smk = 0) = 47/60$. Hence we have $P(Ca = 0, Smk = 0) \neq P(Ca = 0)P(Smk = 0)$. This implies that these two RVs Ca and Smk are not independent.

Joint Distribution

- **Marginal distributions** The probabilities in the subtotal column or subtotal row are the marginal probability. Marginal probs of Ca are $P(Ca = 0) = 50/60 = 5/6$, $P(Ca = 1) = 10/60 = 1/6$.
  and marginal probs of Smk are $P(Smk = 0) = 47/60$, $P(Smk = 1) = 13/60$.

- How to examine if two random variables are independent?
  In the above example, how do we examine if two RVs Ca and Smk are indep?
  Following the definition of indep. RVs.

$$P(Ca = i, Smk = j) = P(Ca = i)P(Smk = j), \text{ for } any\ i, j = 0, 1$$

  Note $P(Ca = 0, Smk = 0) = 40/60 = 2/3$, and $P(Ca = 0) = 5/6, P(Smk = 0) = 47/60$. Hence we have $P(Ca = 0, Smk = 0) \neq P(Ca = 0)P(Smk = 0)$. This implies that these two RVs Ca and Smk are not independent.

- **Independence of RVs** In fact, the dependence of the two RVs can be checked with the conditional prob.
  $P(Ca = 0|Smk = 0) = \frac{P(Ca=0, Smk=0)}{P(Smk=0)} = \frac{40/60}{47/60} = 40/47$,
  $P(Ca = 0) = 50/60 = 5/6$.
  So $P(Ca = 0|Smk = 0) \neq P(Ca = 0)$, thus Ca and Smk are not indep.

- How do we examine the independence of RVs using the table?

Ca $\backslash Smk$	0	1	Total
0	40/60	10/60	50/60
1	7/60	3/60	10/60
Total	47/60	13/60	1

  Compare conditional prob (eg. Cond. on Smk=0) with Marginal Prob. with no condition.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Independent Trials Processes

- **Definition 4.5** A sequence of RVs $X_1, ..., X_n$ mutually indep. and follow the same distribution. We call this a sequence of indep. trials or an indep trials process.
  Indep trials processes arise naturally. A single experiment with sample space $R = \{r_1, r_2, ..., r_s\}$.

$$m_X = \begin{pmatrix} r_1 & r_2 & \dots & r_s \\ p_1 & p_2 & \dots & p_s \end{pmatrix}.$$

  An example is an ordered s-tuple (10-tuple) of fair coin tossing, where s is a fixed number.
  We can repeat this experiments n times. The sample space is $\Omega = R \times R \times ... \times R$, consisting of all possible sequences $\omega = (\omega_1, ..., \omega_n)$, each ω_j chosen from R. Assign a distribution function $m(\omega) = m(\omega_1)m(\omega_2)...m(\omega_n)$ with $m(\omega_2) = p_k, \omega_j = r_k$. Let X_j denote the j-th coordinate of the outcome $(r_1, r_2, ..., r_n)$. The RVs $X_1, X_2, ..., X_n$ form an indep trials process.

Independent Trials Processes

- **Example 4.1.4** A sequence of rolling a die three times. X_i is the outcome of the ith roll, $i = 1, 2, 3$. The common distr function is

$$m_i = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

The sample space is $\Omega = R \times R \times R$ with $R = \{1, 2, 3, 4, 5, 6\}$. If $\omega = (1, 3, 6)$, then $X_1(\omega) = 1, X_2(\omega) = 3, X_3(\omega) = 6$, indicating the first roll is 1, the second is 3 and the third is 6. The prob assigned to any sample point (in the 3 rolls of the die) is

$$m(\omega) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Independent Trials Processes

- **Example 4.15** Bernoulli trials process with prob. p for success. Let $X_j(\omega) = 1$ for a success, and $X_j(\omega) = 0$ for failure. Then $X_1, X_2, ..., X_n$ is an indep trials process. Each X_j has the same distribution function

$$m_j = \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix},$$

If $S_n = X_1 + ... + X_n$, then

$$P(S_n = j) = \binom{n}{j} p^j (1-p)^{n-j},$$

and the random sum S_n has a binomial distribution $Bin(n, p, j)$.

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Bayes' Formula

- **Bayes Probabilities**
  Suppose we have a set of events $H_1, H_2, ..., H_m$ that are pairwise disjoint and exhaustive (exclusive and exhaustive):

$$\Omega = H_1 \cup H_2 \cup ... \cup H_m.$$

  We call these events *hypotheses*. We also have one event E with information. We can event E *evidence*.
  Before receiving evidence E, we have a set of prior probabilities $P(H_1), ..., P(H_m)$ for the hypotheses. If we know the correct hypothesis, we know the probability for the evidence, i.e. $P(E|H_i)$ for all i. Now we want to know $P(H_i|E)$, i.e. the conditional prob of hypotheses given the evidence. This is called the posterior probabilities.

- Very important and useful idea in data science with many applications.

Bayes' Formula

- **Bayes Probabilities**
  For conditional prob

$$P(H_i|E) = \frac{P(H_i \cap E)}{P(E)}.$$

The numerator $P(H_i \cap E) = P(H_i)P(E|H_i)$ by the multiplication rule. Furthermore,

$$P(E) = \sum_{j=1}^{m} P(E \cap H_j) = \sum_{j=1}^{m} P(E|H_j)P(H_j)$$

We have the Bayes' formula

$$P(H_i|E) = \frac{P(H_i)P(E|H_i)}{\sum_{j=1}^{m} P(H_j)P(E|H_j)}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Bayes' Formula

- **Bayes Probabilities**

  Example 4.16. Medical diagnosis. A med doctor to decide if a patient has one of 3 diseases, d_1, d_2, d_3. Two tests to carry out with $+$ or $-$ for each test: $++, +-, -+, --$. National records have the distribution of the diseases and test results in the table.

Disease	# (Pat w Dis)	$++$	$+-$	$-+$	$--$
d_1	3215	2110	301	704	100
d_2	2125	396	132	1187	410
d_3	4660	510	3568	73	509
Total	10000				

test	d_1	d_2	d_3	Posterior prob $P(d_i \mid test)$
$++$	.700	.131	.169	
$+-$	.075	.033	.892	
$-+$	.358	.604	.038	
$--$	.098	.403	.499	

Continuous Conditional Probability

- **Conditional Density Function**
  As in continuous distribution, we need to specify density functions for continuous conditional probability.

- **Continuous Conditional Density function** If Continuous RV X has density $f(x)$, and event E with $P(E) > 0$. We define a conditional density function

$$f(x|E) = \begin{cases} f(x)/P(E), & \text{if } x \in E, \\ 0, & \text{if } x \notin E. \end{cases}$$

Then for any event F, we have

$$P(F|E) = \int_F f(x|E)dx.$$

$P(F|E)$ is called the conditional probability of F given E.

$$P(F|E) = \int_F f(x|E)dx = \int_{E\cap F} \frac{f(x)}{P(E)}dx = \frac{P(E\cap F)}{P(E)}$$

UNIVERSITY of HOUSTON
DEPARTMENT OF MATHEMATICS

Continuous Conditional Probability

- **Conditional Density Function**
  The conditional density function can be regarded as a density function of RV X, which is 0 outside of E, and normalized to the total prob 1 when inside E.

- If the original density is uniform, then the same will be true for the conditional density, but just normalized in E.

  Example. $X \sim Unif(0, 1)$. $E = [0, .5]$ Then the density of $X|E$ is uniform on $[0, .5]$. Just need to normalize the density function so that the total prob is equal to 1. So the constant density of $f = 2$ over $[0, .5]$.

- **Conditional Density Function**
  Example 4.20 Exponential density. Emission experiment of plutonium-239. X is the length of time that passes until the next emission. X follows an exponential distribution with parameter $\lambda > 0$, depending on the size of the lump of plutonium

$$f(x) = \lambda e^{-\lambda x}, \forall x \geq 0$$

  Suppose experiment starts at r seconds. Calculate the prob that there is no emission in a further s seconds.
  Let $G(t)$ be prob next particle is emitted after time t. Then

$$G(t) = \int_t^\infty \lambda e^{-\lambda x} dx = -e^{-\lambda x}|_t^\infty = e^{-\lambda t}.$$

Continuous Conditional Probability

- **Conditional Density Function**
  Example 4.20 Exponential density.
  Let E be the event "the next particle is emitted after time r" and F the event "the next particle is emitted after time $r+s$". Then

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{G(r+s)}{G(r)} = e^{-\lambda(r+s)}/e^{-\lambda r} = e^{-\lambda s}.$$

  The right-hand-side (RHS) is independent of r, and depends only on s. What this implies is that the conditional prob $P(F|E)$ does not depend on the conditional time r in the condition. Which implies that it only depends on the extra time s.

- This special feature of exponential distribution is called "memoryless". Is this a good feature?
  It depends. For radiation, emission, this may be an advantage to make things easier to estimate. But think about human life, survival after a number of years.

- **Independent Events**
  Two events are independent with positive prob, if
  $P(E \cap F) = P(E)P(F)$.
  Example 4.21 Exponential density.
  Let E be the event the dart lands on the upper half of the circle, and F the event that the dart lands on the right half of the circle.

$$P(F \cap E) = \frac{1}{\pi} \int_{E \cap F} 1 dxdy = area(E \cap F)/\pi = P(E)P(F)$$

  Implies indep of E and F.

- **Joint Density and Cumulative Distribution Functions**
  Definition 4.6 Let $X_1, ..., X_n$ be continuous rvs associated with an experiment, and let $X = (X_1, ..., X_n)$. Then the joint cumulative distribution function of X is defined by

$$F(x_1, x_2, .., x_n) = P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n).$$

  The joint density function of X satisfies the following equation

$$F(x_1, ..., x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(t_1, t_2, ..., t_n) dt_n dt_{n-1} ... dt_1.$$

  It can be shown that

$$f(x_1, x_2, .., x_n) = \frac{\partial^n F(x_1, ..., x_n)}{\partial x_1 \partial x_2 ... \partial x_n}.$$

Continuous Conditional Probability

- **Independent Random Variables**
  Definition 4.7 Let $X_1, ..., X_n$ be continuous rvs with CDFs $F_1(x_1), ..., F_n(x_n)$. Then these RVs are mutually indep if

$$F(x_1, x_2, .., x_n) = F_1(x_1)F_2(x_2)...F_n(x_n)$$

  for any choice of $x_1, ..., x_n$.

- If RVs are mutually indep, then the CDF of $X = (X_1, ..., X_n)$ is the product of individual CDFs of $X_1, ..., X_n$.

- **Theorem 4.2** Let $X_1, ..., X_n$ be continuous rvs with density functions $f_1(x_1), .., f_n(x_n)$. Then the RVs are mutually indep iff

$$f(x_1, ..., x_n) = f_1(x_1)...f_n(x_n)$$

  for all choice of $x_1, ., x_n$.

- **Joint distributions**

  Example 4.22 Let (ω_1, ω_2) is randomly selected from the unit square. Let $X_1 = \omega_1^2, X_2 = \omega_2^2$, and $X_3 = \omega_1 + \omega_2$. Find the joint distribution $F_{12}(r_1, r_2)$ and, are X_2 and X_3 indep?
  Q: Are ω_1, ω_2 indep of each other in this case?
  How about if (ω_1, ω_2) are selected randomly from the unit disk? (i.e. uniformly distributed on the disk).
  Since (ω_1, ω_2) is uniformly distributed on the unit square $[0,1] \times [0,1]$. Then (ω_1 and ω_2) must be indep. Why?
  Then we need to prove that X_1 and X_2 are indep.
  How to prove the indep of X_1, X_2?
  Can we do this?

$$F_{X_1X_2}(r_1, r_2) = F_{\omega_1^2\omega_2^2}(r_1, r_2) = ...$$

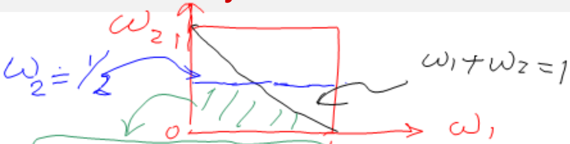
- **Joint distributions**
  Example 4.22 We prove that X_1 and X_2 are indep. $\forall r_1, r_2 \geq 0$,

$$F_{X_1 X_2}(r_1, r_2) = P(X_1 \leq r_1, X_2 \leq r_2) = P(\omega_1^2 \leq r_1, \omega_2^2 \leq r_2)$$

$$= P(\omega_1 \leq \sqrt{r_1}, \omega_2 \leq \sqrt{r_2}) = P(\omega_1 \leq \sqrt{r_1})P(\omega_2 \leq \sqrt{r_2})$$

$$= F_{\omega_1}(\sqrt{r_1})F_{\omega_2}(\sqrt{r_2}) = \sqrt{r_1}\sqrt{r_2} = \sqrt{r_1 r_2}$$

  X_2 and X_3 indep? How to show?
  Use special cases. Take $F_{23}(X_2 \leq 1/4, X_3 \leq 1)$ and see if it is equal to $F_2(1/4)F_3(1)$.
  It is straightforward to see that
  $F_2(1/4) = P(\omega_2^2 \leq 1/4) = P(\omega_2 \leq 1/2) = 1/2$ and from example 2(2.14) on the last page of lecture 2. $F_3(r_3) = r_3^2/2$ for $0 \leq r_3 \leq 1$, and $F_3(r_3) = 2r_3 - r_3^2/2 - 1$ for $1 \leq r_3 \leq 2$, $F_3(1) = 1/2$



- **Joint distributions**
  Example 4.22
  Use special cases. Take $F_{23}(X_2 \leq 1/4, X_3 \leq 1)$ and see if it is equal to $F_2(1/4)F_3(1)$. Now need to figure out $F_{23}(1/4, 1) = P(X_2 \leq 1/4, X_3 \leq 1) = P(\omega_2 \leq 1/2, \omega_1 + \omega_2 \leq 1)$. Notice that (ω_1, ω_2) are uniformly distributed on unit square $[0, 1]^2$. The prob $P(\omega_2 \leq 1/2, \omega_1 + \omega_2 \leq 1)$ is the area of the region defined by the two inequalities $\omega_2 \leq 1/2, \omega_1 + \omega_2 \leq 1$. Hence the area is $1/2 + 1/8 = 3/8$. Consequently,

$$F_2(1/4)F_3(1) = (1/2)(1/2) = 1/4 \neq 3/8 = F_{23}(1/4, 1)$$

  Apparently, it can be seen that the two random variables ω_2^2 and $\omega_1 + \omega_2$ are not indep even though, ω_1 and ω_2 are.

Continuous Conditional Probability

- **Joint distributions**
  Theorem 4.3 Let $X_1, ..., X_n$ be mutually indep continuous rvs and let $\phi_1(x), ..., \phi_n(x)$ be continuous functions. Then $\phi_1(X_1), ..., \phi_n(X_n)$ are mutually indep.

- **Joint distributions**
  Definition 4.8 A sequence of RVs $X_1, ..., X_n$ that are mutually indep and have the same density is called an indep trials process.
  Example. The indep trials processes arise naturally when an experiment by a single random variable is repeated n times.