Random Variables

Fall 2015

Instructor:

Ajit Rajwade

Topic Overview

- Random variable: definition
- Discrete and continuous random variables
- Probability density function (pdf) and cumulative distribution function (cdf)
- Joint and conditional pdfs
- Expectation and its properties
- Variance and covariance
- Markov’s and Chebyshev’s inequality
- Weak law of large numbers

Random variable

- In many random experiments, we are not always interested in the observed values, but in some numerical quantity determined by the observed values.
- Example: we may be interested in the sum of the values of two dice throws, or the number of heads appearing in n consecutive coin tosses.
- Any such quantities determined by the results of random experiments are called as **random variables** (they may also be the observations themselves).

Random variable

Value of *X* (Denoted as *x*) where *X* = sum of 2 dice throws	**P(*X*=*x*)**
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

This is called the **probability mass function** (pmf) table of the random variable X. If S is the sample space, then P(S) = P(union of all events of the form $X = x$) = 1 (verify from table).

Random variable: Notation

- A random variable is usually denoted by an *upper case* alphabet.

- Individual values the random variable can acquire are denoted by *lower case*.

Random variable: discrete

- Random variables whose values can be written as a finite or infinite sequence are called **discrete random variables**.

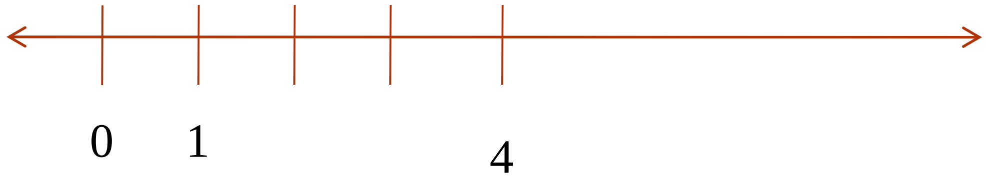
- Example: results of coin toss or random dice experiments

- The probability that a random variable X takes on value x, i.e. $P(X=x)$, is called as the **probability mass function**.

Random variable: continuous

- Random variables that can take on values within a continuum are called **continuous random variables**.

- Example: the dimensions (length, height, width, weight) of an object are usually continuous quantities, direction of a vector, amount of water that can be stored in a 4 litre jar is a continuous random variable in the interval [0,4].



Random variable: continuous

- For a continuous random variable, the probability that it takes on any *particular* value within a continuum is **zero**!

- Why? Because there are infinitely many values – say in the interval [0,4] in the example on the previous slide. Each value will be equally likely.

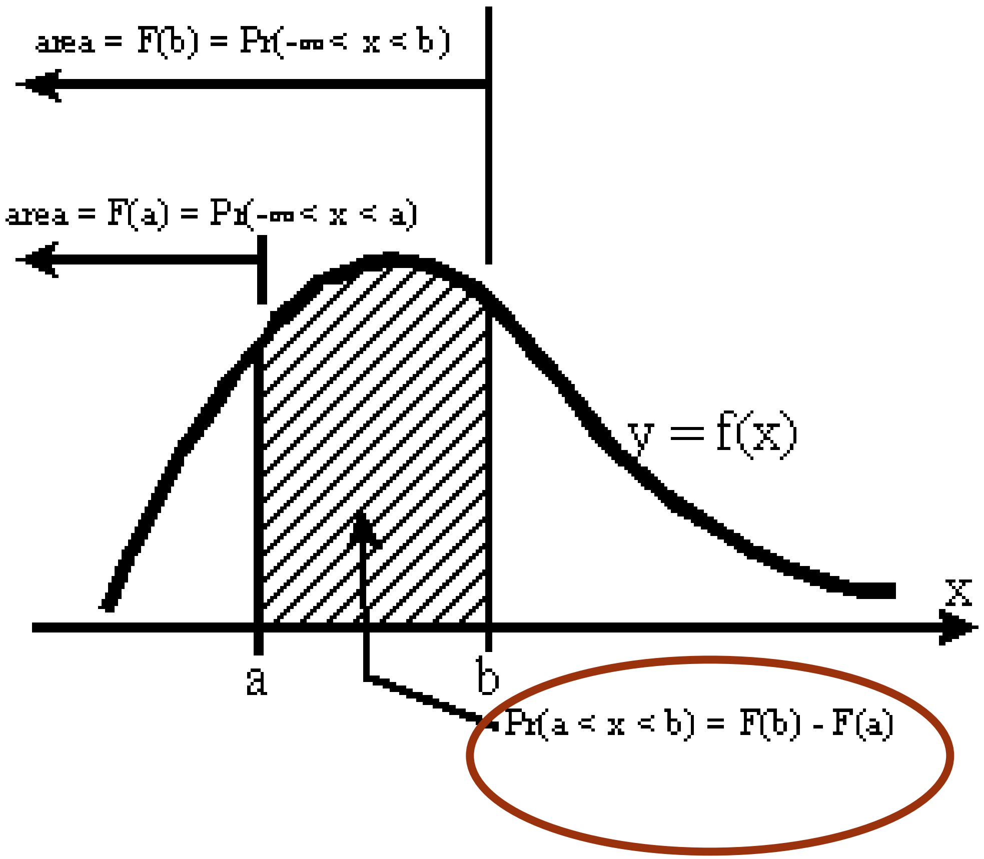
- **Note**: Zero probability in case of continuous random variables does *not* mean the event will *never* occur! This differs from the discrete case.

Random variable: continuous

- Hence for a continuous random variable *X*, we consider the **cumulative distribution function** (cdf) $F_X(x)$ defined as $P\{X \le x\}$.

- The cdf is basically the probability that *X* takes on a value less than or equal to *x*.

- The cdf can be used to compute **cumulative interval measures**, that is the probability that *X* takes on a value greater than *a* and less than or equal to *b*, i.e. $P(a < X \le b) = F_X(b) - F_X(a)$.



Cumulative interval measure

Random variable: continuous - example

- Consider a cdf of the form:

$F_X(x) = 0$ for $x \leq 0$, and

$F_X(x) = 1\text{-}\exp(-x^2)$ otherwise

- To find: probability that X exceeds 1

- $P(X > 1) = 1 - P(X \leq 1) = 1 - F_X(1) = e^{-1}$

Probability Density Function (pdf)

- The pdf of a random variable X at a value x is the *derivative* of its cumulative distribution function (cdf) at that value x.

- It is a non-negative function $f_X(x)$ such that for any set B of real numbers, we have $P\{X \in B\} = \int_B f_X(x)dx$

- Properties:

$$\int_{-\infty}^{+\infty} f_X(x)dx = 1$$

$$P(a \le X \le b) = \int_a^b f_X(x)dx = F_X(b) - F_X(a)$$

$$P(X = a) = \int_a^a f_X(x)dx = 0$$

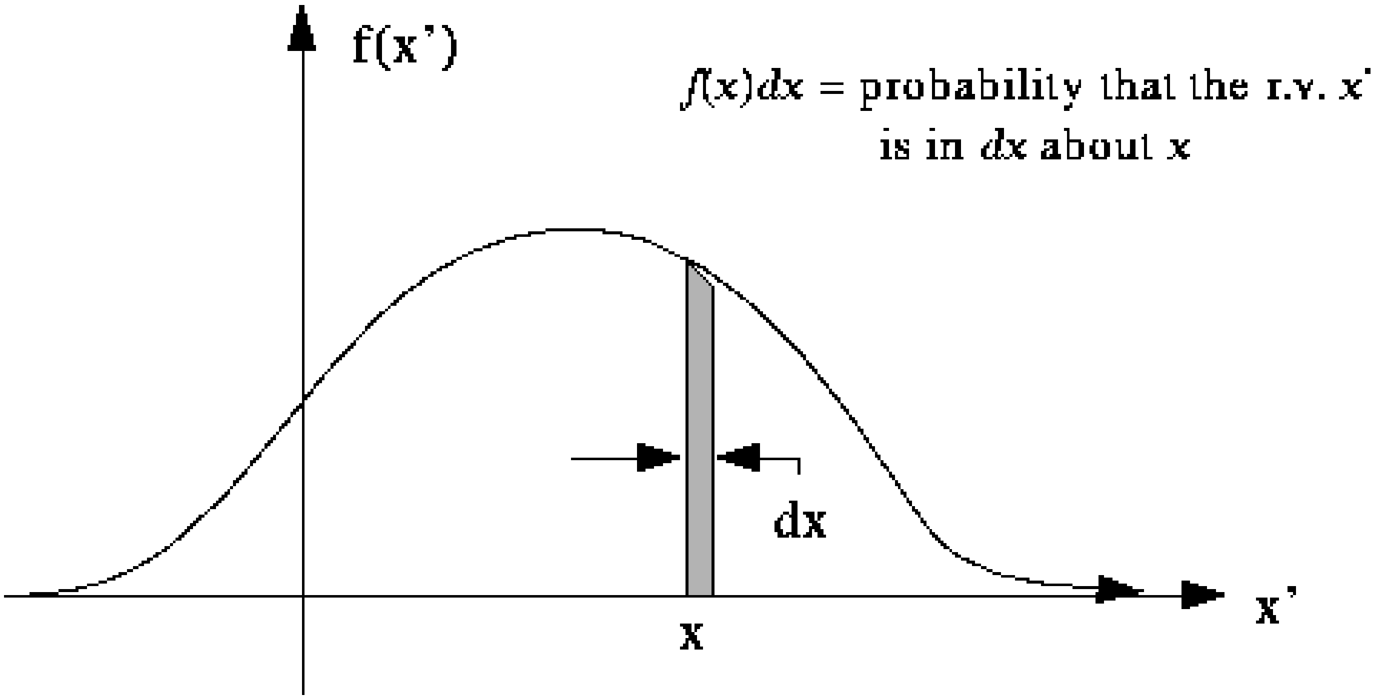


Figure 4. Typical Probability Distribution Function (*pdf*)

Probability Density Function

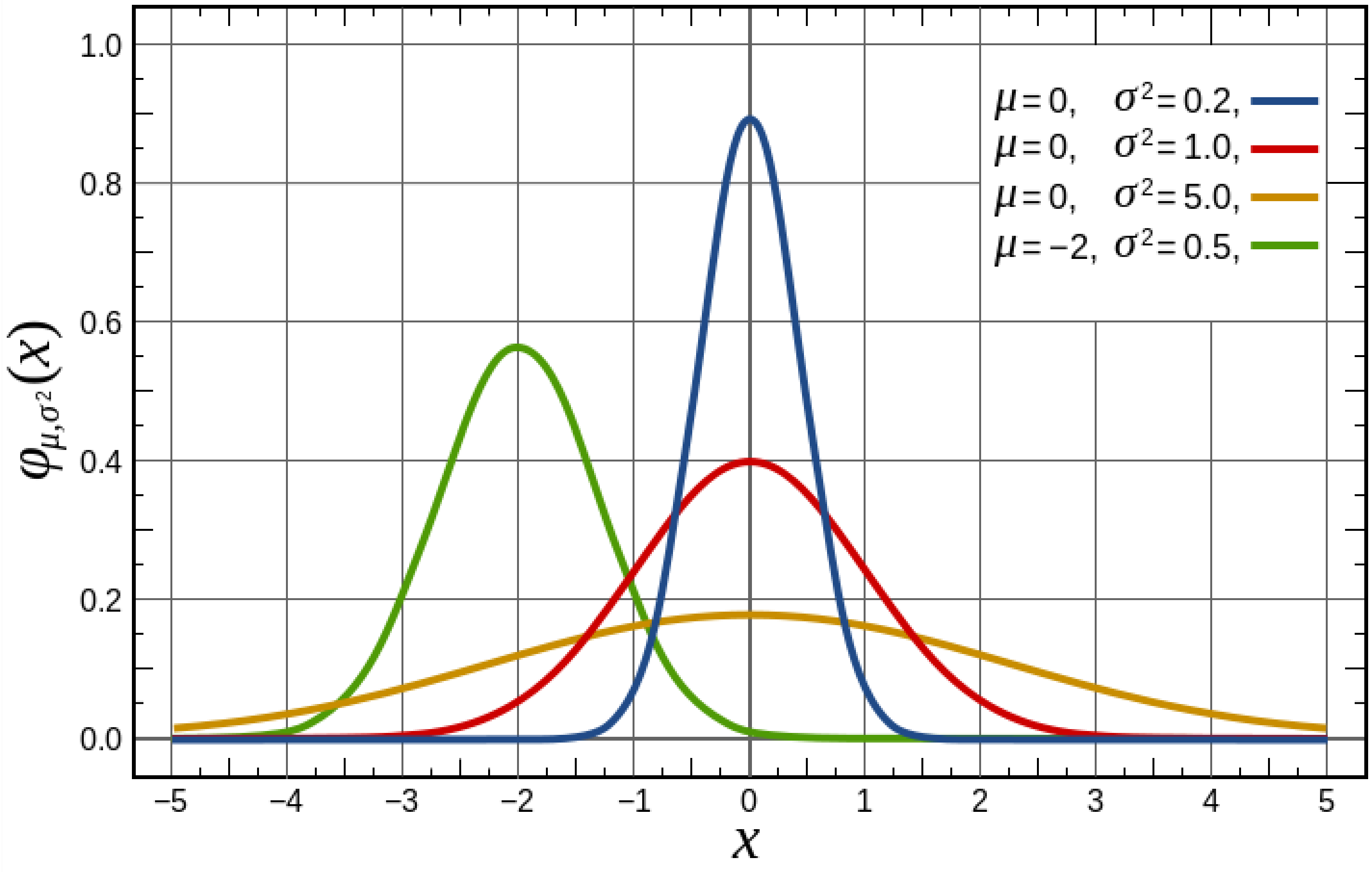
- Another way of looking at this concept:

$$P\{a - \varepsilon/2 \le X \le a + \varepsilon/2\} = \int_{a-\varepsilon/2}^{a+\varepsilon/2} f_X(x)dx \approx \varepsilon f(a)$$

$$f_X(a) = \lim_{\varepsilon \to 0} \frac{P\{a - \varepsilon/2 \le X \le a + \varepsilon/2\}}{\varepsilon}$$

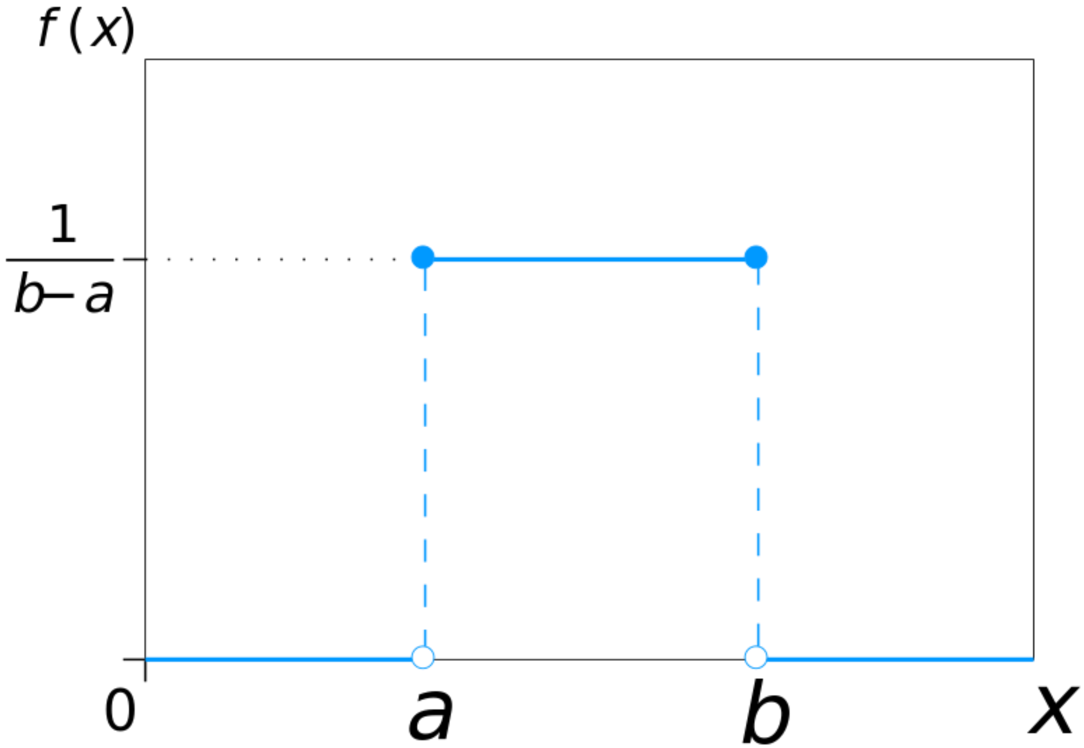
Examples: Popular families of PDFs

- Gaussian (normal) pdf: $f_X(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$



Examples: Popular families of PDFs

- Bounded uniform pdf: $$f_X(x) = \frac{1}{(b-a)}, a \le x \le b$$
$$= 0 \text{ otherwise}$$



Expected Value (Expectation) of a random variable

- It is also called the **mean value** of the random variable.

- For a discrete random variable *X*, it is defined as:

$$E(X) = \sum_i x_i P(X = x_i)$$

- For a continuous random variable *X*, it is defined as:

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx$$

- The expected value should not be (mis)interpreted to be the value that *X* *usually* takes on – it's the average value, *not* the "most frequently occurring value".

Expected Value (Expectation) of a random variable

- For some pdfs, the expected value is not always defined, i.e. the integral below may not have a finite value.

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)\,dx$$

Expected Value: examples

- The expected value that shows up when you throw a die is 1/6(1+2+3+4+5+6) = 3.5.

- The game of roulette consists of a ball and wheel with 38 numbered pockets on its side. The ball rolls and settles on one of the pockets. If the number in the pocket is the same as the one you guessed, you win $35 (probability 1/38), otherwise you lose $1 (probability 37/38). The expected value of the amount you earn after one trial is: (-1)37/38 +(35)1/38 = $-0.0526

A Game of Roulette

Expected value of a function of random variable

- Consider a function $g(X)$ of a discrete random variable X. The expected value of $g(X)$ is defined as:

$$E(g(X)) = \sum_i g(x_i) P(X = x_i)$$

- For a continuous random variable, the expected value of $g(X)$ is defined as:

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f_X(x) dx$$

Properties of expected value

$$E(ag(X)+b)=\int_{-\infty}^{+\infty}(ag(x)+b)f_X(x)dx$$

$$=\int_{-\infty}^{+\infty}ag(x)f_X(x)dx+\int_{-\infty}^{+\infty}bf_X(x)dx$$

$$=aE(g(X))+b---why?$$

This property is called the **linearity** of the expected value. In general, a function $f(x)$ is said to be linear in x is $f(\mathrm{a}x+\mathrm{b}) = \mathrm{a}f(x)+\mathrm{b}$ where a and b are constants. In this case, the expected value is not a function but an operator (it takes a function as input). An operator E is said to be linear if $\mathrm{E}(\mathrm{a}f(x) + \mathrm{b}) = \mathrm{a}\,\mathrm{E}(f(x)) + \mathrm{b}$.

Properties of expected value

Suppose you want to predict the value of a random variable with a known mean. On an average, what value will yield the least squared error?

Let X be the random variable and c be its predicted value.

We want to find c such that $\mathrm{E}((X - c)^2)$ is minimized.

Let μ be the mean of X.

Then

$$E((X\text{-}c)^2) = E((X - \mu + \mu - c)^2)$$

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

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

$$= E((X - \mu)^2) + (\mu - c)^2 + 0$$

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

The expected value is the value that yields the least mean squared prediction error!

The median

- What minimizes the following quantity?

$$J(c)=\int_{-\infty}^{+\infty}|x-c|\,f_X(x)dx$$

$$F(c)=\int_{-\infty}^{c}|x-c|\,f_X(x)dx+\int_{c}^{\infty}|x-c|\,f_X(x)dx$$

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

$$=\int_{-\infty}^{c}cf_X(x)dx-\int_{-\infty}^{c}xf_X(x)dx+\int_{c}^{\infty}xf_X(x)dx-\int_{c}^{\infty}cf_X(x)dx$$

$$=cF_X(c)-\int_{-\infty}^{c}xf_X(x)dx+\int_{c}^{\infty}xf_X(x)dx-c(1-F_X(c))$$

The median

$$J(c) = cF_X(c) - \int_{-\infty}^{c} xf_X(x)dx + \int_{c}^{\infty} xf_X(x)dx - c(1 - F_X(c))$$

$$J(c) = cF_X(c) - \int_{-\infty}^{c} q(x)dx + \int_{c}^{\infty} q(x)dx - c(1 - F_X(c))$$

$$q(x) = xf_X(x)$$

In this derivation, we are assuming that the two definite integrals of $q(x)$ exist! This proof won't go through otherwise.

$$J(c) = cF_X(c) - (Q(c) - Q(-\infty)) + (Q(\infty) - Q(c)) - c(1 - F_X(c))$$

$$(Q(x) = \int xf_X(x)dx)$$

$$= 2cF_X(c) - c - 2Q(c) + Q(\infty) + Q(-\infty)$$

The median

$$J(c) = 2cF_X(c) - c - 2Q(c) + Q(\infty) + Q(-\infty)$$

$$J'(c) = 0$$
$$\therefore 2cf_X(c) + 2F_X(c) - 1 - 2q(c) = 0$$
$$\therefore 2cf_X(c) + 2F_X(c) - 1 - 2cf_X(c) = 0$$
$$\therefore 2F_X(c) - 1 = 0$$
$$\therefore F_X(c) = 1/2$$

This is the median – by definition and it minimizes J(c). We can double check that J''(c) >= 0. Notice the peculiar definition of the median for the continuous case here! This definition is not conceptually different from the discrete case, though. Also, note that the median will not be unique if F_X is not differentiable at *c*.

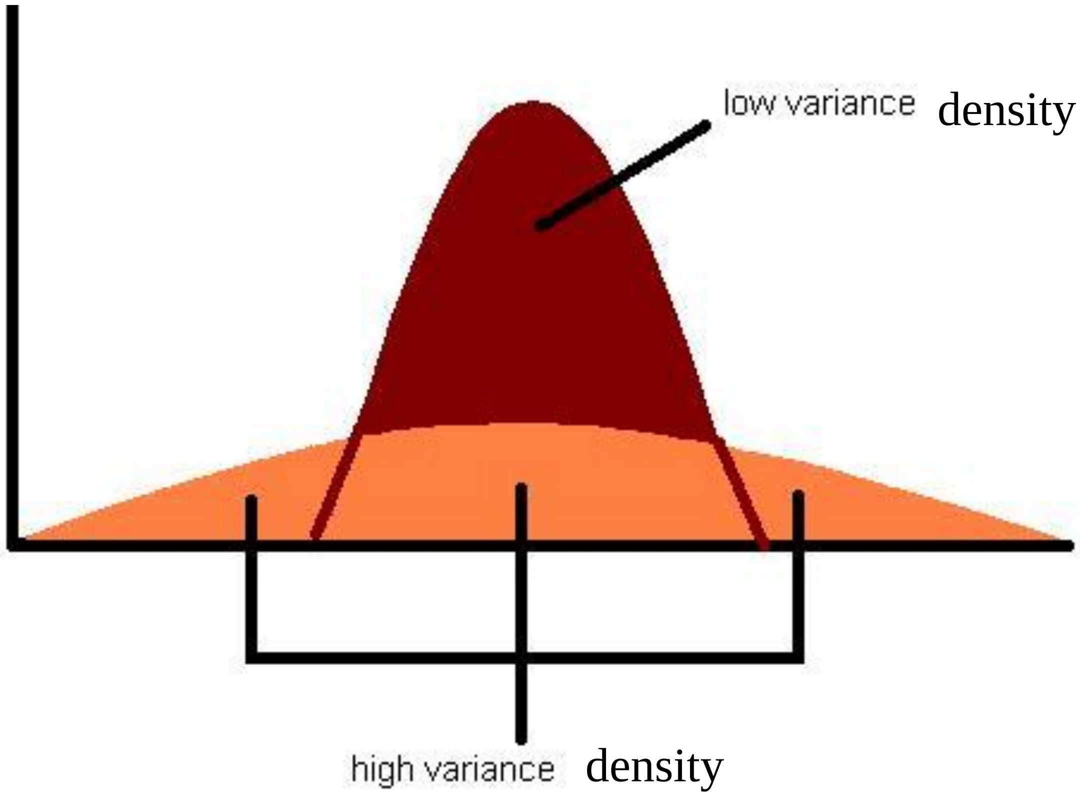
Variance

- The **variance** of a random variable X tells you how much its values deviate from the mean – on an average.
- The definition of variance is:

$$Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx$$

- The positive square-root of the variance is called the **standard deviation**.
- For some distributions, the variance (and hence standard deviation) may not be defined, because the integral may not have a finite value.

Variance: some intuition



Low-variance probability mass functions or probability densities tend to be concentrated around one point. High variance densities are spread out.

Variance: Alternative expression

- The definition of variance is:

$$Var(X) = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx$$

- Alternative expression:

$$Var(X) = E[(X-\mu)^2] = E[X^2 + \mu^2 - 2X\mu]$$
$$= E[X^2] + \mu^2 - 2E[X]\mu$$
$$= E[X^2] + \mu^2 - 2\mu^2 - - - why?$$
$$= E[X^2] - \mu^2$$
$$= E[X^2] - (E[X])^2$$

Variance: properties

- Property:

$$
\begin{aligned}
Var(aX+b) &= E[(aX+b-E(aX+b))^2] \\
&= E[(aX+b-(a\mu+b))^2] \\
&= E[a^2(X-\mu)^2] \\
&= a^2E[(X-\mu)^2] = a^2Var(X)
\end{aligned}
$$

Probabilistic inequalities

- Sometimes we know the mean or variance of a random variable, and want to guess the probability that the random variable can take on a certain value.

- The exact probability can usually not be computed as the information is too less. But we can get upper or lower bounds on this probability which can influence our decision-making processes.

Probabilistic inequalities

- Example: Let's say the average annual salary offered to a CSE Btech-4 student at IITB is \$100,000. What's the probability that you (i.e. a randomly chosen student) will get an offer of \$110,000? **Additionally, if you were told that the variance of the salary was 50,000, what's the probability that your package is between \$90,000 and \$110,000?**

Markov's inequality

- Let X be a random variable that takes only *non-negative* values. For any $a > 0$, we have

  $$P\{X \geq a\} \leq E[X]/a$$

- Proof: next slide

Markov's inequality

- Proof:

$$E[X] = \int_0^{\infty} x f_X(x) dx$$

$$= \int_0^{a} x f_X(x) dx + \int_a^{\infty} x f_X(x) dx$$

$$\geq \int_a^{\infty} x f_X(x) dx$$

$$\geq \int_a^{\infty} a f_X(x) dx$$

$$= a \int_a^{\infty} f_X(x) dx$$

$$= aP\{X \geq a\}$$

Chebyshev's inequality

- For a random variable X with mean μ and variance σ^2, we have for any value $k > 0$,

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

- Proof: follows from Markov's inequality

$(X-\mu)^2$ is a non-negative random variable

$$\therefore P\{(X-\mu)^2 \geq k^2\} \leq E[(X-\mu)^2]/k^2 = \sigma^2/k^2$$

$$\therefore P\{|X-\mu| \geq k\} \leq \sigma^2/k^2$$

Chebyshev's inequality: another form

- For a random variable X with mean μ and variance σ^2, we have for any value $k > 0$,

$$P\{|X-\mu| \geq k\} \leq \frac{\sigma^2}{k^2}$$

- If I replace k by $k\sigma$, I get the following:

$$P\{|X-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$$

Back to counting money! ☺

- Let X be the random variable indicating the annual salary offered to you when you reach Btech-4 ☺

- Then

$$P\{X \geq 110K\} \leq \frac{100K}{110K} = 0.9090 \approx 90\%$$

$$P\{|X - 100K| \geq 10K\} \leq \frac{50K}{10K \times 10K} = 0.0005 \approx 0.05\%$$

$$\therefore P\{|X - 100K| < 10K\} = 1 - 0.05\% = 99.5\%$$

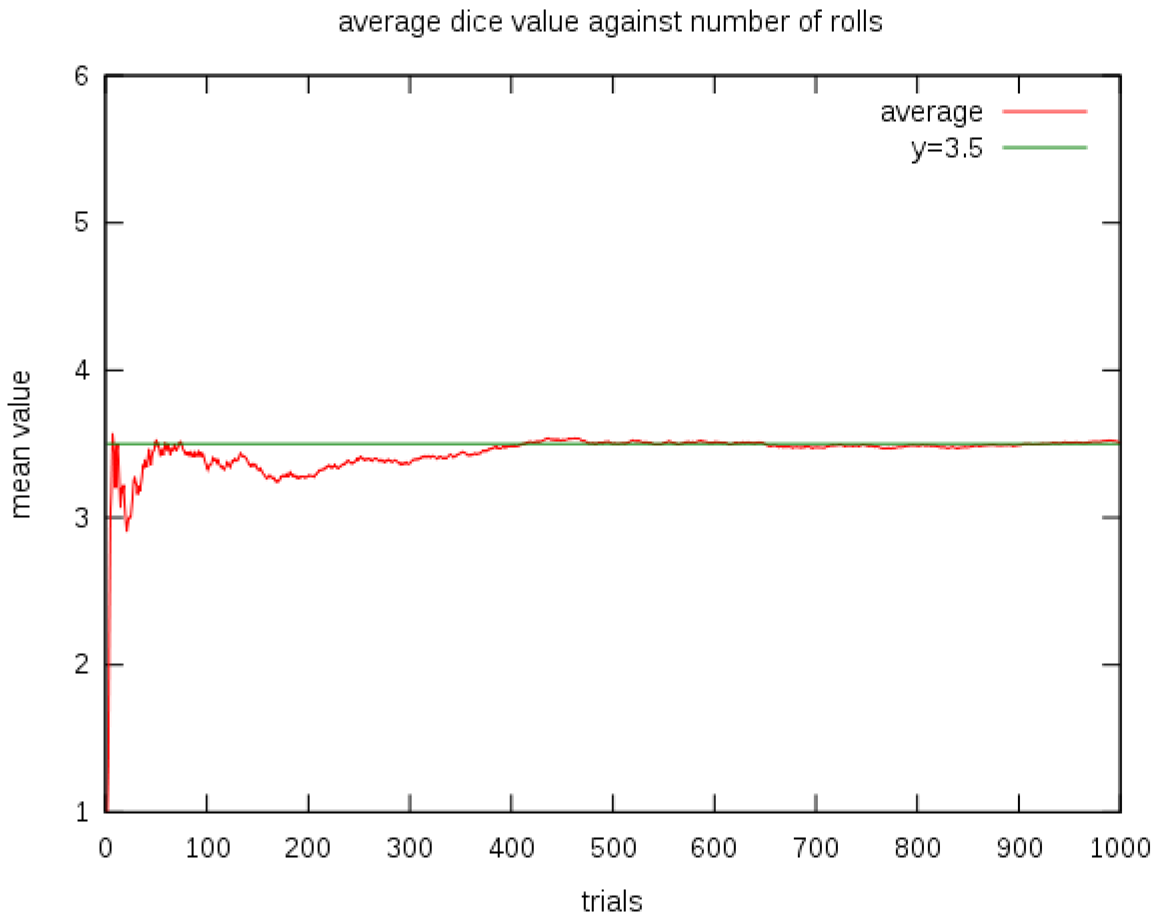
Back to the expected value

- When I tell you that the expected value of a random die variable is 3.5, what does this mean?

- If I throw the die n times, and average the results, I should get a value close to 3.5 provided n is very large (not valid if n is small).

- As n increases, the average value should move closer and closer towards 3.5.

- That's our basic intuition!



https://en.wikipedia.org/wiki/Law_of_large_numbers

Back to the expected value: weak law of large numbers

- This intuition has a rigorous theoretical justification in a theorem known as the **weak law of large numbers**.

- Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables each having mean μ. Then for any $\varepsilon > 0$, we have:

$$P\{|\frac{X_1 + X_2 + \ldots + X_n}{n} - \mu| > \varepsilon\} \to 0 \text{ as } n \to \infty$$

Back to the expected value: weak law of large numbers

- Let $X_1, X_2,\ldots,X_n$ be a sequence of independent and identically distributed random variables each having mean μ. Then for any ε > 0, we have:

$$P\{|\frac{X_1+X_2+\ldots+X_n}{n}-\mu|>\varepsilon\}\to 0 \text{ as } n\to\infty$$

Empirical (or sample) mean

- Proof: follows immediately from Chebyshev's inequality

$$E(\frac{X_1+X_2+\ldots+X_n}{n})=\mu, Var(\frac{X_1+X_2+\ldots+X_n}{n})=\frac{n\sigma^2}{n^2}=\frac{\sigma^2}{n},$$

$$\therefore P\{|\frac{X_1+X_2+\ldots+X_n}{n}-\mu|>\varepsilon\}\le\frac{\sigma^2}{n\varepsilon^2}$$

$$\therefore \lim_{n\to\infty} P\{|\frac{X_1+X_2+\ldots+X_n}{n}-\mu|>\varepsilon\}=0$$

The strong law of large numbers

- The strong law of large numbers states the following:

$$P(\lim_{n\to\infty} \frac{X_1 + X_2 + \ldots + X_n}{n} = \mu) = 1$$

- This is stronger than the weak law because this states that the probability of the desired event (that the empirical mean is equal to the actual mean) is equal to 1 given enough samples. The weak laws states that it tends to 1.

- The proof of the strong law is formidable and beyond the scope of our course.

(The incorrect) Law of averages

- As laymen we tend to believe that if something has been going wrong for quite some time, it will suddenly turn right – using the law of averages.

- This supposed law is actually a fallacy – it reflects wishful thinking, and the core mistake is that we mistake the distribution of samples among a small set of outcomes for the distribution of a larger set.

- This is also called as **Gambler's fallacy**.

(The incorrect) Law of averages

- Let's say a gambler *independently* tosses an unbiased coin 20 times, and gets a head each time. He now applies the "law of averages" and believes that it is more likely that the next coin toss will yield a tail.

- The mistake is as follows: The probability of getting all 21 heads = $(1/2)^{21}$. The probability of getting 20 heads and 1 tail also = $(1/2)^{21}$.

Joint distributions/pdfs/pmfs

Jointly distributed random variables

- Many times in statistics, one needs to model relationships between two or more random variables – for example, your CPI at IITB and the annual salary offered to you during placements!

- Another example: average amount of sugar consumed per day and blood sugar level recorded in a blood test.

- Another example: literacy level and crime rate.

Joint CDFs

- Given continuous random variables *X* and *Y*, their **joint cumulative distribution function** (cdf) is defined as:

$$F_{XY}(x, y) = P(X \leq x, Y \leq y)$$

- The distribution of either random variable (called as marginal cdf) can be obtained from the joint distribution as follows:

$$F_X(x) = P(X \leq x, Y \leq \infty) = F_{XY}(x, \infty)$$
$$F_Y(y) = P(X \leq \infty, Y \leq y) = F_{XY}(\infty, y)$$

I'll explain this a few slides further down

- These definitions can extended to handle more than two random variables as well.

Joint PMFs

- Given two discrete random variables X and Y, their **joint probability mass function** (pmf) is defined as:

$$p_{XY}(x_i, y_j) = P(X = x_i, Y = y_j)$$

- The pmf of either random variable (called as marginal pmf) can be obtained from the joint distribution as follows:

$$P\{X = x_i\} = P(\bigcup_j \{X = x_i, Y = y_j\})$$

$$= \sum_j P\{X = x_i, Y = y_j\} = \sum_j p(x_i, y_j)$$

Why?

Joint PMFs: Example

- Consider that in a city 15% of the families are childless, 20% have only one child, 35% have two children and 30% have three children. Let us suppose that male and female child are equally likely and independent.
- What is the probability that a randomly chosen family has no children?
- $P(B = 0, G = 0) = 0.15 = P(\text{no children})$
- Has 1 girl child?
- $P(B=0,G=1)=P(1 \text{ child})\ P(G=1|1 \text{ child}) = 0.2 \times 0.5 = 0.1$
- Has 3 girls?
- $P(B = 0, G = 3) = P(3 \text{ children})\ P(G=3 \mid 3 \text{ Children}) = 0.3 \times (0.5)^3$
- Has 2 boys and 1 girl?
- $P(B = 2, G = 1) = P(3 \text{ children})\ P(B = 2, G = 1| 3 \text{ children}) = 0.3 \times (1/8) \times 3 = 0.1125$ (all 8 combinations of 3 children are equally likely. Out of these there are 3 of the form 2 boys + 1 girl)

Joint PDFs

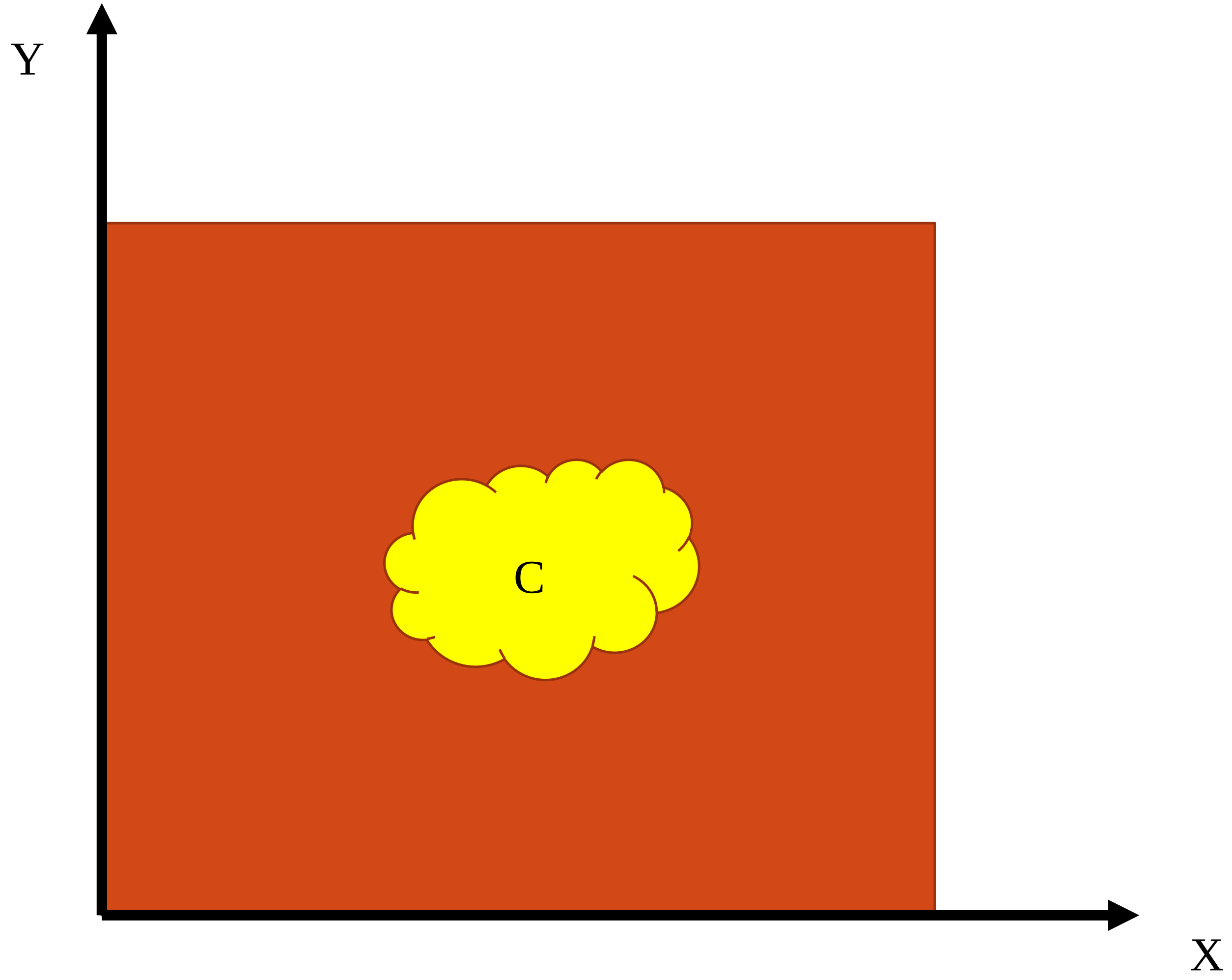
- For two jointly continuous random variables X and Y, the joint pdf is a non-negative function $f_{XY}(x,y)$ such that for *any* set C in the two-dimensional plane, we have:

$$P\{(X,Y)\in C\} = \iint\limits_{(x,y)\in C} f_{XY}(x,y)\,dxdy$$

- The joint CDF can be obtained from the joint PDF as follows:

$$F_{XY}(a,b) = \int_{-\infty}^{a}\int_{-\infty}^{b} f_{XY}(x,y)\,dxdy$$

$$f_{XY}(a,b) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)\Big|_{x=a,y=b}$$



The joint probability that (X,Y) belongs to any arbitrary-shaped region in the XY-plane is obtained by integrating the joint pdf of (X,Y) over that region (eg: region C)

Joint and **marginal** PDFs

- The **marginal pdf** of a random variable can be obtained by integrating the joint pdf w.r.t. the other random variable(s):

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

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

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

$$\int_{-\infty}^{a} f_X(x)dx = \int_{-\infty}^{a}\int_{-\infty}^{\infty} f_{XY}(x,y)dydx$$

$$\therefore F_X(a) = F_{XY}(a,\infty)$$

Independent random variables

- Two continuous random variables are said to be **independent** if and only if:

  $$\forall x, \forall y,\ f_{XY}(x,y) = f_X(x)\, f_Y(y)$$

  i.e., the joint pdf is equal to the product of the marginal pdfs.

- For independent random variables, the joint CDF is also equal to the product of the marginal CDFs:

  $$F_{XY}(x,y) = F_X(x) F_Y(y)$$ Try proving this yourself!

Independent random variables

- Some n continuous random variables are said to be **mutually independent** if and only if:

$$\forall x_1, x_2, \ldots, x_n,$$
$$f_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1) f_{X_2}(x_2) \ldots f_{X_n}(x_n)$$

  i.e., the joint pdf is equal to the product of all n marginal pdfs.

- Note that this condition is *stronger* than pairwise independence!

$$\forall (x_i, x_j), 1 \le i \le n, 1 \le j \le n, i \ne j,$$
$$f_{X_i, X_j}(x_i, x_j) = f_{X_i}(x_i) f_{X_j}(x_j)$$

Independent random variables

- Mutual independence between n random variables implies that they are pairwise independent, or in fact, k-wise independent for any $k < n$.

- But pairwise independence does not necessarily imply mutual independence.

Concept of covariance

- The covariance of two random variables *X* and *Y* is defined as follows:

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

- Further expansion:

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$
$$= E[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y]$$
$$= E[XY] - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y --- why?$$
$$= E[XY] - \mu_X \mu_Y$$
$$= E[XY] - E[X]E[Y]$$

Concept of covariance: properties

- Cov(X,Y) = Cov(Y, X)
- Cov(X, X) = Var(X) [verify this yourself!]
- Cov(aX,Y) = aCov(X,Y) [prove this!]
- Relationship with correlation coefficient:

$$r(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Concept of covariance: properties

$$Cov(X+Z,Y)=Cov(X,Y)+Cov(Z,Y)$$

$\mathrm{Pr}\,oof$:

$$Cov(X+Z,Y)=E[(X+Z)Y]-E[X+Z]E[Y]$$

$$=E[XY+ZY]-E[X]E[Y]-E[Z]E[Y]$$

$$=E[XY]-E[X]E[Y]+E[ZY]-E[Z]E[Y]$$

$$=Cov(X,Y)+Cov(Z,Y)$$

$$Cov(\sum_i X_i,Y)=\sum_i Cov(X_i,Y)$$

$$Cov(\sum_i X_i,\sum_j Y_j)=\sum_i\sum_j Cov(X_i,Y_j)$$

Try proving this yourself! Along similar lines as the previous one.

Concept of covariance: properties

$$Cov(\sum_i X_i, Y) = \sum_i Cov(X_i, Y)$$

$$Cov(\sum_i X_i, \sum_j Y_j) = \sum_i \sum_j Cov(X_i, Y_j)$$

$$Var(\sum_i X_i) = Cov(\sum_i X_i, \sum_i X_i)$$
$$= \sum_i \sum_j Cov(X_i, X_j)$$
$$= \sum_i Cov(X_i, X_i) + \sum_i \sum_{j \neq i} Cov(X_i, X_j)$$
$$= \sum_i Var(X_i) + \sum_i \sum_{j \neq i} Cov(X_i, X_j)$$

Notice that the variance of the sum of random variables is not equal to the sum of their individual variances. This is quite unlike the mean!

Concept of covariance: properties

- For independent random variables X and Y, Cov(X,Y) = 0, i.e. E[XY] = E[X]E[Y].

- Proof:

$$
\begin{aligned}
E[XY] &= \sum_i \sum_j x_i y_j P\{X = x_i, Y = y_j\} \\
&= \sum_i \sum_j x_i y_j P\{X = x_i\} P\{Y = y_j\} \\
&= \sum_i x_i P\{X = x_i\} \sum_j y_j P\{Y = y_j\} \\
&= E[X]E[Y]
\end{aligned}
$$

$$
\begin{aligned}
Cov(X,Y) &= E[(X - \mu_X)(Y - \mu_Y)] \\
&= E[XY] - \mu_X E[X] - \mu_Y E[Y] + \mu_X \mu_Y \\
&= E[XY] - E[X]E[Y] = 0
\end{aligned}
$$

Concept of covariance: properties

- Given random variables X and Y, $\mathrm{Cov}(X,Y) = 0$ does not necessarily imply that X and Y are independent!

- Proof: Construct a counter-example yourself!

Conditional pdf/cdf/pmf

- Given random variables X and Y with joint pdf $f_{XY}(x,y)$, then the conditional pdf of X given $Y = y$ is defined as follows:

$$f_{X|Y}(x \mid y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{\partial}{\partial x}\left(F_{X|Y}(x \mid y)\right)$$

- Conditional cdf $F_{X|Y}(x,y)$:

$$F_{X|Y}(x \mid y) = \lim_{\delta \to 0} P(X \le x \mid y \le Y \le y + \delta) = \int_{-\infty}^{x} f_{X|Y}(z \mid y)\,dz$$

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

http://math.arizona.edu/~jwatkins/m-conddist.pdf

Conditional mean and variance

- Conditional densities or distributions can be used to define the **conditional mean** (also called **conditional expectation**) or **conditional variance** as follows:

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

$$Var(X \mid Y = y) = \int_{-\infty}^{\infty} (x - E(X \mid Y = y))^2 f_{X|Y}(x, y)\,dx$$

Example

$$f(x,y) = 2.4x(2-x-y), 0 < x < 1, 0 < y < 1$$
$$= 0 \text{ otherwise}$$

Find conditional density of X given $Y = y$.

Find conditional mean of X given $Y = y$.