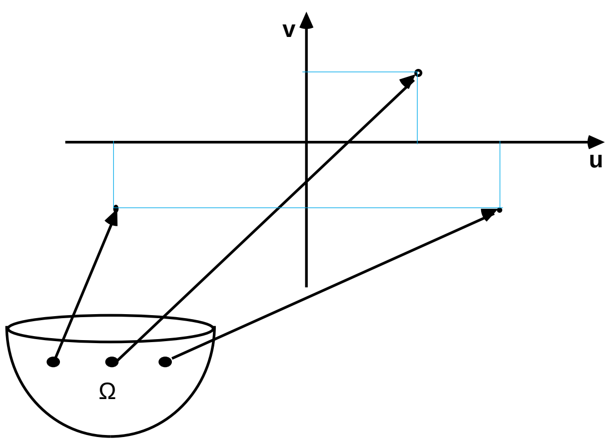
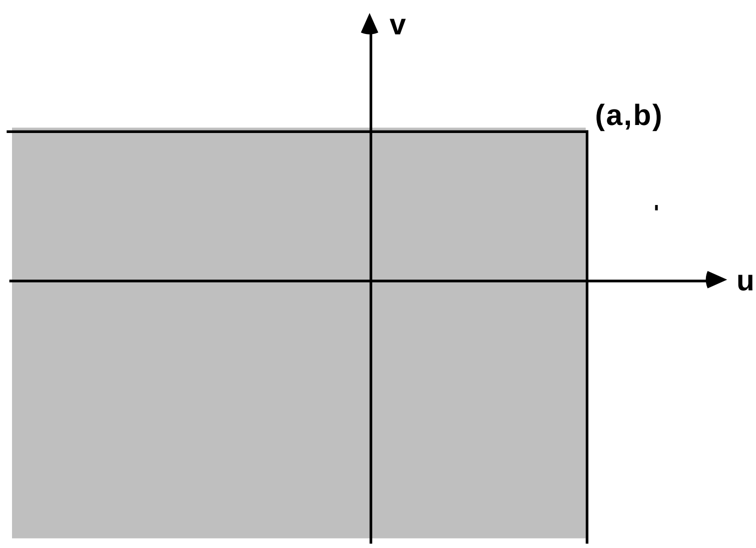
# Chapter 5 Many Random Variables

- A random variable models phenomena in which the experimental outcomes are *numbers*
- A random variable measures one physical parameter
- Different random variables measure different parameters
- Example: Requests for different files arrive at a Web server. Requests have two parameters of interest: the time of arrival and the length of the data file requested
- Sunspot activity causes different levels of fading of radio signals in different frequency bands
- Different random variables measure different parameters
- Experimental observations yield *vectors* or *sequences* or *arrays* of numbers
- The observed vectors are random: we cannot predict beforehand which vector will be observed next
- The mathematical model of a random variable associates numbers with outcomes $\omega$ in a sample space
- Different random variables associate different numbers with each outcome $\omega$
- $[\mathbf{X}(\omega), \mathbf{Y}(\omega), \mathbf{Z}(\omega), \ldots]$ is a *vector* of observations
- Example: $\Omega$ = set of people. One is chosen at random $\mathbf{X}$ = height of person chosen $\mathbf{Y}$ = weight of person chosen
- $(\mathbf{X}, \mathbf{Y})$ is a *random vector* or a *bivariate random variable*
- $(\mathbf{X}, \mathbf{Y})$ maps $\omega \quad \Omega$ onto the *pair* of numbers $(\mathbf{X}(\omega), \mathbf{Y}(\omega))$
- $(\mathbf{X}, \mathbf{Y})$ maps $\omega \quad \Omega$ onto *point* $(\mathbf{X}(\omega), \mathbf{Y}(\omega))$ in the plane
- Coordinate axes are u and v with value of $\mathbf{X}$ plotted along u axis and of $\mathbf{Y}$ along v axis



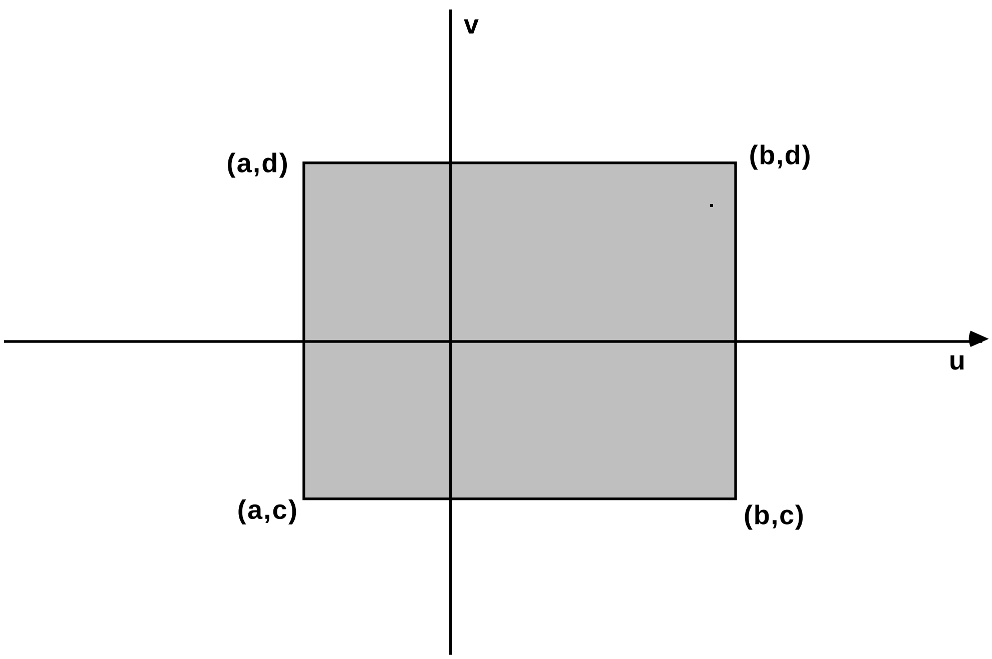

- $\mathbf{X}$ takes on three values, $\mathbf{Y}$ takes on only two values
- Both $\mathbf{X}$ and $\mathbf{Y}$ individually are random variables
- *Together*, $\mathbf{X}$ and $\mathbf{Y}$ create a joint mapping of $\omega \quad \Omega$ onto the *random point* $(\mathbf{X}(\omega), \mathbf{Y}(\omega))$
- As usual, we drop the explicit dependence on $\omega$ when talking about random variables
- $(\mathbf{X}, \mathbf{Y})$ is a random point in the plane
- On succesive trials of the experiment, the outcomes are $\omega_1, \omega_2, \omega_3, \ldots$ and we observe the random points $(\mathbf{X}(\omega_1), \mathbf{Y}(\omega_1)), (\mathbf{X}(\omega_2), \mathbf{Y}(\omega_2)), (\mathbf{X}(\omega_3), \mathbf{Y}(\omega_3)), \ldots$
- The u-coordinate of the random point is $\mathbf{X}(\omega)$ and the v-coordinate is $\mathbf{Y}(\omega)$

- The random point $(\mathbf{X}, \mathbf{Y})$ has more information than either $\mathbf{X}$ or $\mathbf{Y}$; it describes the *joint behavior* of these random variables
- **Example:** Let $\mathbf{X}$ and $\mathbf{Y}$ be discrete random variables with $P\{\mathbf{X} = 0\} = \{\mathbf{Y} = 0\} = 1/2$ and $P\{\mathbf{X} = 1\} = \{\mathbf{Y} = 1\} = 1/2$
- What is the probability that the random point $(\mathbf{X}, \mathbf{Y})$ has value (1,1)? that is, what is $P(\{\mathbf{X} = 1\} \cap \{\mathbf{Y} = 1\})$?
- Given events A and B with probabilities P(A) and P(B), what is $P(A \cap B)$?
- We might have $P(\{\mathbf{X} = 1\}\cap\{\mathbf{Y} = 1\}) = 1/4$; $P(\{\mathbf{X} = 1\}\cap\{\mathbf{Y} = 0\}) = 1/4$
  $P(\{\mathbf{X} = 0\}\cap\{\mathbf{Y} = 1\}) = 1/4$; $P(\{\mathbf{X} = 0\}\cap\{\mathbf{Y} = 0\}) = 1/4$
- We might have $P(\{\mathbf{X} = 1\}\cap\{\mathbf{Y} = 1\}) = 1/3$; $P(\{\mathbf{X} = 1\}\cap\{\mathbf{Y} = 0\}) = 1/6$;
  $P(\{\mathbf{X} = 0\}\cap\{\mathbf{Y} = 1\}) = 1/6$; $P(\{\mathbf{X} = 0\}\cap\{\mathbf{Y} = 0\}) = 1/3$
- Thus, the individual behavior of $\mathbf{X}$ and $\mathbf{Y}$ cannot completely specify their joint behavior
- The joint CDF of $\mathbf{X}$ and $\mathbf{Y}$, or the CDF of the bivariate random variable $(\mathbf{X}, \mathbf{Y})$, or the joint CDF of the random vector $(\mathbf{X}, \mathbf{Y})$, or the CDF of the random point $(\mathbf{X}, \mathbf{Y})$, is
  $F_{\mathbf{X},\mathbf{Y}}(u,v) = P(\{\mathbf{X} \le u\}\cap\{\mathbf{Y} \le v\})$ for all u, v, $-\infty < u, v < \infty$
- Convention: the $\cap$ is always replaced by a comma
- $\{\mathbf{X} \le u, \mathbf{Y} \le v\}$ means that the events $\{\mathbf{X} \le u\}$ **AND** $\{\mathbf{Y} \le v\}$ both have occurred
- DO NOT interpret the event $\{\mathbf{X} \le u, \mathbf{Y} \le v\}$ to mean that at least one of $\{\mathbf{X} \le u\}$ or $\{\mathbf{Y} \le v\}$ occurred
- $F_{\mathbf{X},\mathbf{Y}}(a,b)$ = P{random point $(\mathbf{X},\mathbf{Y})$ is in the region shown}



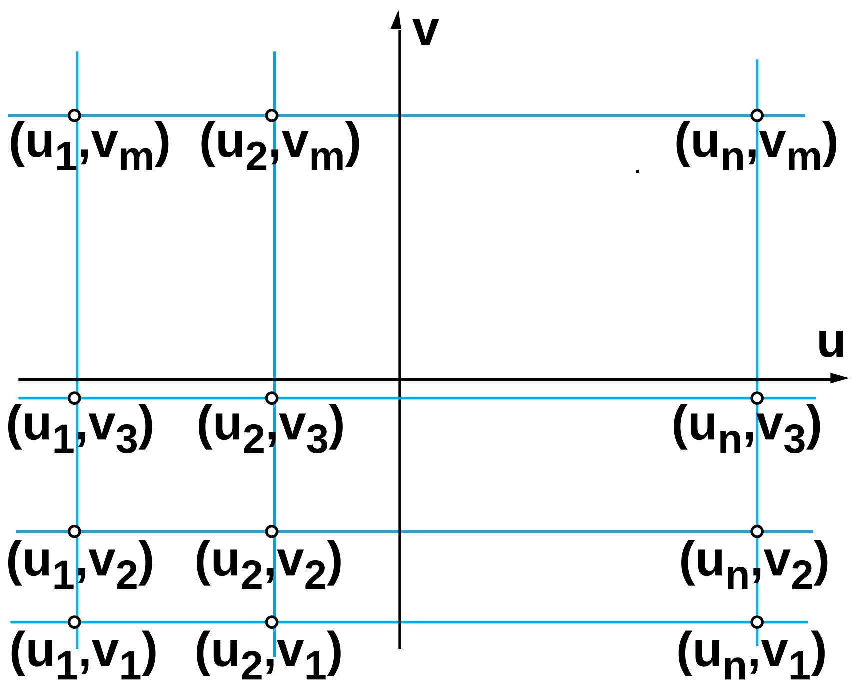

- $F_{\mathbf{X},\mathbf{Y}}(u,v)$ is a function of two variables u and v
- A “graph” of $F_{\mathbf{X},\mathbf{Y}}(u,v)$ is a surface in three-dimensional space (axes u, v, w)
- Since $0 \le F_{\mathbf{X},\mathbf{Y}}(u,v) \le 1$, this surface lies between the planes w = 0 and w = 1
- $F_{\mathbf{X},\mathbf{Y}}(a,b)$ = P{random point $(\mathbf{X},\mathbf{Y})$ is in the region shown}
- $F_{\mathbf{X},\mathbf{Y}}(u,v) = P\{\mathbf{X} \le u, \mathbf{Y} \le v\}$
- If $u \to -\infty$, $\{\mathbf{X} \le u\} \to \varnothing$ and hence $F_{\mathbf{X},\mathbf{Y}}(u,v) \to 0$
- If $v \to -\infty$, $F_{\mathbf{X},\mathbf{Y}}(u,v) \to 0$

- If $v \to \infty$, $\{\mathbf{Y} \le v\} \to \Omega$ and hence $F_{\mathbf{X},\mathbf{Y}}(u,v) \to F_{\mathbf{X}}(u)$
- If $u \to \infty$, $F_{\mathbf{X},\mathbf{Y}}(u,v) \to F_{\mathbf{Y}}(v)$
- Thus, we can obtain $F_{\mathbf{X}}(u)$ and $F_{\mathbf{Y}}(v)$, the individual CDFs of **X** and **Y**, from the joint CDF $F_{\mathbf{X},\mathbf{Y}}(u,v)$
- However, it is not possible to determine $F_{\mathbf{X},\mathbf{Y}}(u,v)$ from $F_{\mathbf{X}}(u)$ and $F_{\mathbf{Y}}(v)$ alone
- When $F_{\mathbf{X}}(u)$ and $F_{\mathbf{Y}}(v)$ are derived from $F_{\mathbf{X},\mathbf{Y}}(u,v)$ by "setting the unneeded variable to ∞", they are called *marginal* CDFs
- Marginal does NOT mean barely making the grade; it means the values of the joint CDF "along the margins"
- Let $v = \beta$ be fixed. Then, $F_{\mathbf{X},\mathbf{Y}}(u,\beta)$ is a nondecreasing right-continuous function of u
- Let $u = \alpha$ be fixed. Then, $F_{\mathbf{X},\mathbf{Y}}(\alpha,v)$ is a nondecreasing right-continuous function of v
- $P(\{\mathbf{X} \le u\} \cup \{\mathbf{Y} \le v\}) = ?$
- $P(\{\mathbf{X} \le u\} \cup \{\mathbf{Y} \le v\}) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  $= P\{\mathbf{X} \le u\} + P\{\mathbf{Y} \le v\} - P\{\mathbf{X} \le u, \mathbf{Y} \le v\} = F_{\mathbf{X}}(u) + F_{\mathbf{Y}}(v) - F_{\mathbf{X},\mathbf{Y}}(u,v)$
- $F_{\mathbf{X}}(u) = F_{\mathbf{X},\mathbf{Y}}(u,\infty)$
- $P\{a < \mathbf{X} \le b, c < \mathbf{Y} \le d\} = ?$




- $P\{a < \mathbf{X} \le b, c < \mathbf{Y} \le d\} = F_{\mathbf{X},\mathbf{Y}}(b,d) - F_{\mathbf{X},\mathbf{Y}}(a,d) - F_{\mathbf{X},\mathbf{Y}}(b,c) + F_{\mathbf{X},\mathbf{Y}}(a,c) \ge 0$
- This inequality must be satisfied for all a < b and all c < d
- Testing that this inequality holds in all cases is nontrivial
- **Generalization:** $\underline{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ is a random vector or n-variate random variable
- $\underline{u} = (u_1, u_2, \ldots, u_n)$
- $\{\underline{\mathbf{X}} \le \underline{u}\} = \{\mathbf{X}_1 \le u_1, \mathbf{X}_2 \le u_2, \ldots, \mathbf{X}_n \le u_n\}$
- Joint CDF of $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ or CDF of $\underline{\mathbf{X}}$ is $F_{\underline{\mathbf{X}}}(\underline{u}) = P\{\underline{\mathbf{X}} \le \underline{u}\}$
  $= P\{\mathbf{X}_1 \le u_1, \mathbf{X}_2 \le u_2, \ldots, \mathbf{X}_n \le u_n\} = F_{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n}(u_1, u_2, \ldots, u_n)$

- If any $u_i \to -\infty$, $F_{\underline{\mathbf{X}}}(\underline{u}) \to 0$
- If some of the $u_i \to +\infty$, the corresponding random variables disappear and we get the marginal CDF of the remaining random variables
- Example: $F_{\mathbf{X}_1,\mathbf{X}_2,\mathbf{X}_3,\mathbf{X}_4}(u_1,\infty,\infty,u_4) = F_{\mathbf{X}_1,\mathbf{X}_4}(u_1,u_4)$
- $F_{\underline{\mathbf{X}}}(\underline{u})$ is a surface in (n+1)-dimensional space
- Fix all $u_i$ except one. $F_{\underline{\mathbf{X}}}(\underline{u})$ is a nondecreasing right-continuous function of the remaining variable
- Generalized rectangle inequality also holds
- In most cases, the joint CDF is cumbersome to work with
- Simpler descriptions of the probabilistic behavior are more commonly used
- Many special cases
- We shall concentrate on the case of two random variables **X** and **Y**
- Degenerate case: If **Y** is a function of **X**, say $\mathbf{Y} = g(\mathbf{X})$, then the coordinates of the random point are $(\mathbf{X},\mathbf{Y}) = (\mathbf{X},g(\mathbf{X}))$
- The random point $(\mathbf{X},\mathbf{Y})$ always lies on the curve $v = g(u)$ in the u-v plane
- All questions about this pair of random variables can be expressed in terms of **X** alone
- $P\{\mathbf{X} \le u, \mathbf{Y} \le v\} = P\{\mathbf{X} \le u, g(\mathbf{X}) \le v\} = P\{\mathbf{X} \le u'\}$ or $P\{a \le \mathbf{X} \le b\}$ or $P\{a \le \mathbf{X} \le b\} + P\{c \le \mathbf{X} \le d$, etc
- **Discrete Random Variables**
- **X** and **Y** are discrete random variables taking on values $\{u_1, u_2, \ldots, u_n\}$ and $\{v_1, v_2, \ldots, v_m\}$ respectively
- The random point $(\mathbf{X},\mathbf{Y})$ can only be one of the m×n points $(u_i,v_j)$, $1 \le i \le n$, $1 \le j \le m$, lying on a m×n (non-uniform) grid



- The joint pmf of $\mathbf{X}$ and $\mathbf{Y}$ is $p_{\mathbf{X},\mathbf{Y}}\{u_i,v_j\} = P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\}$, i = 1, 2, … n; j = 1, 2, … m **(remember commas mean intersection)**
- Think of the joint pmf as an array or m×n matrix

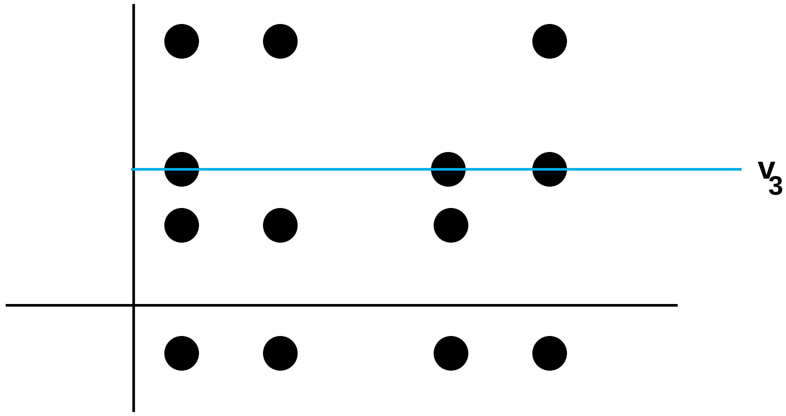
$$\begin{array}{llllllll}
p_{\mathbf{X},\mathbf{Y}}\{u_1,v_m\} & p_{\mathbf{X},\mathbf{Y}}\{u_2,v_m\} & \cdots & \cdots & \cdots & \cdots & \cdots & p_{\mathbf{X},\mathbf{Y}}\{u_n,v_m\} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
p_{\mathbf{X},\mathbf{Y}}\{u_1,v_3\} & p_{\mathbf{X},\mathbf{Y}}\{u_2,v_3\} & \cdots & \cdots & \cdots & \cdots & \cdots & p_{\mathbf{X},\mathbf{Y}}\{u_n,v_3\} \\
p_{\mathbf{X},\mathbf{Y}}\{u_1,v_2\} & p_{\mathbf{X},\mathbf{Y}}\{u_2,v_2\} & \cdots & \cdots & \cdots & \cdots & \cdots & p_{\mathbf{X},\mathbf{Y}}\{u_n,v_2\} \\
p_{\mathbf{X},\mathbf{Y}}\{u_1,v_1\} & p_{\mathbf{X},\mathbf{Y}}\{u_2,v_1\} & \cdots & \cdots & \cdots & \cdots & \cdots & p_{\mathbf{X},\mathbf{Y}}\{u_n,v_1\}
\end{array}$$

- The joint pmf of $\mathbf{X}$ and $\mathbf{Y}$ is $p_{\mathbf{X},\mathbf{Y}}(u_i,v_j) = P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\}$ (remember commas mean intersection)
- Properties of pmfs:
- $p_{\mathbf{X},\mathbf{Y}}(u,v) \geq 0$ for all u and v
- $\sum_{i=1}^{n} \sum_{j=1}^{m} p_{\mathbf{X},\mathbf{Y}}(u_i,v_j) = 1$
- $p_{\mathbf{X}}(u_i) = P\{\mathbf{X} = u_i\} = P\{\mathbf{X} = u_i, \mathbf{Y} = v_1\} + P\{\mathbf{X} = u_i, \mathbf{Y} = v_2\} + \ldots + P\{\mathbf{X} = u_i, \mathbf{Y} = v_m\}$
- $p_{\mathbf{X}}(u_i) = \sum_{j=1}^{m} P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\}$
- For notational convenience, we use $p_{ij}$ to denote $p_{\mathbf{X},\mathbf{Y}}(u_i,v_j) = P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\}$ for i = 1,2,…, n, j = 1,2,…, m
- Think of the joint pmf as an array or table or m×n matrix specifying the probability mass at each of the m×n grid points

| column sums | | | | | | row sums |
|---|---|---|---|---|---|---|
| $\sum\rightarrow$ | $p_{\mathbf{X}}(u_1)$ | $p_{\mathbf{X}}(u_2)$ | • | • | $p_{\mathbf{X}}(u_n)$ | $\sum\downarrow$ |
| $v_m$ | $p_{1m}$ | $p_{2m}$ | • | • | $p_{nm}$ | $p_{\mathbf{Y}}(v_m)$ |
| • | • | • | • | • | • | • |
| • | • | • | • | • | • | • |
| $v_2$ | $p_{12}$ | $p_{22}$ | • | • | $p_{n2}$ | $p_{\mathbf{Y}}(v_2)$ |
| $v_1$ | $p_{11}$ | $p_{21}$ | • | • | $p_{n1}$ | $p_{\mathbf{Y}}(v_1)$ |
| $v\uparrow$ $u\rightarrow$ | $u_1$ | $u_2$ | • | • | $u_n$ | |

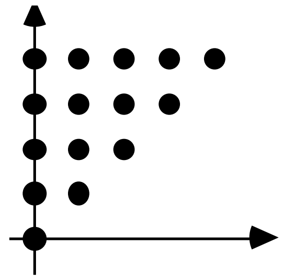
- These notions generalize obviously and naturally to the joint pmf of many discrete random variables
- Joint pmfs are thought of as multidimensional arrays and easily handled by computers
- There are few profoundly interesting joint pmfs
- Suppose that we know that $\mathbf{Y}$ has taken on value $v_3$ on this trial of the experiment
- This fact allows us to deduce additional information about the value taken on by $\mathbf{X}$ on this trial of the experiment

- **Example:** The joint pmf of **X** and **Y** is as shown below



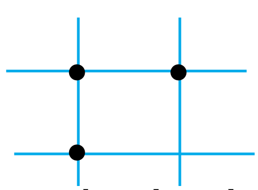

- If $\mathbf{Y} = v_3$, $\mathbf{X}$ can have values $u_1$, $u_3$, or $u_4$, but not $u_2$
- Knowing the value of **Y**, we should *update* the pmf of **X** to *conditional* probabilities (given **Y**) instead of using the unconditional probabilities $p_{\mathbf{X}}(u_i)$
- The conditional probabilities are called the conditional pmf
- The *conditional* pmf of **X** *given* $\mathbf{Y} = v_j$ is
  $$p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j) = P\{\mathbf{X} = u_i|\mathbf{Y} = v_j\} = \frac{P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\}}{P\{\mathbf{Y} = v_j\}} = \frac{p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)}{p_{\mathbf{Y}}(v_j)}$$
- $p_{\mathbf{X},\mathbf{Y}}(u_i,v_j) = 0 \qquad p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j) = 0$
- Conditional pmfs are valid pmfs: given $\mathbf{Y} = v_j$, $p_{\mathbf{X}|\mathbf{Y}}(u|v_j) \geq 0$ for all u, and $\sum_{i=1}^{n} p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j) = 1$
- Unconditional pmfs can be found via the theorem of total probability
- Events $\{\mathbf{Y} = v_1\}$, $\{\mathbf{Y} = v_2\}$, … , $\{\mathbf{Y} = v_m\}$ are a partition
- $P\{\mathbf{X} = u_i\} = \sum_{j=1}^{m} P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\}$
- $p_{\mathbf{X}}(u_i) = \sum_{j=1}^{m} p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j)p_{\mathbf{Y}}(v_j)$
- $p_{\mathbf{X},\mathbf{Y}}(u_i,v_j) = p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j)p_{\mathbf{Y}}(v_j)$
- Bayes’ formula “turns the conditioning around”
- $$p_{\mathbf{Y}|\mathbf{X}}(v_j|u_i) = \frac{P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\}}{P\{\mathbf{X} = u_i\}} = \frac{p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j)p_{\mathbf{Y}}(v_j)}{p_{\mathbf{X}}(u_i)}$$
- **Example:** A Geiger counter detects an $\alpha$-particle with probability p. The emissions are modeled as a Poisson process of rate $\lambda$ per second. **X** = number of particles detected per second **Y** = number of $\alpha$-particles emitted per second
- **Y** is a Poisson random variable with parameter $\lambda$
- *Given* that $\mathbf{Y} = N$, we know that $0 \leq \mathbf{X} \leq N$
- Assumption: detections are independent events
- **X** is a conditionally binomial RV with parameters (N,p)

- *Given* that $\mathbf{Y} = N$, $p_{\mathbf{X}|\mathbf{Y}}(k|N) = \binom{N}{k} p^k(1-p)^{N-k}$, $0 \le k \le N$

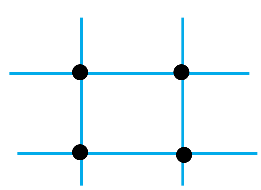


- *Unconditionally*, $\mathbf{X}$ takes on all nonnegative integer values
- $p_{\mathbf{X}}(k) = \sum_{N=k}^{\infty} p_{\mathbf{X}|\mathbf{Y}}(k|N)p_{\mathbf{Y}}(N) = \sum_{N=k}^{\infty} \binom{N}{k} p^k(1-p)^{N-k}\frac{\lambda^N}{N!}\exp(-\lambda) = \frac{(\lambda p)^k}{k!}\exp(-\lambda)\sum_{N=k}^{\infty}\frac{[(1-p)\lambda]^{N-k}}{(N-k)!}$
- $p_{\mathbf{X}}(k) = \frac{(\lambda p)^k}{k!}\exp(-\lambda p)$, $k \ge 0$
- The unconditional pmf of $\mathbf{X}$ is Poisson with parameter $\lambda p$
- The conditional pmf of $\mathbf{X}$ given that $\mathbf{Y} = N$ is binomial with parameter $(N,p)$
- Given that $\mathbf{X} = k$, what can we say about $\mathbf{Y}$? Well, obviously $\mathbf{Y} \ge k$
- $p_{\mathbf{Y}|\mathbf{X}}(N|k) = \frac{p_{\mathbf{X}|\mathbf{Y}}(k|N)p_{\mathbf{Y}}(N)}{p_{\mathbf{X}}(k)} = \frac{[\lambda(1-p)]^{N-k}}{(N-k)!}\exp(-\lambda(1-p))$, $N \ge k$

  = Poisson with parameter $\lambda(1-p)$ but moved right by k
- This formulation also applies to other physical situations, e.g. packet arrivals are a Poisson process; packets are lost if they are not serviced by a deadline; processor can only provide service with probability p.
- **Independent Discrete Random Variables**
- Discrete random variables $\mathbf{X}$ and $\mathbf{Y}$ are independent if knowing value of $\mathbf{Y}$ tells you nothing about the value of $\mathbf{X}$
- If for all $v_j$, the conditional pmf $p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j)$ equals $p_{\mathbf{X}}(u_i)$, we have learnt nothing new
- Equivalently, $p_{\mathbf{X}|\mathbf{Y}}(u_i|v_j) = p_{\mathbf{X}}(u_i)$ iff $p_{\mathbf{X},\mathbf{Y}}(u_i,v_j) = p_{\mathbf{X}}(u_i)p_{\mathbf{Y}}(v_j)$,
  that is, $P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\} = P\{\mathbf{X} = u_i\}P\{\mathbf{Y} = v_j\}$ for *all* $i = 1,2,\ldots, n$, and *all* $j = 1,2,\ldots, m$
- Random variables $\mathbf{X}$ and $\mathbf{Y}$ are independent if their joint pmf equals the product of the marginal pmfs
- Also, the joint CDF equals the product of the marginal CDFs
- $F_{\mathbf{X},\mathbf{Y}}(u,v) = F_{\mathbf{X}}(u)F_{\mathbf{Y}}(v)$ for all u and v, $-\infty < u, v < \infty$
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent, their joint pmf cannot be zero at any of the $m \times n$ grid points
- This gives an eyeball test for deciding that $\mathbf{X}$ and $\mathbf{Y}$ cannot possibly be independent



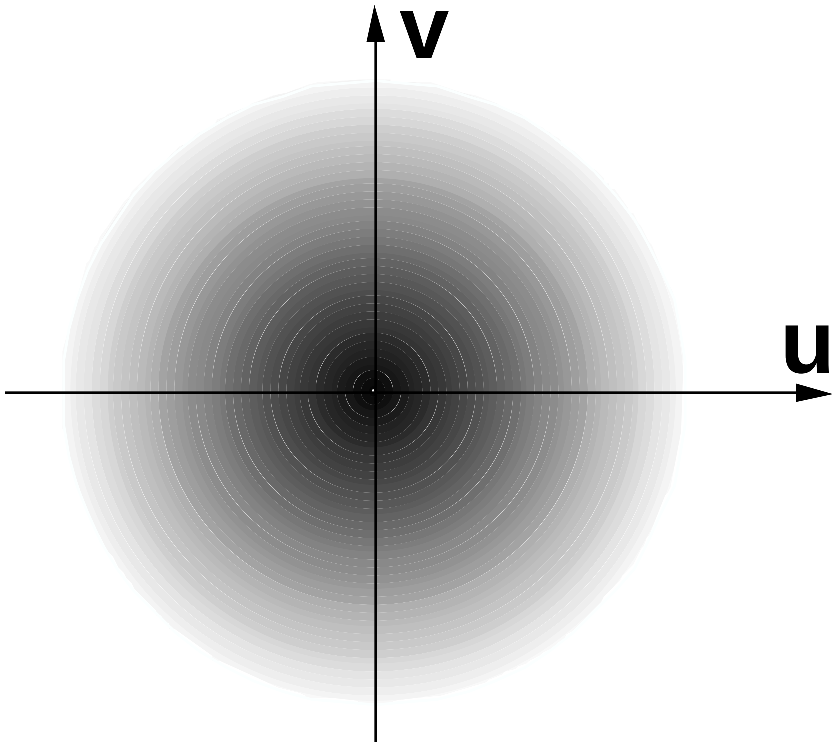
*Cannot* be independent



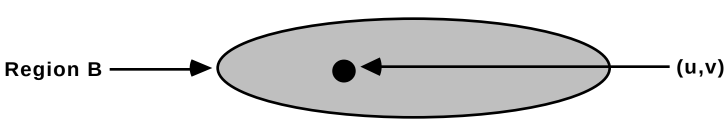
Possibly independent

- Discrete random variables $\mathbf{X}_1$, $\mathbf{X}_2$, …, $\mathbf{X}_n$ are said to be independent if their joint pmf equals the product of the marginal pmfs, that is,
  $p_{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n}(u_1,u_2,\ldots,u_n) = p_{\mathbf{X}_1}(u_1)\, p_{\mathbf{X}_2}(u_2)\ldots p_{\mathbf{X}_n}(u_n)$ for all $u_1, u_2, \ldots, u_n$
- Equivalently, $F_{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n}(u_1,u_2,\ldots,u_n) = F_{\mathbf{X}_1}(u_1)F_{\mathbf{X}_2}(u_2)\ldots F_{\mathbf{X}_n}(u_n)$ for all $u_1, u_2, \ldots, u_n$
- When are random vectors $\underline{\mathbf{X}} = (\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n)$ and $\underline{\mathbf{Y}} = (\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_m)$ independent?
- Random vectors $\underline{\mathbf{X}} = (\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n)$ and $\underline{\mathbf{Y}} = (\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_m)$ are said to be independent if $F_{\underline{\mathbf{X}},\underline{\mathbf{Y}}}(\underline{u},\underline{v}) = F_{\underline{\mathbf{X}}}(\underline{u})F_{\underline{\mathbf{Y}}}(\underline{v})$
- Note that the $\mathbf{X}_i$'s or the $\mathbf{Y}_i$'s need not be independent among themselves
- **Jointly continuous random variables**
- Suppose that $\mathbf{X}$ and $\mathbf{Y}$ are random variables
- $(\mathbf{X}, \mathbf{Y})$ is a *random vector* or a *bivariate random variable* that maps $\omega$ $\Omega$ onto the *pair* of numbers $(\mathbf{X}(\omega), \mathbf{Y}(\omega))$
- Equivalently, $(\mathbf{X}, \mathbf{Y})$ maps $\omega$ $\Omega$ onto *point* $(\mathbf{X}(\omega), \mathbf{Y}(\omega))$ in the plane
- $(\mathbf{X},\mathbf{Y})$ is called a *random point in the plane*
- If $\mathbf{X}$ and $\mathbf{Y}$ are discrete, $(\mathbf{X},\mathbf{Y})$ takes on a discrete set of positions in the plane
- If $\mathbf{X}$ and $\mathbf{Y}$ are continuous random variables, what can be said about the random point $(\mathbf{X},\mathbf{Y})$?
- **Degenerate case:** If $\mathbf{Y} = g(\mathbf{X})$, then $(\mathbf{X},\mathbf{Y}) = (\mathbf{X},g(\mathbf{X}))$ always lies on the curve defined by $v = g(u)$ in the u-v plane
- All probabilities involving this bivariate random variable $(\mathbf{X},\mathbf{Y})$ can be expressed in terms of $\mathbf{X}$ alone
- More generally, $(\mathbf{X},\mathbf{Y})$ can be any point in a *region* of the plane
- **Example:** $(\mathbf{X},\mathbf{Y})$ can be any point inside the square region with vertices (0,0), (0,1), (1,0) and (1,1)
- **Example:** $(\mathbf{X},\mathbf{Y})$ can be any point in the plane
- If $(\mathbf{X},\mathbf{Y})$ can be any point in a *region* of the plane, $\mathbf{X}$ and $\mathbf{Y}$ are called *jointly continuous* random variables
- Other verbiage: $(\mathbf{X},\mathbf{Y})$ is a jointly continuous random variable; $(\mathbf{X},\mathbf{Y})$ is a jointly continuous bivariate random variable
- If $(\mathbf{X},\mathbf{Y})$ is jointly continuous, then $\mathbf{X}$ and $\mathbf{Y}$ are individually also continuous random variables
- The degenerate case discussed above shows that continuous random variables $\mathbf{X}$ and $\mathbf{Y}$ need not be jointly continuous as well — the random point $(\mathbf{X},\mathbf{Y})$ could well always lie on a curve instead of anywhere in a region. This is the definitive characteristic of joint continuity
- If $(\mathbf{X},\mathbf{Y})$ is jointly continuous, the random point $(\mathbf{X},\mathbf{Y})$ may occur *anywhere* in a region
- Repeated trials may show more heavy clustering in some parts of the region than in other parts
- The random point may be more likely to be in some locations than in others

- The random point may be more likely to be in some locations than in others




- Discrete random variables create point (probability) masses at discrete points: $p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$ = probability mass at point $(u_i,v_j)$
- Jointly continuous random variables *spread* the probability mass in the plane
- There is no *point mass* at any point in the plane: instead, the probability mass is spread with varying *density* $f_{\mathbf{X},\mathbf{Y}}(u,v)$ over a region of the plane
- $f_{\mathbf{X},\mathbf{Y}}(u,v)$ = density of the mass at the point (u,v) in the plane
- The joint probability density function (joint pdf) $f_{\mathbf{X},\mathbf{Y}}(u,v)$, defined for all u and v, $-\infty < u, v < \infty$, describes the variation in the *density* of the probability mass in the plane
- $f_{\mathbf{X},\mathbf{Y}}(u,v) \geq 0$ for all u and v
- $f_{\mathbf{X},\mathbf{Y}}(u,v)$ describes a surface that lies above the u-v plane
- For our picture, $f_{\mathbf{X},\mathbf{Y}}(u,v)$ is maximum at (0,0) indicating that, on repeated trials, the heaviest clustering of the random points was observed at the origin
- $f_{\mathbf{X},\mathbf{Y}}(u,v)$ is not a probability; it is the *density* of the probability mass at (u,v)
- $f_{\mathbf{X},\mathbf{Y}}(u,v)$ is measured in units of probability mass/area
- To get a probability, we must multiply $f_{\mathbf{X},\mathbf{Y}}(u,v)$ by an area
- All this is similar to the pdf of a (single) continuous random variable:
  $P\{u \leq \mathbf{X} \leq u+\delta\} = f_{\mathbf{X}}(u)\times\delta$
- B = small area including the point (u,v)



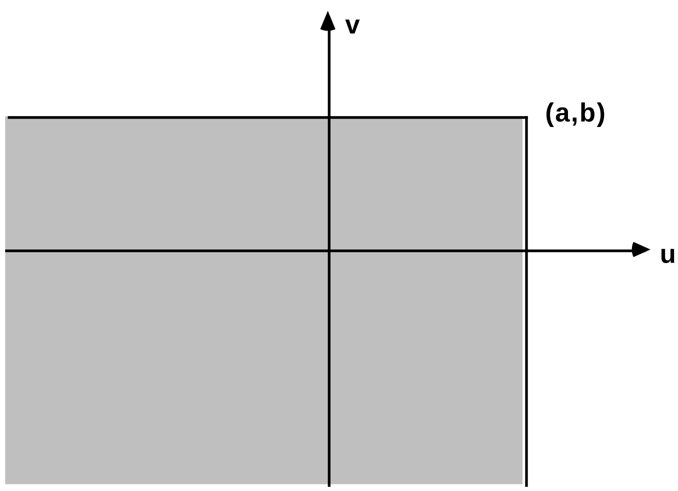

- $P\{(\mathbf{X},\mathbf{Y})\quad B\} \approx f_{\mathbf{X},\mathbf{Y}}(u,v)\times$Area of region B

- Approximation is good if B has small area, and poor if the area of B is large
- $P\{(\mathbf{X},\mathbf{Y}) \in B\} \approx f_{\mathbf{X},\mathbf{Y}}(u,v) \times$ Area of region B
- $f_{\mathbf{X},\mathbf{Y}}(u,v)$ is a surface above the u-v plane
- $f_{\mathbf{X},\mathbf{Y}}(u,v) \times$ Area of region B ≈ volume between $f_{\mathbf{X},\mathbf{Y}}(u,v)$ surface and region B in the u-v plane

  $P\{u \le \mathbf{X} \le u + \delta u,\ v \le \mathbf{Y} \le v + \delta v\}$

  $= P\{(\mathbf{X},\mathbf{Y}) \in \text{rectangle of area } \delta u \times \delta v\} \approx f_{\mathbf{X},\mathbf{Y}}(u,v) \times \delta u \times \delta v$
- Poor approximation if area is large
- For larger areas, take the sum of such approximations
- Let B denote an arbitrary region in the plane
- The limiting form of the sum is

  $$P\{(\mathbf{X},\mathbf{Y}) \in B\} = \iint_{(u,v) \in B} f_{\mathbf{X},\mathbf{Y}}(u,v)\,du\,dv = \text{volume between the pdf surface and the region B}$$
- $F_{\mathbf{X},\mathbf{Y}}(a,b) = P\{$random point $(\mathbf{X},\mathbf{Y})$ is in the region shown$\}$




- $F_{\mathbf{X},\mathbf{Y}}(a,b)$ = volume below pdf surface in the shaded region
- $$F_{\mathbf{X},\mathbf{Y}}(a,b) = P\{\mathbf{X} \le a,\ \mathbf{Y} \le b\} = \int_{-\infty}^{a} \left[ \int_{-\infty}^{b} f_{\mathbf{X},\mathbf{Y}}(u,v)\,dv \right] du = \int_{-\infty}^{b} \left[ \int_{-\infty}^{a} f_{\mathbf{X},\mathbf{Y}}(u,v)\,du \right] dv$$
- Integration can be done in any order
- $$F_{\mathbf{X},\mathbf{Y}}(\infty,\infty) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)\,du\,dv = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)\,dv\,du = 1$$
- Total volume between the surface $f_{\mathbf{X},\mathbf{Y}}(u,v)$ and the u-v plane equals 1
- All valid joint pdfs satisfy the following two properties:
- $f_{\mathbf{X},\mathbf{Y}}(u,v) \ge 0$ for all u and v
- $$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)\,dv\,du = 1$$

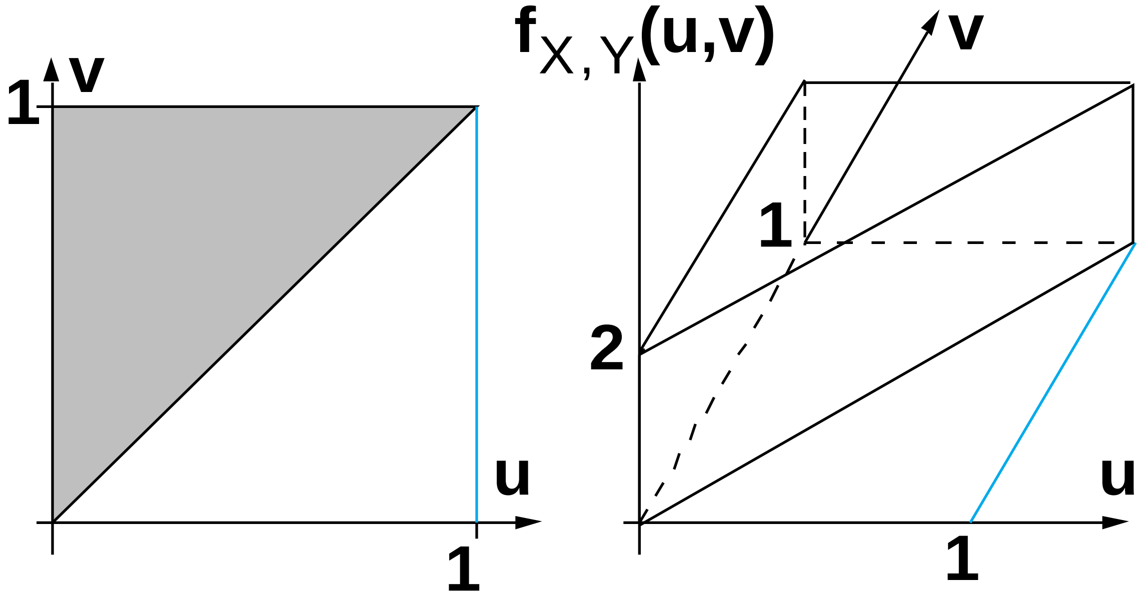
- $F_{\mathbf{X},\mathbf{Y}}(a,b) = P\{\mathbf{X} \le a,\ \mathbf{Y} \le b\} = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{\mathbf{X},\mathbf{Y}}(u,v)\,dv\,du$
- Hence, $f_{\mathbf{X},\mathbf{Y}}(u,v) = \frac{\partial^2}{\partial u \partial v} F_{\mathbf{X},\mathbf{Y}}(u,v)$
- **What is the derivative of an integral?**
- An integral is an area, so its value is a constant, and the derivative is therefore 0 ??
- $g(x;\alpha)$ = function of x with $\alpha$ as a parameter e.g. $\exp(-\alpha x)$
- $a(\alpha)$ and $b(\alpha)$ are given functions of $\alpha$
- $G(\alpha) = \int_{a(\alpha)}^{b(\alpha)} g(x;\alpha)\,dx$
- $\frac{\partial}{\partial \alpha} G(\alpha) = \int_{a(\alpha)}^{b(\alpha)} \frac{\partial}{\partial \alpha} g(x;\alpha)\,dx + g(b(\alpha);\alpha)\frac{db(\alpha)}{d\alpha} - g(a(\alpha);\alpha)\frac{da(\alpha)}{d\alpha}$
- All derivatives are assumed to exist
- **Special case:** If limits do not depend on $\alpha$, then
- $G(\alpha) = \int_{a}^{b} g(x;\alpha)\,dx$
- $\frac{\partial}{\partial \alpha} G(\alpha) = \int_{a}^{b} \frac{\partial}{\partial \alpha} g(x;\alpha)\,dx$
- **Special case:** If integrand does not depend on $\alpha$, then
- $G(\alpha) = \int_{a(\alpha)}^{b(\alpha)} g(x)\,dx$
- $\frac{\partial}{\partial \alpha} G(\alpha) = g(b(\alpha))\frac{db(\alpha)}{d\alpha} - g(a(\alpha))\frac{da(\alpha)}{d\alpha}$
- $F_{\mathbf{X},\mathbf{Y}}(a,b) = P\{\mathbf{X} \le a,\ \mathbf{Y} \le b\} = \int_{-\infty}^{a} \left[ \int_{-\infty}^{b} f_{\mathbf{X},\mathbf{Y}}(u,v)\,dv \right] du = \int_{-\infty}^{a} g(u;b)\,du$
  where $g(u;b)$ = inner integral is *not* a function of a
- $F_{\mathbf{X},\mathbf{Y}}(a,b) = \int_{-\infty}^{a} g(u;b)\,du$
- $\frac{\partial}{\partial a} F_{\mathbf{X},\mathbf{Y}}(a,b) = g(a;b)\frac{\partial a}{\partial a} = g(a;b) = \int_{-\infty}^{b} f_{\mathbf{X},\mathbf{Y}}(a,v)\,dv$
- Differentiate this with respect to b

- $f_{\mathbf{X},\mathbf{Y}}(a,b) = \frac{\partial^2}{\partial a \partial b} F_{\mathbf{X},\mathbf{Y}}(a,b)$
- $F_{\mathbf{X}}(a) = F_{\mathbf{X},\mathbf{Y}}(a,\infty) = \int_{-\infty}^{a} \left[ \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)dv \right] du$
- $F_{\mathbf{Y}}(b) = \int_{-\infty}^{b} \left[ \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)du \right] dv$
- $\frac{\partial}{\partial a} F_{\mathbf{X}}(a) = f_{\mathbf{X}}(a) = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(a,v)dv$
- $\frac{\partial}{\partial a} F_{\mathbf{Y}}(b) = f_{\mathbf{Y}}(b) = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,b)du$
- Moral: to find the marginal pdfs, integrate w.r.t. to the unwanted variable
- **Generalization**:
- $\underline{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$ is a random vector or n-variate random variable
- Joint pdf of $\mathbf{X}_1$, $\mathbf{X}_2$, … , $\mathbf{X}_n$ or pdf of $\underline{\mathbf{X}}$ is denoted by $f_{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n}(u_1,u_2,\ldots,u_n)$
- $P\{u \leq \mathbf{X} \leq u+\delta u\} = P\{\cap\ (u_i \leq \mathbf{X}_i \leq u_i + \delta u_i)\} = f_{\underline{\mathbf{X}}}(\underline{u})\delta u_1 \delta u_2 \ldots \delta u_n$
- $P\{\underline{\mathbf{X}} \in B\} = \iint\cdots\iint_{B} f_{\underline{\mathbf{X}}}(\underline{u})d\underline{u}$ = n-dimensional volume
- $F_{\underline{\mathbf{X}}}(\underline{a}) = F_{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n}(a_1,a_2,\ldots,a_n) = \int_{-\infty}^{a_1}\int_{-\infty}^{a_2}\cdots\int_{-\infty}^{a_n} f_{\underline{\mathbf{X}}}(\underline{u})d\underline{u}$
- To find the joint pdf of a subset of $\mathbf{X}_1$, $\mathbf{X}_2$, … , $\mathbf{X}_n$, integrate $f_{\underline{\mathbf{X}}}(\underline{u})$ from -∞ to ∞ w.r.t the unwanted $u_i$'s
- Recall that all valid joint pdfs satisfy
- $f_{\mathbf{X},\mathbf{Y}}(u,v) \geq 0$ for all u and v
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)dvdu = 1$
- $F_{\mathbf{X},\mathbf{Y}}(a,b) = P\{\mathbf{X} \leq a, \mathbf{Y} \leq b\} = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{\mathbf{X},\mathbf{Y}}(u,v)dv\,du$
- $f_{\mathbf{X},\mathbf{Y}}(u,v) = \frac{\partial^2}{\partial u \partial v} F_{\mathbf{X},\mathbf{Y}}(u,v)$
- $f_{\mathbf{X}}(u) = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)dv$
- $f_{\mathbf{Y}}(v) = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)du$
- The random point $(\mathbf{X},\mathbf{Y})$ is uniformly distributed on a region B if the joint pdf is of the form

$$f_{\mathbf{X},\mathbf{Y}}(u,v) = \begin{cases} c, & (u,v) \in B, \\ 0, & \text{elsewhere.} \end{cases}$$

- Volume = c×Area(B) equals 1. Hence, $c = [\text{Area}(B)]^{-1}$

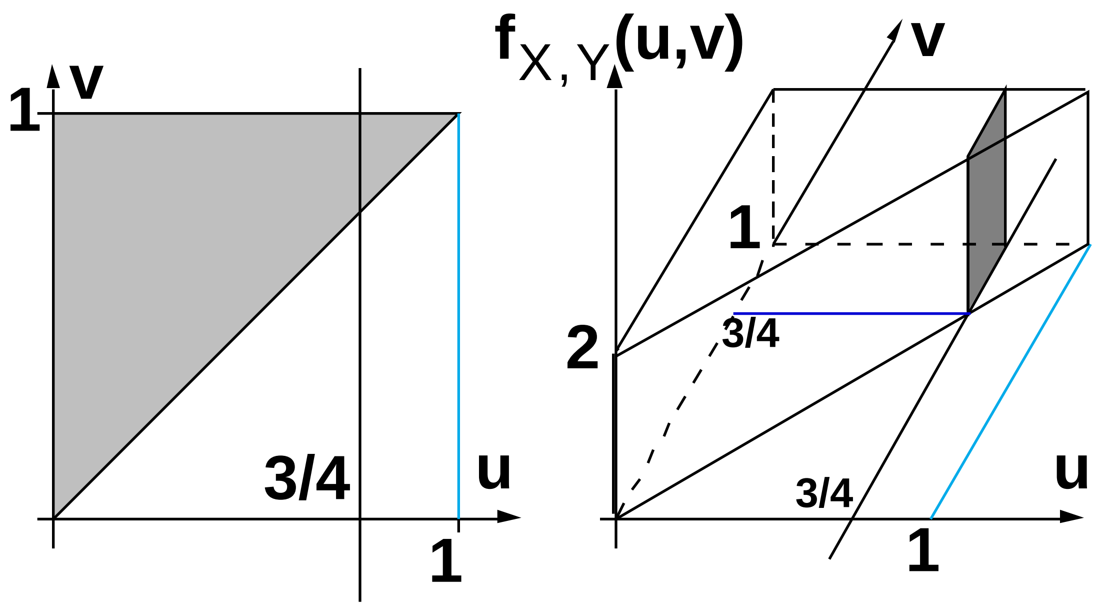
- **Example:** $f_{\mathbf{X},\mathbf{Y}}(u,v) = \begin{cases} 2, & 0 < u < v < 1, \\ 0, & \text{elsewhere.} \end{cases}$




- $f_{\mathbf{X}}(u) = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)dv$

- Fix a value for u, say u = 3/4. Then, $f_{\mathbf{X}}(3/4) = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(3/4,v)dv$

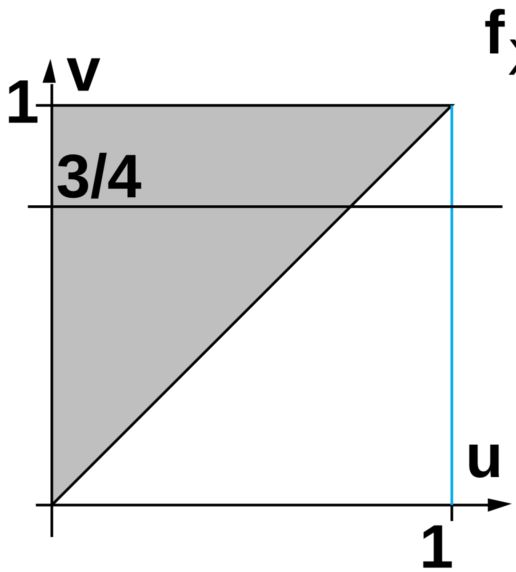
- u = 3/4 is the equation of a vertical line in the u-v plane

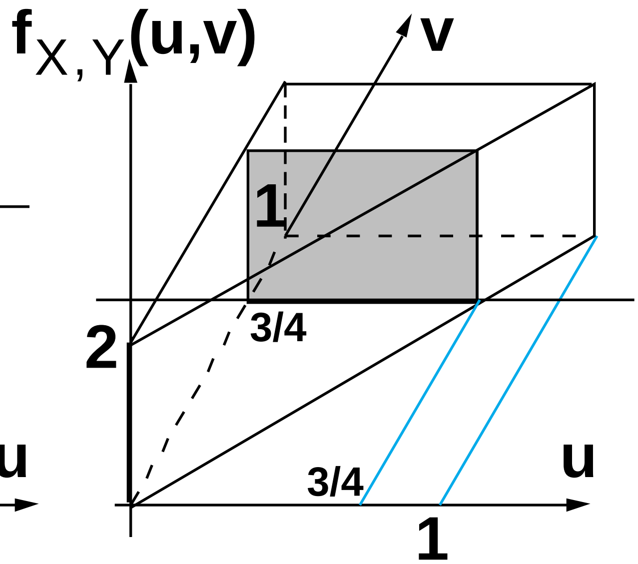



- $f_{\mathbf{X}}(3/4)$ = Area of rectangle = $2\times(1-3/4) = 1/2$

- More generally, for $0 < u < 1$, $f_{\mathbf{X}}(u) = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)dv$ = area of rectangle = $2(1-u)$

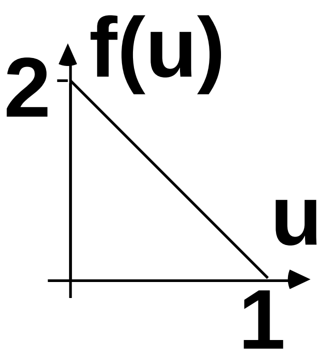
  For $u \le 0$ or $u \ge 1$, $f_{\mathbf{X},\mathbf{Y}}(u,v) = 0$ and hence integral is 0
- $f_{\mathbf{X}}(u) = \begin{cases} 2(1-u), & 0 < u < 1, \\ 0, & \text{elsewhere.} \end{cases}$

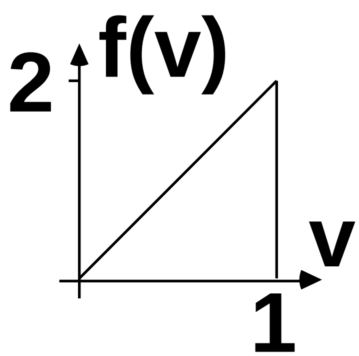



- $f_{\mathbf{Y}}(3/4)$ = Area of rectangle = $2 \times 3/4 = 3/2$
- More generally, for $0 < v < 1$, $f_{\mathbf{Y}}(v) = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)du$ = area of rectangle = $2v$

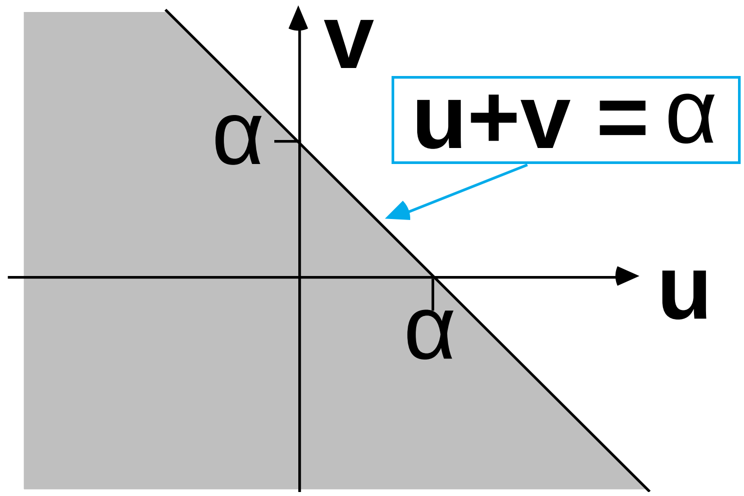
  For $v \le 0$ or $v \ge 1$, $f_{\mathbf{X},\mathbf{Y}}(u,v) = 0$ and hence integral is 0
- $f_{\mathbf{Y}}(v) = \begin{cases} 2v, & 0 < v < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- $f_{\mathbf{X}}(u) = \begin{cases} 2(1-u), & 0 < u < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- It is easily verified that these are valid pdfs



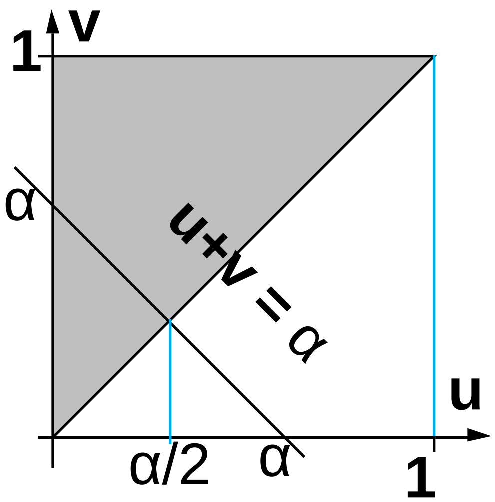



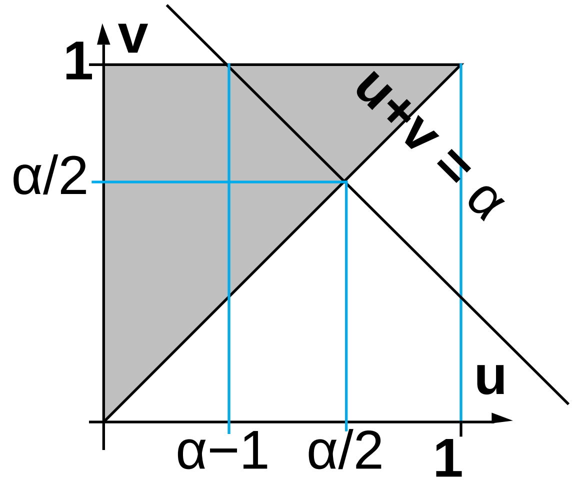

- For *any* joint pdf $f_{\mathbf{X},\mathbf{Y}}(u,v)$,
  $f_{\mathbf{X}}(u_0)$ = area of *cross-section* of $f_{\mathbf{X},\mathbf{Y}}(u,v)$ (parallel to v-axis) at point $u = u_0$
  $f_{\mathbf{Y}}(v_0)$ = area of *cross-section* of $f_{\mathbf{X},\mathbf{Y}}(u,v)$ (parallel to u-axis) at point $v = v_0$

- **Functions of random variables**
- What is $P\{\mathbf{X} + \mathbf{Y} \le \alpha\}$?



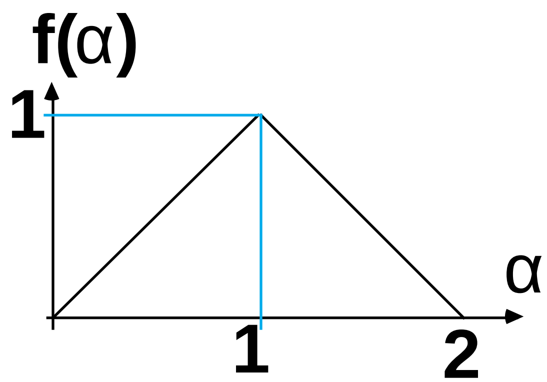

- $P\{\mathbf{X} + \mathbf{Y} \le \alpha\} = P\{(\mathbf{X},\mathbf{Y})$ shaded region$\}$
- **Example:** $f_{\mathbf{X},\mathbf{Y}}(u,v) = \begin{cases} 2, & 0 < u < v < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- Since $\mathbf{X}$ and $\mathbf{Y}$ have values in the range (0,1), $0 < \mathbf{X} + \mathbf{Y} < 2$
- $P\{\mathbf{X} + \mathbf{Y} \le \alpha\} = \begin{cases} 0 & \text{if } \alpha \le 0, \\ 1 & \text{if } \alpha \ge 2. \end{cases}$
- What if $0 < \alpha < 2$?
- For $0 < \alpha \le 1$, $P\{\mathbf{X}+\mathbf{Y} \le \alpha\} = 2\left[\frac{1}{2}\times\alpha\times(\alpha/2)\right] = \frac{\alpha^2}{2}$



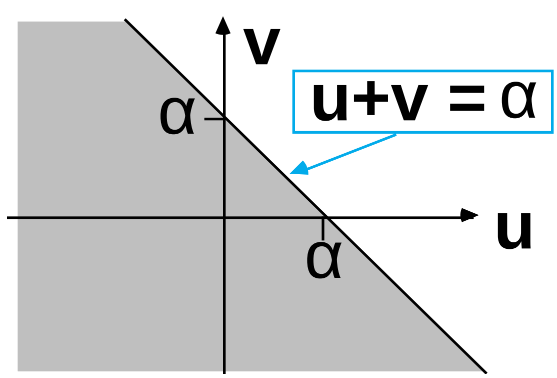



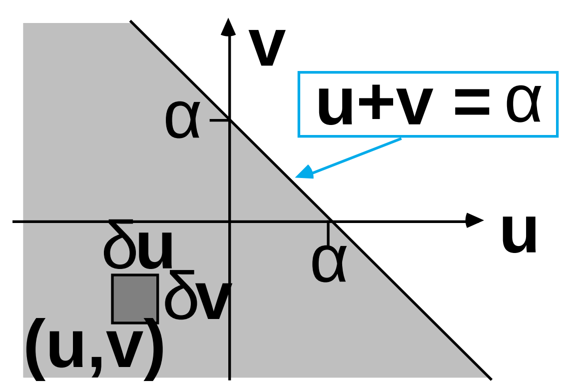

- For $1 \le \alpha < 2$, $P\{\mathbf{X}+\mathbf{Y} > \alpha\} = 2\left[\frac{1}{2}(2-\alpha)(1-\alpha/2)\right] = \frac{(2-\alpha)^2}{2}$

- $$P\{\mathbf{X} + \mathbf{Y} \le \alpha\} = \begin{cases} 0, & \text{if } \alpha \le 0, \\ \dfrac{\alpha^2}{2}, & \text{if } 0 < \alpha \le 1, \\ 1 - \dfrac{(2-\alpha)^2}{2}, & \text{if } 1 \le \alpha < 2, \\ 1, & \text{if } \alpha \ge 2. \end{cases}$$
- Let $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$
- $$F_{\mathbf{Z}}(\alpha) = P\{\mathbf{Z} \le \alpha\} = P\{\mathbf{X}+\mathbf{Y} \le \alpha\} = \begin{cases} 0, & \text{if } \alpha \le 0, \\ \dfrac{\alpha^2}{2}, & \text{if } 0 < \alpha \le 1, \\ 1 - \dfrac{(2-\alpha)^2}{2}, & \text{if } 1 \le \alpha < 2, \\ 1, & \text{if } \alpha \ge 2. \end{cases}$$
- Differentiate to get pdf of $\mathbf{Z}$: $$f_{\mathbf{Z}}(\alpha) = \begin{cases} \alpha, & \text{if } 0 < \alpha \le 1, \\ 2-\alpha, & \text{if } 1 \le \alpha < 2, \\ 0, & \text{elsewhere.} \end{cases}$$



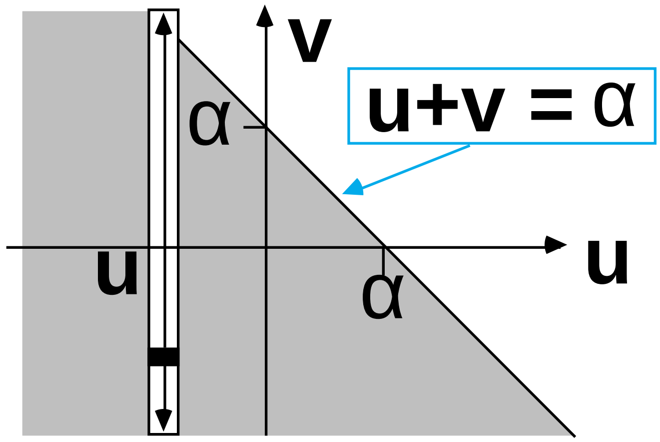

- More generally, $P\{\mathbf{X} + \mathbf{Y} \le \alpha\} = P\{(\mathbf{X},\mathbf{Y})$ shaded region$\}$ and we need to compute an integral. This is how we set it up.




- $P\{(\mathbf{X},\mathbf{Y})$ shaded region$\} = \sum\sum f_{\mathbf{X},\mathbf{Y}}(u,v)\times\delta u\times\delta v$

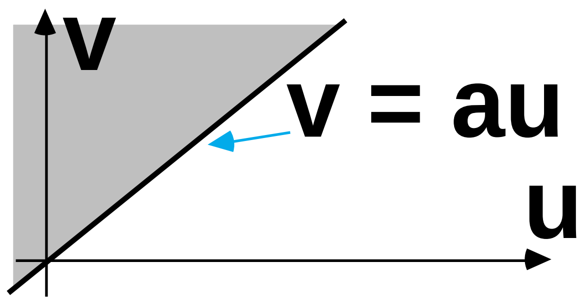
- Move little box to cover the entire shaded region
- Fix the value of u and let v vary, carrying box with it




- Limits on v are –∞ to α–u
- Repeat with different u values: $P\{\mathbf{X} + \mathbf{Y} \le \alpha\} = \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{\alpha-u} f_{\mathbf{X},\mathbf{Y}}(u,v)dvdu$
- $P\{\mathbf{X} + \mathbf{Y} \le \alpha\} = \int_{v=-\infty}^{\infty} \int_{u=-\infty}^{\alpha-v} f_{\mathbf{X},\mathbf{Y}}(u,v)dudv$ if we fix v and let u vary first
- Let $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$
- $F_{\mathbf{Z}}(\alpha) = P\{\mathbf{Z} \le \alpha\} = P\{\mathbf{X}+\mathbf{Y} \le \alpha\} = \int_{u=-\infty}^{\infty} \int_{v=-\infty}^{\alpha-u} f_{\mathbf{X},\mathbf{Y}}(u,v)dvdu$
- $f_{\mathbf{Z}}(\alpha) = \frac{d}{d\alpha} F_{\mathbf{Z}}(\alpha) = ?$
- Differentiate the integral (in two steps)!
- $\frac{\partial}{\partial\alpha} G(\alpha) = \int_a^b \frac{\partial}{\partial\alpha} g(x;\alpha)dx$
- $f_{\mathbf{Z}}(\alpha) = \int_{u=-\infty}^{\infty} \frac{\partial}{\partial\alpha} \int_{v=-\infty}^{\alpha-u} f_{\mathbf{X},\mathbf{Y}}(u,v)dvdu$
- $\frac{\partial}{\partial\alpha} \int_a^{b(\alpha)} g(x)dx = g(b(\alpha))\frac{db(\alpha)}{d\alpha}$
- $f_{\mathbf{Z}}(\alpha) = f_{\mathbf{X}+\mathbf{Y}}(\alpha) = \int_{u=-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,\alpha-u)du = \int_{v=-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(\alpha-v,v)dv$
- This magical formula is usually too messy to apply to any particular problem
- $f_{\mathbf{X}+\mathbf{Y}}(\alpha) = \int_{u=-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,\alpha-u)du$
- **Example:** $f_{\mathbf{X},\mathbf{Y}}(u,v) = \begin{cases} 2, & 0 < u < v < 1, \\ 0, & \text{elsewhere.} \end{cases}$

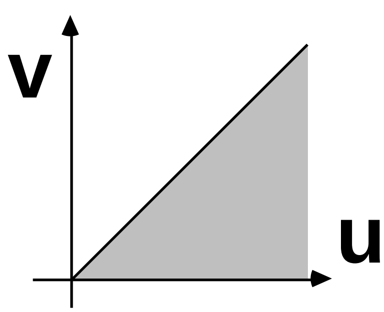
  $f_{\mathbf{X},\mathbf{Y}}(u,\alpha-u) = \begin{cases} 2, & 0 < u < \alpha-u < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- $0 < u < \alpha-u$ $\quad 0 < u < \alpha/2$ $u < \alpha-u < 1$ $\quad \alpha-1 < u < \alpha/2$

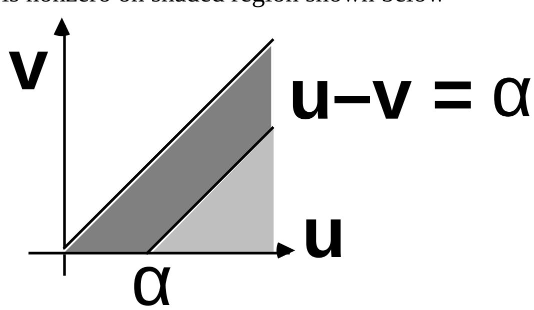
- $0 < u < \alpha/2$ and $\alpha-1 < u < \alpha/2$
- If $0 < \alpha \le 1$, $\alpha-1 \le 0$ and thus $f_{\mathbf{X},\mathbf{Y}}(u,\alpha-u) = 2$ for $0 < u < \alpha/2$
- $f_{\mathbf{X}+\mathbf{Y}}(\alpha) = \int_{u=-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,\alpha-u)du = \int_{u=0}^{\alpha/2} 2\,du = \alpha$
- $0 < u < \alpha/2$ and $\alpha-1 < u < \alpha/2$
- If $1 \le \alpha < 2$, $0 \le \alpha-1 < 1$ $f_{\mathbf{X},\mathbf{Y}}(u,\alpha-u) = 2$ for $\alpha-1 < u < \alpha/2$
- $f_{\mathbf{X}+\mathbf{Y}}(\alpha) = \int_{u=-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,\alpha-u)du = \int_{u=\alpha-1}^{\alpha/2} 2\,du = 2-\alpha$
- $P\{\mathbf{Y} > a\mathbf{X}\} = P\{(\mathbf{X},\mathbf{Y})$ shaded region$\} = \int_{u=-\infty}^{\infty} \int_{v=au}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)dvdu$ if we fix u and let v vary first




- **Fundamental notion:** to find the probability that the random point $(\mathbf{X},\mathbf{Y})$ satisfies some condition such as $\{\mathbf{X}+\mathbf{Y} \le \alpha\}$ or $\{\mathbf{Y} > a\mathbf{X}\}$ or $\{\mathbf{X}^2+\mathbf{Y}^2 \le 1\}$, *identify* the region in the u-v plane, and *find* the *volume under the pdf* in this region
- In some cases, no integration is necessary because pdf is simple function
- Otherwise, set up limits on the integral of the pdf and evaluate
- Discrete random variables? Just add all the probability mass in the region of interest
- **Example:** $(\mathbf{X},\mathbf{Y})$ has joint pdf $f_{\mathbf{X},\mathbf{Y}}(u,v) = \begin{cases} e^{-u}, & 0 < v < u < \infty, \\ 0, & \text{elsewhere.} \end{cases}$

  Find the pdf of $\mathbf{Z} = \mathbf{X}-\mathbf{Y}$ Joint pdf is nonzero on shaded region shown below

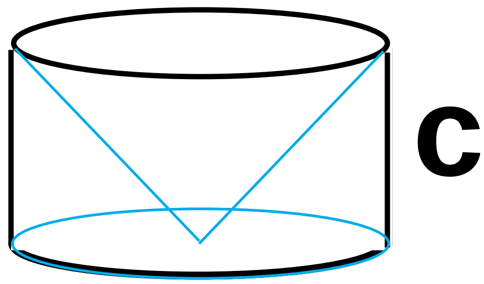


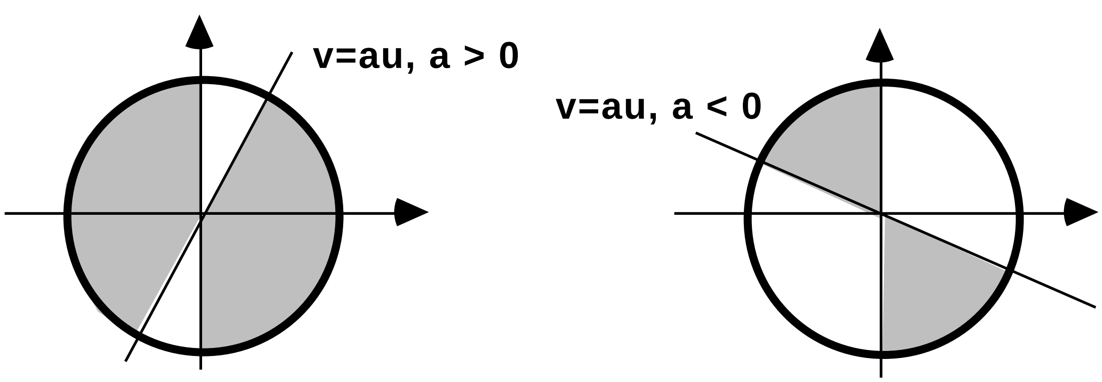

- $\mathbf{X} > \mathbf{Y}$ always so that $\mathbf{Z} > 0$
- For $\alpha > 0$, $P\{\mathbf{Z} < \alpha\} = P\{\mathbf{X}-\mathbf{Y} < \alpha\}$ = volume in deep shaded region above

- For $\alpha > 0$, $P\{\mathbf{Z} < \alpha\} = \int_{v=0}^{\infty} \int_{u=v}^{v+\alpha} \exp(-u)du = \int_{v=0}^{\infty} \exp(-v) - \exp(-v-\alpha)\, dv = 1 - \exp(-\alpha)$

  **Z** is exponential RV with parameter 1
- **Example:** $(\mathbf{X},\mathbf{Y})$ has joint pdf $f_{\mathbf{X},\mathbf{Y}}(u,v) = \begin{cases} c\sqrt{u^2+v^2}, & 0 < u^2+v^2 < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- What is the value of c?
- What is the pdf of $\mathbf{R} = \sqrt{\mathbf{X}^2+\mathbf{Y}^2}$

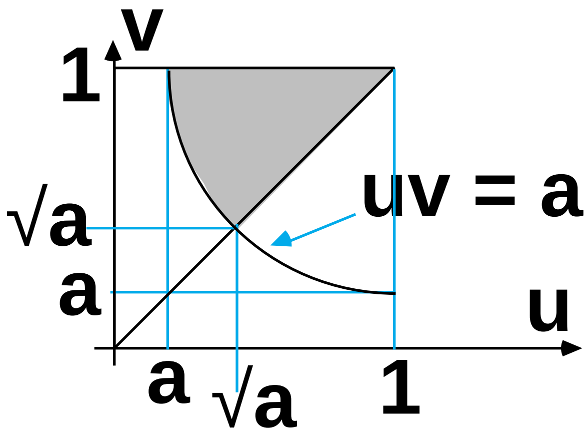



- Cylinder – cone must have volume = 1. But, volume = $\pi c - \pi c/3 = 2\pi c/3$ $c = 3/2\pi$
- **R** has value between 0 and 1
- For $0 < \alpha < 1$, $P\{\mathbf{R} < \alpha\} = \int\int_{u^2+v^2 \le \alpha^2} \frac{3}{2\pi}\sqrt{u^2+v^2}dudv = \int_{r=0}^{\alpha} \int_{\theta=0}^{2\pi} \frac{3}{2\pi} r\, rd\theta\, dr = r^3\Big|_0^{\alpha} = \alpha^3$
- Thus, $f_{\mathbf{R}}(\alpha) = \begin{cases} 3\alpha^2, & 0 < \alpha < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- What is the pdf of $\mathbf{Z} = \mathbf{Y}/\mathbf{X}$?
- **Z** takes on values in $(-\infty, \infty)$

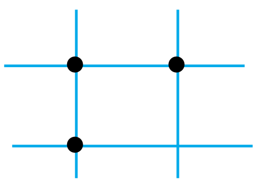
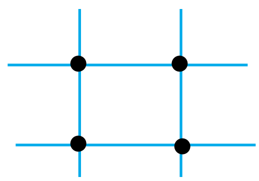


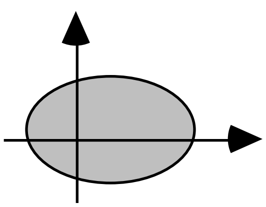
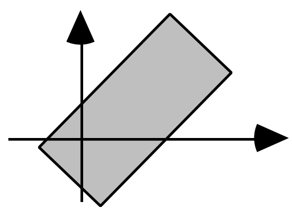
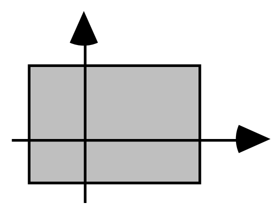

- $P\{\mathbf{Z} < a\} = \frac{1}{2} + \frac{1}{\pi}\arctan(a)$
- $P\{\mathbf{Z} < a\} = \frac{1}{2} + \frac{1}{\pi}\arctan(a)$ for all values of a

- Hence, $f_{\mathbf{Z}}(a) = \dfrac{1}{\pi(1 + a^2)}$, $-\infty < a < \infty$ which is a Cauchy pdf
- Note that *all circularly symmetric* pdfs $f_{\mathbf{X},\mathbf{Y}}(u,v)$ will give the same result!
- **Example:** $f_{\mathbf{X},\mathbf{Y}}(u,v) = \begin{cases} 2, & 0 < u < v < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- What is the pdf of $\mathbf{Z} = \mathbf{XY}$?
- Since $0 < \mathbf{X}, \mathbf{Y} < 1$, $0 < \mathbf{Z} < 1$
- For $0 < a < 1$, we will find $P\{\mathbf{Z} > a\}$ and obtain the pdf $f_{\mathbf{Z}}(a)$ from this

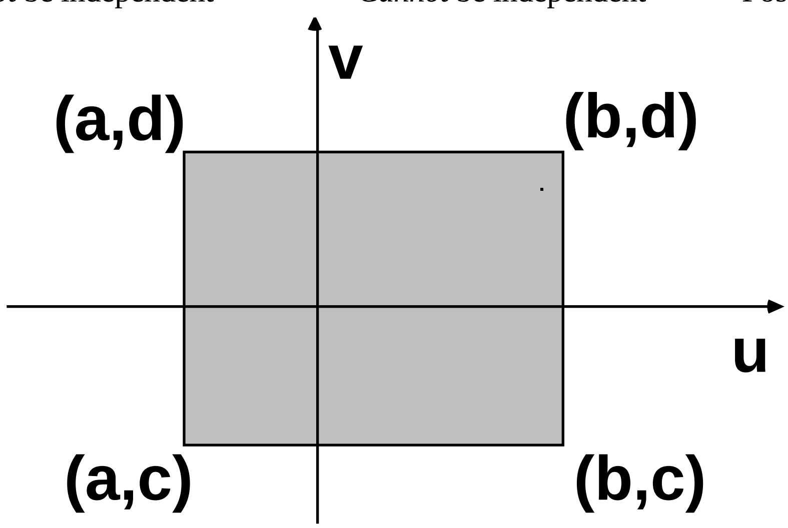



- $P\{\mathbf{Z} > a\} = \displaystyle\int_{v=\sqrt{a}}^{1} \int_{u=a/v}^{v} 2 \, du dv = \int_{v=\sqrt{a}}^{1} 2(v - a/v) \, dv = v^2 - 2a \ln v \Big|_{\sqrt{a}}^{1} = 1 - a - a \ln a$
- Thus $f_{\mathbf{Z}}(a) = -\ln a$, $0 < a < 1$.
- **Independent Random Variables**
- Two random variables $\mathbf{X}$ and $\mathbf{Y}$ are said to be independent if their joint CDF equals the product of the marginal CDFs, that is, if $F_{\mathbf{X},\mathbf{Y}}(u,v) = F_{\mathbf{X}}(u)F_{\mathbf{Y}}(v)$ for all u and v, $-\infty < u, v < \infty$
- Definition applies to all types of random variables: discrete, jointly continuous, mixed, etc
- Discrete random variables are independent if $p_{\mathbf{X},\mathbf{Y}}(u_i,v_j) = p_{\mathbf{X}}(u_i)p_{\mathbf{Y}}(v_j)$, that is, if $P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\} = P\{\mathbf{X} = u_i\}P\{\mathbf{Y} = v_j\}$ for all i = 1,2,…, n, and all j = 1,2,…, m
- Joint pmf = product of marginal pmfs
- Jointly continuous random variables are independent if
  $f_{\mathbf{X},\mathbf{Y}}(u,v) = \dfrac{\partial^2}{\partial u \partial v} F_{\mathbf{X},\mathbf{Y}}(u,v) = \dfrac{\partial^2}{\partial u \partial v} F_{\mathbf{X}}(u)F_{\mathbf{Y}}(v) = f_{\mathbf{X}}(u)f_{\mathbf{Y}}(v)$ for all u and v, $-\infty < u, v < \infty$
- Joint pdf = product of marginal pdfs
- $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are said to be independent if the joint CDF equals the product of the marginal CDFs, that is, if $F_{\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n}(u_1,u_2,\ldots,u_n) = F_{\mathbf{X}_1}(u_1)\, F_{\mathbf{X}_2}(u_2) \ldots F_{\mathbf{X}_n}(u_n)$
- Also holds if CDF is replaced by pdf or pmf
- $\underline{\mathbf{X}} = (\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n)$ and $\underline{\mathbf{Y}} = (\mathbf{Y}_1,\mathbf{Y}_2,\ldots,\mathbf{Y}_m)$ are independent if $F_{\underline{\mathbf{X}},\underline{\mathbf{Y}}}(\underline{u},\underline{v}) = F_{\underline{\mathbf{X}}}(\underline{u})F_{\underline{\mathbf{Y}}}(\underline{v})$
- Note that the $\mathbf{X}_i$'s or the $\mathbf{Y}_i$'s need not be independent among themselves
- Physical independence is assumed from problem, and stochastic independence reflects this assumption
- We *assume* that $\mathbf{X}$ and $\mathbf{Y}$ are physically independent and *set* the joint CDF/pdf/pmf to be the product of the marginal CDFs/pdfs/pmfs

- Meaning of independence: knowing which value of **Y** occurred on a trial tells you nothing new about **X**
- For all a < b and any c, $P\{a \le \mathbf{X} < b \mid \mathbf{Y} = c\} = P\{a \le \mathbf{X} < b\}$
- Knowing **Y** = c doesn't help
- **Example:** Suppose that $f_{\mathbf{X},\mathbf{Y}}(u,v) = \begin{cases} 2, & 0 < u < v < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- Both **X** and **Y** can take on values in (0,1)
- If we know that **Y** = 3/4 on a trial, we know that 0 < **X** < 3/4
- Hence: not independent
- Eyeball test for independence
- It is *necessary* that the joint pdf/pmf be nonzero on a region/grid that is a rectangle with sides parallel to the axes
- Rectangle test is *not* sufficient: we also need to check $F_{\mathbf{X},\mathbf{Y}}(u,v) = F_{\mathbf{X}}(u)F_{\mathbf{Y}}(v)$ for all u and v, $-\infty < u, v < \infty$
-  

  Cannot be independent Possibly independent
- 
-   

  *Cannot* be independent *Cannot* be independent Possibly independent




- $f_{\mathbf{X},\mathbf{Y}}(u,v) = 1/[(b-a)(d-c)]$
- (**X**,**Y**) is uniformly distributed on (a,b)×(c,d), (i.e. rectangle with vertices at (a,c), (b,c), (b,d) and (a,d)) if and only if **X** and **Y** are independent and uniformly distributed on (a,b) and (c,d) respectively

- **Sums of independent random variables:** $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$
- If discrete random variables $\mathbf{X}$ and $\mathbf{Y}$ are integer-valued, then

$$P\{\mathbf{X} + \mathbf{Y} = n\} = P\{\mathbf{Z} = n\} = \sum_k P\{\mathbf{X}=k,\ \mathbf{Y}=n-k\} = \sum_k p_{\mathbf{X},\mathbf{Y}}(k,n-k)$$

- If $\mathbf{X}$ and $\mathbf{Y}$ are independent, then $P\{\mathbf{X} + \mathbf{Y} = n\} = p_{\mathbf{Z}}(n) = \sum_k p_{\mathbf{X},\mathbf{Y}}(k,n-k) = \sum_k p_{\mathbf{X}}(k)p_{\mathbf{Y}}(n-k)$ is the *discrete convolution* of the marginal pmfs of $\mathbf{X}$ and $\mathbf{Y}$
- **Important Special Cases:**
- **Binomial** = # of successes on n trials where P(success) = p
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent binomial random variables with parameters (m, p) and (n, p), then $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ is a binomial random variable with parameters (m + n, p)
- See text for proof (Example 3e, p. 271)
- **Negative binomial** = waiting time for the n-th success
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent negative binomial random variables with parameters (m, p) and (n, p), then $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ is a negative binomial random variable with parameters (m + n, p)
- If $\mathbf{X}_i$, $1 \le i \le n$ are independent binomial (or negative binomial) random variables with parameters $(N_i, p)$, then $\sum_i \mathbf{X}_i$ is binomial (negative binomial) with parameters $(\sum_i N_i, p)$
- **Poisson** = number of random points in given time interval
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent Poisson random variables with parameters $\lambda_{\mathbf{X}}$ and $\lambda_{\mathbf{Y}}$, then $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ is a Poisson random variable with parameter $\lambda_{\mathbf{X}} + \lambda_{\mathbf{Y}}$
- $\mathbf{X}$ and $\mathbf{Y}$ are the numbers of random points in (0, t] and (t, t + s] respectively
- $\mathbf{X}$ and $\mathbf{Y}$ are independent Poisson parameters $\mu t$ and $\mu s$
- $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ = number of random points in (0, t + s] = Poisson with parameter $\mu(t + s)$
- If $\mathbf{X}_i$, $1 \le i \le n$ are independent Poisson random variables with parameters $\lambda_i$, then $\sum_i \mathbf{X}_i$ is a Poisson random variable with parameter $\sum_i \lambda_i$
- If $(\mathbf{X},\mathbf{Y})$ is jointly continuous, $f_{\mathbf{Z}}(\alpha) = \int_{u=-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,\alpha-u)\,du$
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent, $f_{\mathbf{Z}}(\alpha) = \int_{u=-\infty}^{\infty} f_{\mathbf{X}}(u)f_{\mathbf{Y}}(\alpha-u)\,du = f_{\mathbf{X}}*f_{\mathbf{Y}} = f_{\mathbf{Y}}*f_{\mathbf{X}}$ = convolution of marginal pdfs
- Note: On page 265 of the text it is shown that $F_{\mathbf{X}+\mathbf{Y}}(\alpha) = \int_{u=-\infty}^{\infty} F_{\mathbf{X}}(\alpha-v)f_{\mathbf{Y}}(v)\,dv$
- $F_{\mathbf{X}+\mathbf{Y}}$ is called the convolution of the CDFs $F_{\mathbf{X}}$ and $F_{\mathbf{Y}}$. Not so! $F_{\mathbf{X}+\mathbf{Y}} = F_{\mathbf{X}}*f_{\mathbf{Y}} \neq F_{\mathbf{X}}*F_{\mathbf{Y}}$

- If $\mathbf{X}$ and $\mathbf{Y}$ are independent **gamma random variables** with parameters (s, λ) and (t, λ), then $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ is a gamma random variable with parameters (s + t, λ)
- Generalization to the sum of many independent gamma random variables with same scale parameter λ but (possibly) different order parameters: the result is a gamma random variable with same scale parameter λ and order parameter = sum of order parameters

- $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent $N(\mu_i, \sigma_i^2)$ random variables. Then $\sum_i \mathbf{X}_i$ is a $N(\sum_i \mu_i, \sum_i \sigma_i^2)$ random variable
- The sum of **Gaussian random variables** is a Gaussian random variable

- More generally, $\sum_i a_i \mathbf{X}_i$ is a $N(\sum_i a_i\mu_i, \sum_i a_i^2\sigma_i^2)$ random variable
- The sum of non-independent Gaussian random variables is also Gaussian
- $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent random variables
- Then, $g_1(\mathbf{X}_1), g_2(\mathbf{X}_2), \ldots, g_n(\mathbf{X}_n)$ are also independent random variables
- $g(\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_k)$ and $h(\mathbf{X}_{k+1}, \ldots \mathbf{X}_n)$ are independent
- **Functions of random variables**
- **Example:** $\mathbf{X}$ and $\mathbf{Y}$ are independent unit Gaussian random variables
- $f_{\mathbf{X},\mathbf{Y}}(u,v) = \frac{1}{2\pi}\exp\left[-\frac{u^2+v^2}{2}\right]$
- $\mathbf{R} = \sqrt{\mathbf{X}^2+\mathbf{Y}^2}$; $\mathbf{Z} = \mathbf{R}^2 = \mathbf{X}^2+\mathbf{Y}^2$
- $f_{\mathbf{R}}(\alpha) = ?$
- $f_{\mathbf{Z}}(\beta) = ?$

- For $\alpha \geq 0$, $P\{\mathbf{R} > \alpha\} = 1 - F_{\mathbf{R}}(\alpha) = \iint_{u^2+v^2 > \alpha^2} \frac{1}{2\pi}\exp\left[-\frac{u^2+v^2}{2}\right]du\,dv$

$$= \int_{r=\alpha}^{\infty}\int_{\theta=0}^{2\pi} \frac{1}{2\pi}\exp(-r^2/2)\, r\,d\theta\, dr = \exp(-\alpha^2/2)$$

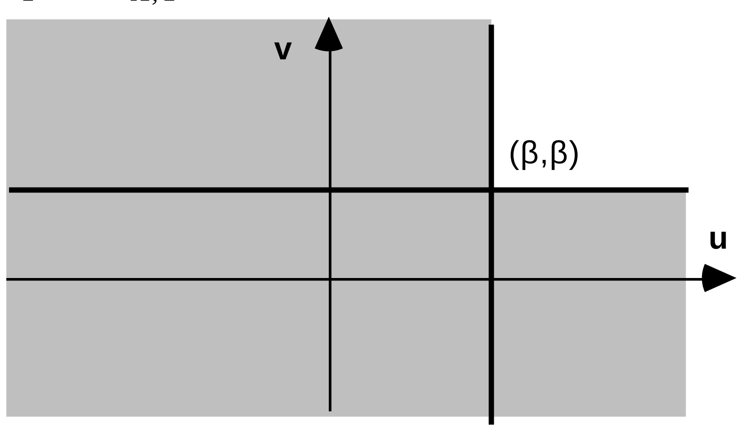
- $f_{\mathbf{R}}(\alpha) = \begin{cases} \alpha\exp(-\alpha^2/2), & \alpha \geq 0, \\ 0, & \alpha < 0. \end{cases}$
- This is a Rayleigh pdf (which also models the lifetimes of systems with linearly increasing hazard rates)
- What if we began with $N(0,\sigma^2)$ instead of $N(0,1)$ variables?
- Pdf of $a\mathbf{X}$ is $f_{a\mathbf{X}}(v) = \frac{1}{|a|} f_{\mathbf{X}}\left(\frac{v}{a}\right)$
- If $\mathbf{X}$ is $N(0,1)$, $\sigma\mathbf{X}$ is $N(0,\sigma^2)$
- $\sqrt{(\sigma\mathbf{X})^2+(\sigma\mathbf{Y})^2} = \sigma\mathbf{R}$ where $\mathbf{R}$ has pdf $(\alpha)\exp(-\alpha^2/2)$
- $f_{\sigma\mathbf{R}}(v) = (v/\sigma^2)\exp(-v^2/2\sigma^2)$ is the general Rayleigh pdf
- DSP microprocessors work on complex baseband signals in I and Q channels

- Noise voltages in channels are modeled as independent $N(0,\sigma^2)$ random variables
- Noise amplitude is Rayleigh random variable with pdf $(v/\sigma^2)\exp(-v^2/2\sigma^2)$
- For $\beta \geq 0$, $P\{\mathbf{Z} > \beta\} = 1 - F_{\mathbf{Z}}(\beta) = \int\int_{u^2+v^2 > \beta} \frac{1}{2\pi}\exp\left[-\frac{u^2+v^2}{2}\right]dudv$

  $= \int_{r=\sqrt{\beta}}^{\infty} \int_{\theta=0}^{2\pi} \frac{1}{2\pi}\exp(-r^2/2)\, rd\theta\, dr = \exp(-\beta/2)$
- $f_{\mathbf{Z}}(\beta) = \begin{cases} (1/2)\exp(-\beta/2), & \beta \geq 0, \\ 0, & \beta < 0. \end{cases}$
- This is an exponential pdf with parameter 1/2, a.k.a. a gamma pdf with parameters (1,1/2)
- If $\mathbf{X}$ and $\mathbf{Y}$ were $N(0,\sigma^2)$, we would have obtained gamma pdf with parameters $(1,1/2\sigma^2)$
- If $\mathbf{X}$ is $N(0,\sigma^2)$, $\mathbf{X}^2$ has gamma pdf with parameter $(1/2,1/2\sigma^2)$
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent, so are $\mathbf{X}^2$ and $\mathbf{Y}^2$
- $\mathbf{X}^2 + \mathbf{Y}^2$ has gamma pdf with parameters $(1,1/2\sigma^2)$
- If $\mathbf{X}_i$, $1 \leq i \leq n$, are independent $N(0,\sigma^2)$ $\mathbf{X}_i^2$ are independent gamma random variables

  with parameters $(1/2,1/2\sigma^2)$ $\sum_i \mathbf{X}_i^2$ has gamma pdf with parameters $(n/2,1/2\sigma^2)$.
- This is often called a $\chi^2$ pdf with n degrees of freedom
- **Example:** $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ are independent $N(0,\sigma^2)$ random variables
- $\mathbf{W} = \mathbf{X}^2 + \mathbf{Y}^2 + \mathbf{Z}^2$ has a gamma pdf with parameters $(3/2,1/2\sigma^2)$
- $f_{\mathbf{W}}(\alpha) = \sqrt{\frac{\alpha}{2\pi\sigma^3}}\ \exp(-\alpha/2\sigma^2)$
- Gamma random variable with parameters $(t, \lambda)$ has expected value $t/\lambda$
- $E[\mathbf{W}] = (3/2)/(1/2\sigma^2) = 3\sigma^2$
- $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$: the velocity of a molecule as measured along three perpendicular axes
- $(1/2)m\mathbf{W}$ is the kinetic energy of the particle
- Average kinetic energy $= E[(1/2)m\mathbf{W}] = (1/2)mE[\mathbf{W}] = (3/2)m\sigma^2 = (3/2)kT$
- $f_{\mathbf{W}}(\alpha) = \sqrt{\frac{\alpha}{2\pi\sigma^3}}\ \exp(-\alpha/2\sigma^2)$ • $\sigma^2 = kT/m$
- Kinetic energy $\mathbf{H} = (1/2)m\mathbf{W}$ has Maxwell-Boltzmann pdf $f_{\mathbf{H}}(\beta) = \frac{2}{\sqrt{\pi}}(kT)^{-3/2}\sqrt{\beta}\exp(-\beta/kT)$
- This is a gamma pdf; what are its parameters?
- $\mathbf{V} = \sqrt{\mathbf{W}} = \sqrt{\mathbf{X}^2 + \mathbf{Y}^2 + \mathbf{Z}^2}$ is the "speed" of the molecule

- The pdf of **V** is $f_{\mathbf{V}}(\gamma) = \frac{4}{\sqrt{\pi}}\left(\frac{m}{2kT}\right)^{3/2}\gamma^2\exp\left(-\frac{m\gamma^2}{2kT}\right)$ cf. Theoretical Exercise 1, p. 237 of the text
- **Exercise:** Find E[**V**].
- **Example:** **X** and **Y** are random variables with joint CDF $F_{\mathbf{X},\mathbf{Y}}(u,v)$
- $\mathbf{Z} = \max(\mathbf{X},\mathbf{Y}) = \begin{cases} \mathbf{X}, & \text{if } \mathbf{X} > \mathbf{Y}, \\ \mathbf{Y} & \text{if } \mathbf{X} \le \mathbf{Y}. \end{cases}$
- $F_{\mathbf{Z}}(\alpha) = P\{\mathbf{Z} \le \alpha\} = P\{\max(\mathbf{X},\mathbf{Y}) \le \alpha\} = P\{\mathbf{X} \le \alpha, \mathbf{Y} \le \alpha\} = F_{\mathbf{X},\mathbf{Y}}(\alpha,\alpha)$
- If **X** and **Y** are independent, $F_{\mathbf{Z}}(\alpha) = F_{\mathbf{X},\mathbf{Y}}(\alpha,\alpha) = F_{\mathbf{X}}(\alpha)F_{\mathbf{Y}}(\alpha)$
- If **X** and **Y** are independent and jointly continuous, then
  $f_{\mathbf{Z}}(\alpha) = \frac{d}{d\alpha}F_{\mathbf{Z}}(\alpha) = \frac{d}{d\alpha}F_{\mathbf{X}}(\alpha)F_{\mathbf{Y}}(\alpha) = f_{\mathbf{X}}(\alpha)F_{\mathbf{Y}}(\alpha) + F_{\mathbf{X}}(\alpha)f_{\mathbf{Y}}(\alpha)$
- $P\{\alpha \le \mathbf{Z} \le \alpha+\delta\alpha\} = f_{\mathbf{Z}}(\alpha)\delta\alpha = P\{\alpha \le \mathbf{X} \le \alpha+\delta\alpha, \mathbf{Y} \le \alpha\} + P\{\mathbf{X} \le \alpha, \alpha \le \mathbf{Y} \le \alpha+\delta\alpha\}$
  $= f_{\mathbf{X}}(\alpha)\delta\alpha F_{\mathbf{Y}}(\alpha) + F_{\mathbf{X}}(\alpha)f_{\mathbf{Y}}(\alpha)\delta\alpha$
- If **X** and **Y** also have identical marginal CDF and pdf, then $F_{\mathbf{Z}}(\alpha) = F^2(\alpha)$; $f_{\mathbf{Z}}(\alpha) = 2f(\alpha)F(\alpha)$
- **Example:** If **X** and **Y** are independent and uniformly distributed on [0,1], then
  $f(\alpha) = \begin{cases} 1, & 0 \le \alpha \le 1, \\ 0, & \text{elsewhere.} \end{cases}$
- Hence, $f_{\mathbf{Z}}(\alpha) = 2f(\alpha)F(\alpha) = \begin{cases} 2\alpha, & 0 \le \alpha \le 1, \\ 0, & \text{elsewhere.} \end{cases}$
- **X** and **Y** are random variables with joint CDF $F_{\mathbf{X},\mathbf{Y}}(u,v)$
- $\mathbf{W} = \min(\mathbf{X},\mathbf{Y}) = \begin{cases} \mathbf{Y}, & \text{if } \mathbf{X} > \mathbf{Y}, \\ \mathbf{X} & \text{if } \mathbf{X} \le \mathbf{Y}. \end{cases}$
- $F_{\mathbf{W}}(\beta) = P\{\mathbf{W} \le \beta\} = P\{\min(\mathbf{X},\mathbf{Y}) \le \beta\} = P(\{\mathbf{X} \le \beta\} \cup \{\mathbf{Y} \le \beta\})$
- $F_{\mathbf{W}}(\alpha) = P\{\mathbf{W} \le \beta\} = P(\{\mathbf{X} \le \beta\} \cup \{\mathbf{Y} \le \beta\}) = P\{\mathbf{X} \le \beta\} + P\{\mathbf{Y} \le \beta\} - P\{\mathbf{X} \le \beta, \mathbf{Y} \le \beta\}$
  $= F_{\mathbf{X}}(\beta) + F_{\mathbf{Y}}(\beta) - F_{\mathbf{X},\mathbf{Y}}(\beta,\beta)$




- If **X** and **Y** are independent, then
  $F_{\mathbf{W}}(\beta) = F_{\mathbf{X}}(\beta) + F_{\mathbf{Y}}(\beta) - F_{\mathbf{X}}(\beta)F_{\mathbf{X}}(\beta) = 1 - [1 - F_{\mathbf{X}}(\beta)][1 - F_{\mathbf{Y}}(\beta)]$
- $F_{\mathbf{W}}(\beta) = 2F(\beta) - [F(\beta)]^2$ and $f_{\mathbf{W}}(\beta) = 2f(\beta)[1 - F(\beta)]$ if the marginal CDFs are the same

- **Example:** If $\mathbf{X}$ and $\mathbf{Y}$ are independent and uniformly distributed on [0,1],

  then $f(\alpha) = \begin{cases} 1, & 0 \le \alpha \le 1, \\ 0, & \text{elsewhere.} \end{cases}$

  $f_{\mathbf{W}}(\beta) = 2f(\beta)[1 - F(\beta)] = \begin{cases} 2(1-\beta), & 0 \le \beta \le 1, \\ 0, & \text{elsewhere.} \end{cases}$

- $F_{\mathbf{W}}(\beta) = F_{\mathbf{X}}(\beta) + F_{\mathbf{Y}}(\beta) - F_{\mathbf{X}}(\beta)F_{\mathbf{X}}(\beta) = 1 - [1 - F_{\mathbf{X}}(\beta)][1 - F_{\mathbf{Y}}(\beta)]$
- Note that $[1-F_{\mathbf{X}}(\beta)][1-F_{\mathbf{Y}}(\beta)] = P\{\mathbf{X} > \beta\}P\{\mathbf{Y} > \beta\} = P\{\mathbf{X} > \beta, \mathbf{Y} > \beta\}$
  $= P\{\min(\mathbf{X}, \mathbf{Y}) > \beta\} = 1 - F_{\mathbf{W}}(\beta)$
- **Generalization to maximum and minimum of many random variables**
- $\mathbf{Z} = \max(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$
- $\mathbf{W} = \min(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$
- $F_{\mathbf{Z}}(\alpha) = P\{\mathbf{Z} \le \alpha\} = P\{\cap\, \mathbf{X}_i \le \alpha\} = F_{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n}(\alpha, \alpha, \ldots, \alpha)$

  $= \prod_i F_{\mathbf{X}}(\alpha)$ if independent $= [F(\alpha)]^n$ if the marginal CDFs are also the same
- Thus, if $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ have identical marginal pdf f(u) and they are also independent random variables, then, $\mathbf{Z}$, the largest of these n random variables, has pdf $f_{\mathbf{Z}}(\alpha) = n[F(\alpha)]^{n-1}f(\alpha)$
- $\mathbf{W} = \min(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$
- $P\{\mathbf{W} > \beta\} = 1 - F_{\mathbf{W}}(\beta) = P\{\cap\, \mathbf{X}_i > \beta\}$
- If the $\mathbf{X}_i$'s are independent, $P\{\cap \mathbf{X}_i > \beta\} = \prod_i P\{\mathbf{X}_i > \beta\}$
- If the $\mathbf{X}_i$'s are independent, $1 - F_{\mathbf{W}}(\beta) = \prod_i [1 - F_{\mathbf{X}}(\beta)]$
- If the marginal CDFs are the same, then $1 - F_{\mathbf{W}}(\beta) = [1 - F(\beta)]^n$ and
  $f_{\mathbf{W}}(\beta) = n[1 - F(\beta)]^{n-1}f(\beta)$
- **Example:** $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_5$ are independent and uniformly distributed on [0,1]

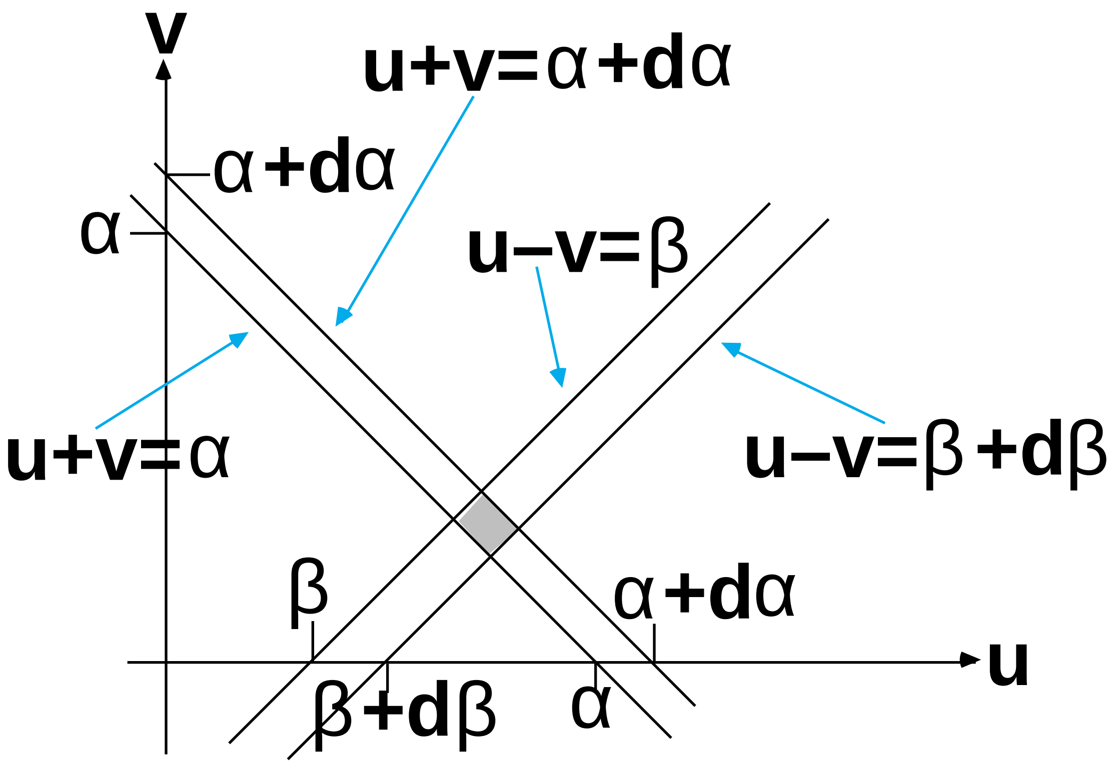
  $f(\beta) = 1$ for $0 < \beta < 1$ $F(\beta) = \beta$ for $0 < \beta < 1$
- $f_{\mathbf{W}}(\beta) = \begin{cases} 5(1-\beta)^4 & \text{if } 0 < \beta < 1, \\ 0, & \text{otherwise.} \end{cases}$
- **Example:** $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ are independent random variables representing the lifetimes of components
- The system fails as soon as a component fails
- System lifetime = $\mathbf{W} = \min(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$
- $1 - F_{\mathbf{W}}(\beta) = \prod_i [1 - F_{\mathbf{X}_i}(\beta)] = [1 - F_{\mathbf{X}_1}(\beta)][1 - F_{\mathbf{X}_2}(\beta)]\ldots[1 - F_{\mathbf{X}_n}(\beta)]$
- $f_{\mathbf{W}}(\beta) = \sum_i f_{\mathbf{X}_i}(\beta)\prod_{j\ne i}[1 - F_{\mathbf{X}_j}(\beta)]$
- What is the hazard rate of $\mathbf{W}$, the lifetime of the system?

- Hazard rate of $\mathbf{W} = h_{\mathbf{W}}(t) = \dfrac{f_{\mathbf{W}}(t)}{1 - F_{\mathbf{W}}(t)} = \displaystyle\sum_i \dfrac{f_{\mathbf{X}_i}(t)}{1 - F_{\mathbf{X}_i}(t)} = \sum_i h_{\mathbf{X}_i}(t)$
- Hazard rate of system lifetime is the *sum* of the hazard rates of the components, and thus is larger than the hazard rate of any one component
- The larger the hazard rate, the sooner the system fails
- Moral: systems with many components (i.e. complex systems) fail sooner than simple systems
- **Multiple functions of multiple random variables**
- We have studied methods for finding the CDF (or pdf or pmf) of a function of two (or more) random variables
- What is the joint CDF/pdf/pmf of $\mathbf{W} = g(\mathbf{X},\mathbf{Y})$ and $\mathbf{Z} = h(\mathbf{X},\mathbf{Y})$?
- The distributions of **W** and **Z** *separately* are insufficient to determine the *joint* distribution
- Suppose that **X** and **Y** are discrete $\mathbf{W} = g(\mathbf{X},\mathbf{Y})$ and $\mathbf{Z} = h(\mathbf{X},\mathbf{Y})$ are also discrete
- Determine the set of values taken on by **W** and **Z**

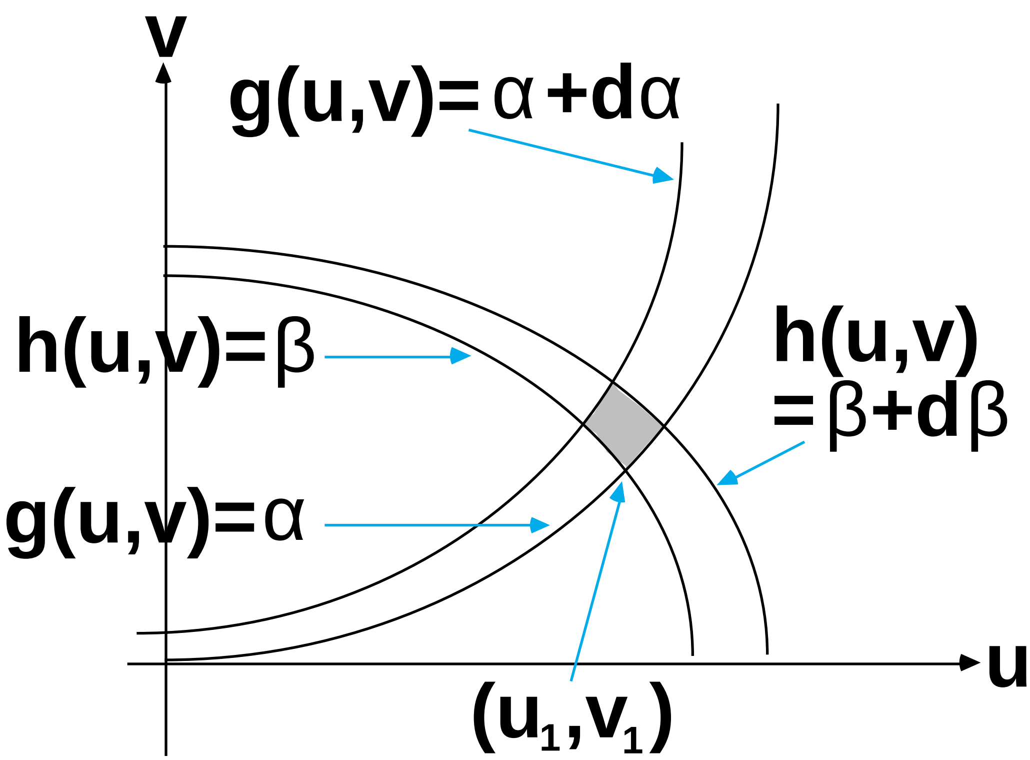
  $\{\alpha_k\} = \{g(u_i,v_j): 1 \le i \le n,\ 1 \le j \le m\}$ $\{\beta_l\} = \{h(u_i,v_j): 1 \le i \le n,\ 1 \le j \le m\}$
- $p_{\mathbf{W},\mathbf{Z}}(\alpha_k,\beta_l) = P\{\mathbf{W} = \alpha_k,\ \mathbf{Z} = \beta_l\} = \displaystyle\sum_i \sum_j p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)\ g(u_i,v_j) = \alpha_k,\ h(u_i,v_j) = \beta_l$
- We are just adding up the probabilities of all $(\mathbf{X},\mathbf{Y})$ such that $g(\mathbf{X},\mathbf{Y}) = \alpha_k$; $h(\mathbf{X},\mathbf{Y}) = \beta_l$
- **Example:** $\mathbf{W} = \mathbf{XY}$, $\mathbf{Z} = \mathbf{X}/\mathbf{Y}$
- **W** and **Z** are both positive or they are both negative
- $p_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = p_{\mathbf{X},\mathbf{Y}}(\sqrt{\alpha\beta},\sqrt{\alpha/\beta}) + p_{\mathbf{X},\mathbf{Y}}(-\sqrt{\alpha\beta},-\sqrt{\alpha/\beta})$ $\boxed{\alpha,\ \beta > 0}$
- $p_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = p_{\mathbf{X},\mathbf{Y}}(\sqrt{\alpha\beta},-\sqrt{\alpha/\beta}) + p_{\mathbf{X},\mathbf{Y}}(-\sqrt{\alpha\beta},\sqrt{\alpha/\beta})$ $\boxed{\alpha,\ \beta < 0}$
- **Example:** $\mathbf{W} = \min(\mathbf{X},\mathbf{Y})$, $\mathbf{Z} = \max(\mathbf{X},\mathbf{Y})$
- $p_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = \begin{cases} p_{\mathbf{X},\mathbf{Y}}(\alpha,\beta) + p_{\mathbf{X},\mathbf{Y}}(\beta,\alpha), & \alpha \le \beta, \\ 0, & \alpha > \beta, \end{cases}$
- All the probability mass lies on or above the line $\alpha = \beta$ in the plane with axes $\alpha$ and $\beta$
- Suppose that $(\mathbf{X},\mathbf{Y})$ is jointly continuous and so is $(\mathbf{W},\mathbf{Z}) = (g(\mathbf{X},\mathbf{Y}),h(\mathbf{X},\mathbf{Y}))$
- Note that even if g and h are continuous, $(\mathbf{W},\mathbf{Z})$ need not be jointly continuous
- **Example:** $\mathbf{W} = \cos(\mathbf{X}+\mathbf{Y})$ and $\mathbf{Z} = \sin(\mathbf{X}+\mathbf{Y})$ are not jointly continuous (why not?)
- Basic idea: we compute the joint pdf $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta)$ directly from $f_{\mathbf{X},\mathbf{Y}}(u,v)$
- $P\{(\mathbf{X},\mathbf{Y})\quad B\} \approx f_{\mathbf{X},\mathbf{Y}}(u,v) \times$ Area of region B
- Approximation is good if B has small area, and poor if the area of B is large
- $P\{u \le \mathbf{X} \le u+\delta u,\ v \le \mathbf{Y} \le v+\delta v\} = P\{(\mathbf{X},\mathbf{Y})$ rectangle of area

  $\delta u \times \delta v\} \approx f_{\mathbf{X},\mathbf{Y}}(u,v) \times \delta u \times \delta v$
- $P\{\alpha \le \mathbf{W} \le \alpha+d\alpha,\ \beta \le \mathbf{Z} \le \beta+d\beta\} \approx f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta)d\alpha d\beta \approx f_{\mathbf{X},\mathbf{Y}}(u,v) \times \text{Area(B)}$ for some B
- **Example:** $\mathbf{W} = \mathbf{X}+\mathbf{Y}$, $\mathbf{Z} = \mathbf{X}-\mathbf{Y}$

- $\mathbf{W} = \alpha$ if $(\mathbf{X},\mathbf{Y})$ lies on $u+v = \alpha$ $\mathbf{Z} = \beta$ if $(\mathbf{X},\mathbf{Y})$ lies on $u-v = \beta$
- $(\mathbf{W},\mathbf{Z}) = (\alpha,\beta)$ iff $(\mathbf{X},\mathbf{Y}) = \left(\frac{\alpha+\beta}{2},\frac{\alpha-\beta}{2}\right)$
- $\{\alpha \le \mathbf{W} \le \alpha+d\alpha,\ \beta \le \mathbf{Z} \le \beta+d\beta\}$ iff $(\mathbf{X},\mathbf{Y})$ shaded rectangle shown

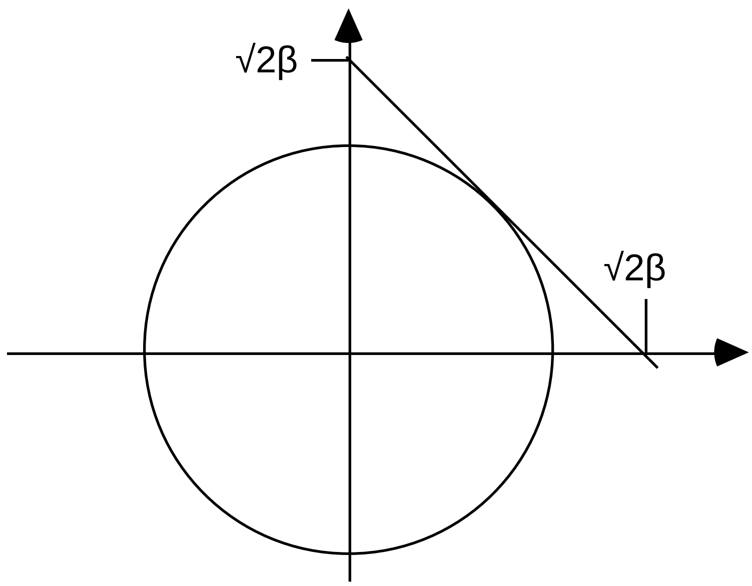



- $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta)d\alpha d\beta \approx P\{\alpha \le \mathbf{W} \le \alpha+d\alpha,\ \beta \le \mathbf{Z} \le \beta+d\beta\} = P\{(\mathbf{X},\mathbf{Y})$ rectangle$\}$
$\approx f_{\mathbf{X},\mathbf{Y}}\left(\frac{\alpha+\beta}{2},\frac{\alpha-\beta}{2}\right)\times$Area(rectangle); where Area(rectangle) $= \frac{d\alpha}{\sqrt{2}}\times\frac{d\beta}{\sqrt{2}} = \frac{d\alpha\, d\beta}{2}$
- $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = \frac{1}{2}f_{\mathbf{X},\mathbf{Y}}\left(\frac{\alpha+\beta}{2},\frac{\alpha-\beta}{2}\right)$
- More generally, $g(u,v) = \alpha$ and $h(u,v) = \beta$ are curves in the u-v plane
- Let $g(u,v) = \alpha$ and $h(u,v) = \beta$ intersect at the point $(u_1,v_1)$
- Thus, $(\mathbf{W},\mathbf{Z}) = (\alpha,\beta)$ whenever $(\mathbf{X},\mathbf{Y}) = (u_1,v_1)$
- More generally, the *rectangle* $\{\alpha \le \mathbf{W} \le \alpha+d\alpha,\ \beta \le \mathbf{Z} \le \beta+d\beta\}$ in the $\alpha$-$\beta$ plane corresponds to a region in the u-v plane bounded by the four *curves* $g(u,v) = \alpha$, $g(u,v) = \alpha+d\alpha$, and $h(u,v) = \beta$, $h(u,v) = \beta+d\beta$
- One corner is at $(u_1,v_1)$




- $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta)d\alpha d\beta \approx P\{\alpha \le \mathbf{W} \le \alpha+d\alpha,\ \beta \le \mathbf{Z} \le \beta+d\beta\} = P\{(\mathbf{X},\mathbf{Y})$ shaded region$\}$
  $\approx f_{\mathbf{X},\mathbf{Y}}(u_1,v_1)\times\text{Area(region)}$
- The area of the region is given by the *Jacobian* of the transformation mapping the u-v plane into the α-β plane: $J(u,v) = \begin{vmatrix} \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \\ \frac{\partial h}{\partial u} & \frac{\partial h}{\partial v} \end{vmatrix} = \frac{\partial g}{\partial u}\frac{\partial h}{\partial v} - \frac{\partial g}{\partial v}\frac{\partial h}{\partial u}$ Area of shaded region $\approx \dfrac{d\alpha d\beta}{|J(u_1,v_1)|}$
- $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = \dfrac{f_{\mathbf{X},\mathbf{Y}}(u_1,v_1)}{|J(u_1,v_1)|}$
- **Example: $\mathbf{W} = \mathbf{X}+\mathbf{Y},\ \mathbf{Z} = \mathbf{X}-\mathbf{Y}$**
- $g(u,v) = u + v$, $h(u, v) = u - v$ $u + v = \alpha$ and $u - v = \beta$ intersect at $(u_1,v_1) = \left(\frac{\alpha+\beta}{2}, \frac{\alpha-\beta}{2}\right)$
- $J(u,v) = \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix} = -2$ for all u and v
- $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = \frac{1}{2} f_{\mathbf{X},\mathbf{Y}}\left(\frac{\alpha+\beta}{2}, \frac{\alpha-\beta}{2}\right)$

- **Caveat:** It is possible that for some values of $\alpha$ and $\beta$, the curves $g(u,v) = \alpha$ and $h(u,v) = \beta$ do not intersect at all
- **Example:** $u + v = 10$ and $u^2 + v^2 = 1$ do not intersect
- $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = 0$ for such $\alpha$ and $\beta$
- **Caveat:** It is possible that for some values of $\alpha$ and $\beta$, the curves $g(u,v) = \alpha$ and $h(u,v) = \beta$ intersect at $(u_1,v_1)$, $(u_2,v_2)$, … $(u_n,v_n)$
- $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = \sum_{i=1}^{n} \frac{f_{\mathbf{X},\mathbf{Y}}(u_i,v_i)}{|J(u_i,v_i)|}$
- **Example:** $\mathbf{W} = \mathbf{X}+\mathbf{Y}$, $\mathbf{Z} = \mathbf{X}^2+\mathbf{Y}^2$
- $u + v = \alpha$ and $u^2 + v^2 = \beta$ do not intersect if $\alpha > \sqrt{2\beta}$

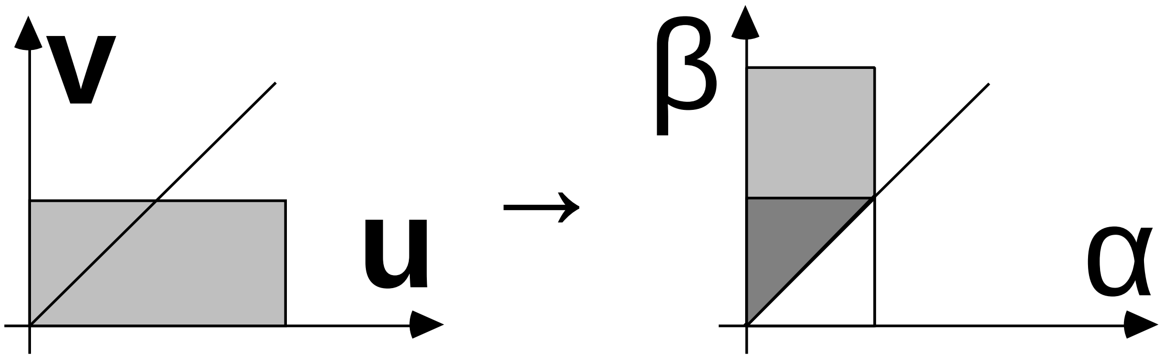



- If $|\alpha| < \sqrt{2\beta}$, there are two intersections at $\left(\frac{\alpha}{2} + \frac{\sqrt{2\beta-\alpha^2}}{2}, \frac{\alpha}{2} - \frac{\sqrt{2\beta-\alpha^2}}{2}\right)$ and $\left(\frac{\alpha}{2} - \frac{\sqrt{2\beta-\alpha^2}}{2}, \frac{\alpha}{2} + \frac{\sqrt{2\beta-\alpha^2}}{2}\right)$
- $J(u,v) = \begin{vmatrix} 1 & 1 \\ 2u & 2v \end{vmatrix} = 2(v-u) = \pm 2\sqrt{2\beta-\alpha^2}$
- **Example:** If $\mathbf{X}$ and $\mathbf{Y}$ are independent unit Gaussian random variables, then $f_{\mathbf{X},\mathbf{Y}}(u,v) = \frac{1}{2\pi}\exp(-(u^2+v^2)/2)$
- $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = \begin{cases} \frac{\exp(-\beta/2)}{2\pi\sqrt{2\beta-\alpha^2}}, & \text{if } |\alpha| < \sqrt{2\beta}, \\ 0 & \text{otherwise.} \end{cases}$
- **Exercise**: Sketch the region where the pdf is nonzero

- **Example:** If **X** and **Y** are independent exponential random variables with parameter 1, then $f_{\mathbf{X},\mathbf{Y}}(u,v) = \exp(-(u+v))$, $u,v > 0$

$$f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = \begin{cases} \dfrac{\exp(-\alpha)}{\sqrt{2\beta-\alpha^2}}, & \text{if } \sqrt{\beta} < \alpha < \sqrt{2\beta}, \\ 0 & \text{otherwise.} \end{cases}$$

- **Exercise:** Why must $\alpha > \sqrt{\beta}$ ??
- Generalization of the use of Jacobians to n functions of n random variables is possible. See the text for details
- Jacobians can be used only when g and h are differentiable functions
- Otherwise, we must work from first principles
- **Example:** $\mathbf{W} = \min(\mathbf{X},\mathbf{Y})$, $\mathbf{Z} = \max(\mathbf{X},\mathbf{Y})$ The minimum function and the maximum function are non-differentiable
- $P\{\alpha \le \mathbf{W} \le \alpha+d\alpha,\ \beta \le \mathbf{Z} \le \beta+d\beta\} = 0$ if $\alpha > \beta$. Why?
- For $\alpha \le \beta$, $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta)d\alpha d\beta = P\{\alpha \le \mathbf{W} \le \alpha+d\alpha,\ \beta \le \mathbf{Z} \le \beta+d\beta\}$
  $= P\{\alpha \le \mathbf{X} \le \alpha+d\alpha,\ \beta \le \mathbf{Y} \le \beta+d\beta\} + P\{\beta \le \mathbf{X} \le \beta+d\beta,\ \alpha \le \mathbf{Y} \le \alpha+d\alpha\}$
  $= f_{\mathbf{X},\mathbf{Y}}(\alpha,\beta)d\alpha d\beta + f_{\mathbf{X},\mathbf{Y}}(\beta,\alpha)d\alpha d\beta$
- $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = f_{\mathbf{X},\mathbf{Y}}(\alpha,\beta) + f_{\mathbf{X},\mathbf{Y}}(\beta,\alpha)$
- If $\alpha \le \beta$, $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = f_{\mathbf{X},\mathbf{Y}}(\alpha,\beta) + f_{\mathbf{X},\mathbf{Y}}(\beta,\alpha)$
  If $\alpha > \beta$, $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta) = 0$
- The joint pdf is nonzero only in the region $\alpha \le \beta$
- $(\alpha,\beta)$ and $(\beta,\alpha)$ are mirror image points about the line $\alpha = \beta$
- The joint pdf of the min and the max can be found by "folding the pdf" over




- Example: If $(\mathbf{X},\mathbf{Y})$ are uniform on unit square, min and max are uniform on $0 < \alpha < \beta < 1$

- **Expectation**
- $E[\mathbf{X}]$ = the expectation of $\mathbf{X}$ $E[\mathbf{X}] = \int_{-\infty}^{\infty} u f_{\mathbf{X}}(u) du$ for a continuous random variable

  $E[\mathbf{X}] = \sum_i u_i p_{\mathbf{X}}(u_i)$ for a discrete random variable
- $f_{\mathbf{X}}(u) = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v) dv$
- $E[\mathbf{X}] = \int_{-\infty}^{\infty} u f_{\mathbf{X}}(u) du = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u f_{\mathbf{X},\mathbf{Y}}(u,v) dv du$
- $E[\mathbf{Y}] = \int_{-\infty}^{\infty} v f_{\mathbf{Y}}(v) dv = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} v f_{\mathbf{X},\mathbf{Y}}(u,v) du dv$
- Similar formulas hold for discrete random variables with sums replacing integrals
- $E[\mathbf{X}] = \sum_i u_i p_{\mathbf{X}}(u_i) = \sum_i u_i \sum_j p_{\mathbf{X},\mathbf{Y}}(u_i,v_j) = \sum_i \sum_j u_i p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$
- LOTUS: $E[g(\mathbf{X})] = \sum_i g(u_i) p_{\mathbf{X}}(u_i) = \sum_i \sum_j g(u_i) p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$

  $E[h(\mathbf{Y})] = \sum_i \sum_j h(v_j) p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$
- $E[g(\mathbf{X})] = \int_{-\infty}^{\infty} g(u) f_{\mathbf{X}}(u) du = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u) f_{\mathbf{X},\mathbf{Y}}(u,v) dv du$
- $E[h(\mathbf{Y})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(v) f_{\mathbf{X},\mathbf{Y}}(u,v) du dv$
- A more general version of LOTUS also holds (GLOTUS?)
- Let $\mathbf{Z} = g(\mathbf{X},\mathbf{Y})$
- $E[\mathbf{Z}] = E[g(\mathbf{X},\mathbf{Y})] = \int_{-\infty}^{\infty} \alpha f_{\mathbf{Z}}(\alpha) d\alpha$
- To find $E[g(\mathbf{X},\mathbf{Y})]$, it is *not* necessary to first find the pdf (or pmf) of $g(\mathbf{X},\mathbf{Y})$: GLOTUS!
- GLOTUS: It is easy to find $E[\mathbf{Z}]$ as $E[\mathbf{Z}] = E[g(\mathbf{X},\mathbf{Y})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(u,v) f_{\mathbf{X},\mathbf{Y}}(u,v) du dv$

  or as $\sum_i \sum_j g(u_i,v_j) p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)$

- $E[g(\mathbf{X})h(\mathbf{Y})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u)h(v)f_{\mathbf{X},\mathbf{Y}}(u,v)dudv$
- Special case:

  $g(\mathbf{X},\mathbf{Y}) = \mathbf{X} + \mathbf{Y}$ $E[g(\mathbf{X},\mathbf{Y})] = E[\mathbf{X} + \mathbf{Y}] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (u+v)f_{\mathbf{X},\mathbf{Y}}(u,v)dudv = E[\mathbf{X}] + E[\mathbf{Y}]$
- More generally, $E[a\mathbf{X} + b\mathbf{Y}] = aE[\mathbf{X}] + bE[\mathbf{Y}]$
- GGLOTUS $E[g(\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n)] = E[g(\underline{\mathbf{X}})] = \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \underline{g(\underline{u})}f_{\underline{\mathbf{X}}}\underline{(\underline{u})}d\underline{u}$

  $= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} g(u_1,u_2,\ldots,u_n)\, f_{\mathbf{X}_1\mathbf{X}_2\ldots\mathbf{X}_n}(u_1,u_2,\ldots,u_n)du_1du_2\ldots du_n$
- $E[a_1\mathbf{X}_1 + a_2\mathbf{X}_2 + \ldots + a_n\mathbf{X}_n] = a_1E[\mathbf{X}_1] + a_2E[\mathbf{X}_2] + \ldots + a_nE[\mathbf{X}_n]$
- The average of a sum is the sum of the averages
- This statement is true for *arbitrary* random variables
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent random variables, then $f_{\mathbf{X},\mathbf{Y}}(u,v) = f_{\mathbf{X}}(u)f_{\mathbf{Y}}(v)$ and hence $E[g(\mathbf{X})h(\mathbf{Y})] = E[g(\mathbf{X})]E[h(\mathbf{Y})]$
- The *covariance* of two random variables generalizes the notion of variance
- $\text{cov}(\mathbf{X},\mathbf{Y}) = E[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})]$
- $\text{cov}(\mathbf{X},\mathbf{X}) = E[(\mathbf{X} - \mu_{\mathbf{X}})^2] = \text{var}(\mathbf{X}) = E[\mathbf{X}^2] - (\mu_{\mathbf{X}})^2 = E[\mathbf{X}^2] - (E[\mathbf{X}])^2$
- $\text{cov}(\mathbf{X},\mathbf{Y}) = E[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})] = E[\mathbf{XY}] - \mu_{\mathbf{X}}\mu_{\mathbf{Y}} = E[\mathbf{XY}] - E[\mathbf{X}]E[\mathbf{Y}]$
- If $\text{cov}(\mathbf{X},\mathbf{Y}) = 0$, then $\mathbf{X}$ and $\mathbf{Y}$ are said to be *uncorrelated* random variables
- $E[g(\mathbf{X})h(\mathbf{Y})] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u)h(v)f_{\mathbf{X},\mathbf{Y}}(u,v)dudv$
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent random variables, then $f_{\mathbf{X},\mathbf{Y}}(u,v) = f_{\mathbf{X}}(u)f_{\mathbf{Y}}(v)$ and $E[g(\mathbf{X})h(\mathbf{Y})] = E[g(\mathbf{X})]E[h(\mathbf{Y})]$
- For independent random variables $\mathbf{X}$ and $\mathbf{Y}$, $E[\mathbf{XY}] = E[\mathbf{X}]E[\mathbf{Y}] = \mu_{\mathbf{X}}\mu_{\mathbf{Y}}$
- $\text{var}(\mathbf{X} + \mathbf{Y}) = E[(\mathbf{X} + \mathbf{Y})^2] - (E[\mathbf{X} + \mathbf{Y}])^2$

  $= E[\mathbf{X}^2] + E[\mathbf{Y}^2] + 2\mu_{\mathbf{X}}\mu_{\mathbf{Y}} - (\mu_{\mathbf{X}})^2 - (\mu_{\mathbf{Y}})^2 - 2\mu_{\mathbf{X}}\mu_{\mathbf{Y}}$
- For *independent* random variables $\mathbf{X}$ and $\mathbf{Y}$, $\text{var}(\mathbf{X} + \mathbf{Y}) = \text{var}(\mathbf{X}) + \text{var}(\mathbf{Y})$
- $E[\mathbf{X} + \mathbf{Y}] = E[\mathbf{X}] + E[\mathbf{Y}]$ for *arbitrary* random variables
- **Exercise:** Show that $\text{var}(\mathbf{X} - \mathbf{Y}) = \text{var}(\mathbf{X}) + \text{var}(\mathbf{Y})$ for independent RVs
- What is $\text{var}(\mathbf{X} + \mathbf{Y})$ for arbitrary random variables?
- We need to study the concept of the covariance of two random variables in order to find $\text{var}(\mathbf{X} + \mathbf{Y})$ or $\text{var}(\mathbf{X} - \mathbf{Y})$
- $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated if $E[\mathbf{XY}] - \mu_{\mathbf{X}}\mu_{\mathbf{Y}} = 0$, i.e. if $E[\mathbf{XY}] = \mu_{\mathbf{X}}\mu_{\mathbf{Y}} = E[\mathbf{X}]E[\mathbf{Y}]$
- If $\mathbf{X}$ and $\mathbf{Y}$ are independent random variables, then $E[g(\mathbf{X})h(\mathbf{Y})] = E[g(\mathbf{X})]E[h(\mathbf{Y})]$
- $\text{cov}(\mathbf{X},\mathbf{Y}) = E[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})] = E[\mathbf{X} - \mu_{\mathbf{X}}]E[\mathbf{Y} - \mu_{\mathbf{Y}}] = 0\times 0 = 0$

- Independent random variables are uncorrelated
- Independent random variables are uncorrelated
- However, uncorrelated random variables need not be independent
- Covariance = 0 a certain integral or sum equals 0
- Independence joint CDF = product of marginal CDFs
- Independence asserts that a certain property holds at every point in the plane
- Covariance is an average over the entire plane
- Covariance = 0 implies that the distribution is "balanced" in some way but says nothing about the distribution values
- **Example:** $(\mathbf{X},\mathbf{Y})$ is uniformly distributed on the unit circle
- $E[\mathbf{X}] = \int\int_{u^2+v^2 \le 1} u\frac{1}{\pi}dudv = \int_{r=0}^{1}\int_{\theta=0}^{2\pi} \frac{1}{\pi} r\cos\theta \, r d\theta \, dr = 0$
- Similarly, $E[\mathbf{Y}] = 0$
- $E[\mathbf{X}\mathbf{Y}] = \int\int_{u^2+v^2 \le 1} \frac{1}{\pi} uv \; dudv = \int_{r=0}^{1}\int_{\theta=0}^{2\pi} \frac{1}{\pi} r\cos\theta \, r\sin\theta \, r \, d\theta \, dr = 0$ $\mathbf{X}$ and $\mathbf{Y}$ uncorrelated
- But, eyeball test shows that $\mathbf{X}$ and $\mathbf{Y}$ cannot be independent
- Normalized version of the covariance is called the *correlation coefficient* $\rho$ or $\rho_{\mathbf{X},\mathbf{Y}}$ or $\rho(\mathbf{X},\mathbf{Y}) = \frac{\mathrm{cov}(\mathbf{X},\mathbf{Y})}{\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}} = \frac{\mathrm{cov}(\mathbf{X},\mathbf{Y})}{\sqrt{\mathrm{var}(\mathbf{X})\mathrm{var}(\mathbf{Y})}}$
- $-1 \le \rho \le 1$
- If $\rho = 1$, $\mathbf{X}$ and $\mathbf{Y}$ are said to be perfectly (positively) correlated
- If $\rho = -1$, $\mathbf{X}$ and $\mathbf{Y}$ are said to be perfectly (negatively) correlated
- If $\rho = 0$, $\mathbf{X}$ and $\mathbf{Y}$ are said to be uncorrelated
- What is $\mathrm{var}(\mathbf{X} + \mathbf{Y})$ for arbitrary random variables?
- $E[\mathbf{X} + \mathbf{Y}] = E[\mathbf{X}]+E[\mathbf{Y}] = \mu_{\mathbf{X}}+\mu_{\mathbf{Y}}$
- $\mathrm{var}(\mathbf{X} + \mathbf{Y}) = E[(\mathbf{X} + \mathbf{Y})^2] - (E[\mathbf{X} + \mathbf{Y}])^2$
  $= E[\mathbf{X}^2] + E[\mathbf{Y}^2] + 2E[\mathbf{X}\mathbf{Y}] - (\mu_{\mathbf{X}})^2 - (\mu_{\mathbf{Y}})^2 - 2\mu_{\mathbf{X}}\mu_{\mathbf{Y}}$
- $\mathrm{var}(\mathbf{X} + \mathbf{Y}) = E[\mathbf{X}^2] + E[\mathbf{Y}^2] + 2E[\mathbf{X}\mathbf{Y}] - (\mu_{\mathbf{X}})^2 - (\mu_{\mathbf{Y}})^2 - 2\mu_{\mathbf{X}}\mu_{\mathbf{Y}}$
  $= \mathrm{var}(\mathbf{X}) + \mathrm{var}(\mathbf{Y}) + 2\,\mathrm{cov}(\mathbf{X},\mathbf{Y})$
- $\mathrm{var}(\mathbf{X} - \mathbf{Y}) = \mathrm{var}(\mathbf{X}) + \mathrm{var}(\mathbf{Y}) - 2\,\mathrm{cov}(\mathbf{X},\mathbf{Y})$
- Special case: If $\mathbf{X}$ and $\mathbf{Y}$ are uncorrelated random variables, then $\mathrm{cov}(\mathbf{X},\mathbf{Y}) = 0$
- $\mathrm{var}(\mathbf{X} \pm \mathbf{Y}) = \mathrm{var}(\mathbf{X}) + \mathrm{var}(\mathbf{Y})$ for uncorrelated random variables also, not just for independent random variables. Independent uncorrelated
- More generally, $\mathrm{var}(a\mathbf{X} + b\mathbf{Y}) = E[(a\mathbf{X}+b\mathbf{Y})^2] - (E[a\mathbf{X}+b\mathbf{Y}])^2$
  $= E[a^2\mathbf{X}^2]+E[b^2\mathbf{Y}^2]+2E[ab\mathbf{X}\mathbf{Y}] - (a\mu_{\mathbf{X}})^2 - (b\mu_{\mathbf{Y}})^2 - 2ab\mu_{\mathbf{X}}\mu_{\mathbf{Y}}$
  $= a^2\mathrm{var}(\mathbf{X}) + b^2\mathrm{var}(\mathbf{Y}) + 2ab\,\mathrm{cov}(\mathbf{X},\mathbf{Y})$

- $\text{var}(a\mathbf{X} + b\mathbf{Y}) = a^2(\sigma_{\mathbf{X}})^2 + b^2(\sigma_{\mathbf{Y}})^2 + 2ab\rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}$
- Set $a = 1/\sigma_{\mathbf{X}}$, $b = \pm 1/\sigma_{\mathbf{Y}}$ $\text{var}\left(\frac{\mathbf{X}}{\sigma_{\mathbf{X}}} \pm \frac{\mathbf{Y}}{\sigma_{\mathbf{Y}}}\right) = 2(1 \pm \rho)$
- Variance $\geq 0$ and hence $1 + \rho \geq 0$; $1 - \rho \geq 0$, i.e. $|\rho| \leq 1$
- $\text{cov}(a\mathbf{X} + b\mathbf{Y}, c\mathbf{X} + d\mathbf{Y}) = E[(a\mathbf{X} + b\mathbf{Y})(c\mathbf{X} + d\mathbf{Y})] - (a\mu_{\mathbf{X}} + b\mu_{\mathbf{Y}})(c\mu_{\mathbf{X}} + d\mu_{\mathbf{Y}})$
  $= ac(E[\mathbf{X}^2] - (\mu_{\mathbf{X}})^2) + bd(E[\mathbf{Y}^2] - (\mu_{\mathbf{Y}})^2) + (ad + bc)(E[\mathbf{XY}] - \mu_{\mathbf{X}}\mu_{\mathbf{Y}})$
  $= ac\ \text{var}(\mathbf{X}) + bd\ \text{var}(\mathbf{Y}) + (ad + bc)\ \text{cov}(\mathbf{X},\mathbf{Y})$
- Since $\text{var}(\mathbf{X}) = \text{cov}(\mathbf{X},\mathbf{X})$ and $\text{var}(\mathbf{Y}) = \text{cov}(\mathbf{Y},\mathbf{Y})$ $\text{cov}(a\mathbf{X} + b\mathbf{Y}, c\mathbf{X} + d\mathbf{Y})$
  $= ac\ \text{var}(\mathbf{X}) + bd\ \text{var}(\mathbf{Y}) + (ad + bc)\ \text{cov}(\mathbf{X},\mathbf{Y})$
  $= ac\ \text{cov}(\mathbf{X},\mathbf{X}) + bd\ \text{cov}(\mathbf{Y},\mathbf{Y}) + (ad + bc)\ \text{cov}(\mathbf{X},\mathbf{Y})$
- **Conditional density functions**
- *Conditional* pmf of $\mathbf{Y}$ *given* $\mathbf{X} = u_i$ is $p_{\mathbf{Y}|\mathbf{X}}(v_j|u_i) = P\{\mathbf{Y} = v_j|\mathbf{X} = u_i\}$
- $p_{\mathbf{Y}|\mathbf{X}}(v_j|u_i) = P\{\mathbf{Y} = v_j|\mathbf{X} = u_i\} = \dfrac{P\{\mathbf{X} = u_i, \mathbf{Y} = v_j\}}{P\{\mathbf{X} = u_i\}} = \dfrac{p_{\mathbf{X},\mathbf{Y}}(u_i,v_j)}{p_{\mathbf{X}}(u_i)}$
- A conditional pmf is a valid pmf just like any other pmf
- If $(\mathbf{X},\mathbf{Y})$ is jointly continuous, we cannot condition on $\mathbf{X} = u$ because this is an event of probability 0 ?

  $P\{u \leq \mathbf{X} \leq u + \delta u, v \leq \mathbf{Y} \leq v + \delta v\} = P\{(\mathbf{X},\mathbf{Y})$ rectangle of area $\delta u \times \delta v\}$

  $\approx f_{\mathbf{X},\mathbf{Y}}(u,v) \times \delta u \times \delta v$

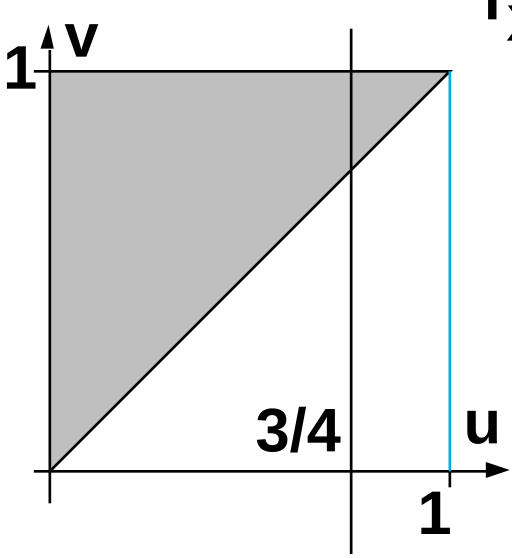
$$P\{v \leq \mathbf{Y} \leq v + \delta v | u \leq \mathbf{X} \leq u + \delta u\} = \frac{f_{\mathbf{X},\mathbf{Y}}(u,v) \times \delta u \times \delta v}{P\{u \leq \mathbf{X} \leq u + \delta u\}} = \frac{f_{\mathbf{X},\mathbf{Y}}(u,v) \times \delta u \times \delta v}{f_{\mathbf{X}}(u) \times \delta u} = \frac{f_{\mathbf{X},\mathbf{Y}}(u,v) \times \delta v}{f_{\mathbf{X}}(u)}$$

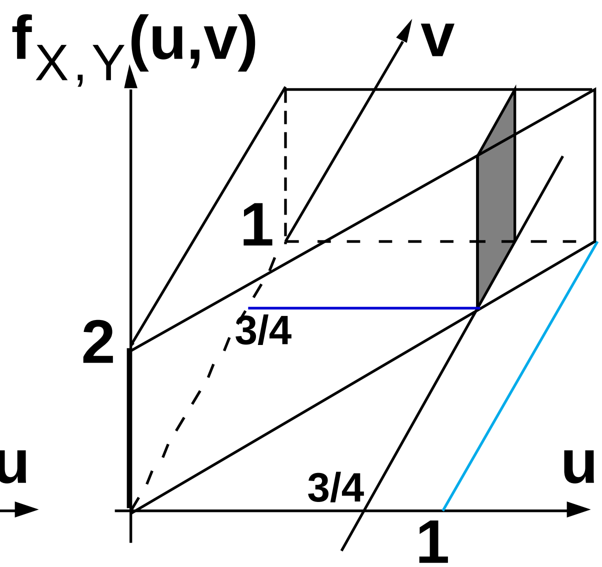
  which does not depend on $\delta u$ at all!

- If u is such that $f_{\mathbf{X}}(u) > 0$, the conditional pdf of $\mathbf{Y}$ given that $\mathbf{X} = u$ is $f_{\mathbf{Y}|\mathbf{X}}(v|u) = \dfrac{f_{\mathbf{X},\mathbf{Y}}(u,v)}{f_{\mathbf{X}}(u)}$

  $P\{v \leq \mathbf{Y} \leq v + \delta v | \mathbf{X} = u\} = f_{\mathbf{Y}|\mathbf{X}}(v|u) \times \delta v$

- In $f_{\mathbf{Y}|\mathbf{X}}(v|u)$, the variable is v. u is just the value of $\mathbf{X}$. It is some number, say 3/4
- The *shape* of the conditional pdf $f_{\mathbf{Y}|\mathbf{X}}(v|3/4)$ is the *shape* of the cross-section $f_{\mathbf{X},\mathbf{Y}}(3/4,v)$ of the joint pdf surface




- Cross-section $f_{\mathbf{X},\mathbf{Y}}(3/4,v)$ of the joint pdf is a rectangle
- The cross-section $f_{\mathbf{X},\mathbf{Y}}(3/4,v)$ is not necessarily a valid pdf because its area may not be 1
- If area of cross-section = A, $f_{\mathbf{X},\mathbf{Y}}(3/4,v)/A$ is a valid pdf
- $A = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(3/4,v)dv = f_{\mathbf{X}}(3/4)$
- $A = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(3/4,v)dv = f_{\mathbf{X}}(3/4)$ and we defined $f_{\mathbf{Y}|\mathbf{X}}(v|3/4) = \dfrac{f_{\mathbf{X},\mathbf{Y}}(3/4,v)}{f_{\mathbf{X}}(3/4)}$
- $f_{\mathbf{X}}(u)$ in the denominator is a normalizing factor
- **Example:** $f_{\mathbf{X},\mathbf{Y}}(u,v) = \begin{cases} 2, & 0 < u < v < 1, \\ 0, & \text{elsewhere.} \end{cases}$
- $f_{\mathbf{X}}(u) = 2(1-u),\ \ 0 < u < 1.$
- $f_{\mathbf{Y}|\mathbf{X}}(v|u) = \dfrac{f_{\mathbf{X},\mathbf{Y}}(u,v)}{f_{\mathbf{X}}(u)} = \dfrac{1}{1-u}$ for $u < v < 1$
- $f_{\mathbf{Y}|\mathbf{X}}(v|u) = \dfrac{1}{1-u}$ for $u < v < 1$
- This is a *uniform pdf* on (u,1) Note that the variable is v, and not u
- Given **X** = u, the conditional pdf of **Y** is uniform on (u,1)
- Conditional pdfs are valid pdfs
- $f_{\mathbf{Y}|\mathbf{X}}(v|u) = \dfrac{f_{\mathbf{X},\mathbf{Y}}(u,v)}{f_{\mathbf{X}}(u)}$
- If **X** and **Y** are independent random variables, then $f_{\mathbf{X},\mathbf{Y}}(u,v) = f_{\mathbf{X}}(u)f_{\mathbf{Y}}(v)$ and hence $f_{\mathbf{Y}|\mathbf{X}}(v|u) = f_{\mathbf{Y}}(v)$
- This is intuitively satisfying
- **Theorem of total probability:** $f_{\mathbf{Y}}(v) = \int_{-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,v)du = \int_{-\infty}^{\infty} f_{\mathbf{Y}|\mathbf{X}}(v|u)f_{\mathbf{X}}(u)du$

- Let T denote some event. Then, $P(T) = \int_{-\infty}^{\infty} P(T | \mathbf{X} = u) f_{\mathbf{X}}(u)du$
- **Example:** $P\{\mathbf{Y} \le \mathbf{X}\} = ?$
- $P\{\mathbf{Y} \le \mathbf{X}\} = \int_{-\infty}^{\infty} P\{\mathbf{Y} \le \mathbf{X} | \mathbf{X} = u\} f_{\mathbf{X}}(u)du = \int_{-\infty}^{\infty} P\{\mathbf{Y} \le u | \mathbf{X} = u\} f_{\mathbf{X}}(u)du$

  $= \int_{-\infty}^{\infty} \int_{-\infty}^{u} f_{\mathbf{Y}|\mathbf{X}}(v|u)\, dv\, f_{\mathbf{X}}(u)du$
- $P\{\mathbf{Y} \le \mathbf{X}\} = \int_{-\infty}^{\infty} \int_{-\infty}^{u} f_{\mathbf{Y}|\mathbf{X}}(v|u)\, dv\, f_{\mathbf{X}}(u)du = \int_{-\infty}^{\infty} \int_{-\infty}^{u} f_{\mathbf{X},\mathbf{Y}}(u,v)\, dv\, du$
- The *covariance* of two random variables generalizes the notion of variance
- $\text{cov}(\mathbf{X},\mathbf{Y}) = E[(\mathbf{X} - \mu_{\mathbf{X}})(\mathbf{Y} - \mu_{\mathbf{Y}})] = E[\mathbf{X}\mathbf{Y}] - \mu_{\mathbf{X}}\mu_{\mathbf{Y}}$
- $\text{cov}(\mathbf{X},\mathbf{X}) = E[(\mathbf{X} - \mu_{\mathbf{X}})^2] = \text{var}(\mathbf{X}) = E[\mathbf{X}^2] - (E[\mathbf{X}])^2$
- $\text{cov}(a\mathbf{X} + b\mathbf{Y}, c\mathbf{X} + d\mathbf{Y}) = ac\,\text{cov}(\mathbf{X},\mathbf{X}) + bd\,\text{cov}(\mathbf{Y},\mathbf{Y}) + (ad + bc)\,\text{cov}(\mathbf{X},\mathbf{Y})$
  $= ac\,\text{var}(\mathbf{X}) + bd\,\text{var}(\mathbf{Y}) + (ad + bc)\,\text{cov}(\mathbf{X},\mathbf{Y})$
- For n random variables, $\text{cov}\left(\sum_i a_i\mathbf{X}_i \;\; \sum_j b_j\mathbf{X}_j\right) = \sum_i \sum_j a_i b_j \text{cov}(\mathbf{X}_i,\mathbf{X}_j)$
- In dealing with n random variables, matrix notation is convenient
- $\underline{\mathbf{X}} = (\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n)$
- $E[\underline{\mathbf{X}}] = (E[\mathbf{X}_1],E[\mathbf{X}_2],\ldots,E[\mathbf{X}_n])$
- $\underline{a} = (a_1,a_2,\ldots,a_n)$
- $\underline{a}\bullet\underline{\mathbf{X}} = \sum_{i=1}^{n} a_i\mathbf{X}_i = \underline{a}\underline{\mathbf{X}}^T$
- $E[\underline{a}\bullet\underline{\mathbf{X}}] = \underline{a}\bullet E[\underline{\mathbf{X}}] = \underline{a}\{E[\underline{\mathbf{X}}]\}^T$
- $R =$ n×n covariance matrix has entries $R_{ij} = \text{cov}(\mathbf{X}_i,\mathbf{X}_j)$
- $R$ is a symmetric positive semidefinite matrix
- $R$ is a diagonal matrix if, for all $i \ne j$, $\mathbf{X}_i$ and $\mathbf{X}_j$ are uncorrelated random variables
- Special case: $R$ is diagonal if the $\mathbf{X}_i$ are independent random variables
- $\text{cov}(\underline{a}\bullet\underline{\mathbf{X}}, \underline{b}\bullet\underline{\mathbf{X}}) = \underline{a}R\underline{b}^T$
- $\text{var}(\underline{a}\bullet\underline{\mathbf{X}}) = \text{cov}(\underline{a}\bullet\underline{\mathbf{X}},\underline{a}\bullet\underline{\mathbf{X}}) = \underline{a}R\underline{a}^T$ is a positive semidefinite quadratic form in the n coefficients $a_1, a_2, \ldots, a_n$
- Let $\underline{\mathbf{Y}} = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m) = \underline{\mathbf{X}}A$ where $A$ is an n×m matrix. Then, $E[\underline{\mathbf{Y}}] = E[\underline{\mathbf{X}}]A$
- Let $B = [B_{ij}] = [\text{cov}(\mathbf{Y}_i,\mathbf{Y}_j)] =$ m×m covariance matrix of $\underline{\mathbf{Y}}$
- Then, $B = A^T R A$

- **Jointly Gaussian random variables**
- If **X** is an $N(\mu_{\mathbf{X}}, \sigma^2_{\mathbf{X}})$ random variable, then its pdf is $f_{\mathbf{X}}(u) = \frac{1}{\sigma_{\mathbf{X}}}\varphi\left(\frac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)$ where $\varphi$ is the unit Gaussian pdf $\frac{1}{\sqrt{2\pi}}\exp\left(\frac{u^2}{2}\right)$
- If $\mathbf{X} \sim N(\mu_{\mathbf{X}},\sigma^2_{\mathbf{X}})$, $\mathbf{Y} \sim N(\mu_{\mathbf{Y}},\sigma^2_{\mathbf{Y}})$ are independent Gaussian random variables, then

  $f_{\mathbf{X},\mathbf{Y}}(u,v) = f_{\mathbf{X}}(u)f_{\mathbf{Y}}(v) = \frac{1}{\sigma_{\mathbf{X}}}\varphi\left(\frac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)\frac{1}{\sigma_{\mathbf{Y}}}\varphi\left(\frac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right) = \frac{1}{2\pi\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}}\exp\left(-\frac{1}{2}\left(\frac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)^2 - \frac{1}{2}\left(\frac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right)^2\right)$
- The pdf is a flattened bell
- A contour of a surface is a line through all points at a given height above the plane
- Contours of the joint pdf are ellipses centered at $(\mu_{\mathbf{X}},\mu_{\mathbf{Y}})$ defined by the equation

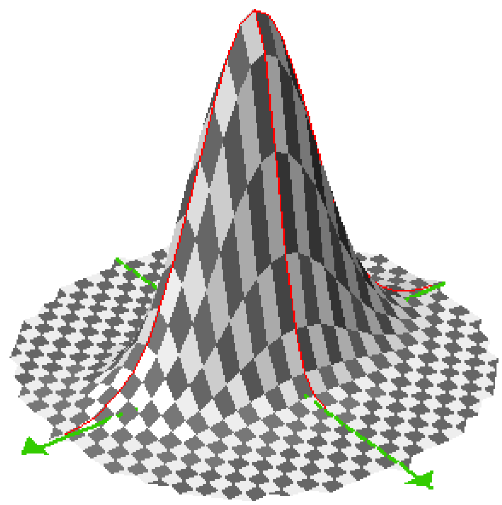
$$\left(\frac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)^2 + \left(\frac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right)^2 = \text{constant.}$$

- The figures on the next page illustrate the shape of the Gaussian pdf for the two cases $\sigma^2_{\mathbf{X}} > \sigma^2_{\mathbf{Y}}$ and $\sigma^2_{\mathbf{X}} < \sigma^2_{\mathbf{Y}}$. In both cases, the point $(\mu_{\mathbf{X}},\mu_{\mathbf{Y}}) = (0,0)$. Notice that the shapes are identical but oriented differently with respect to the coordinate axes.

- The figures below illustrate the shape of the Gaussian pdf for the cases $\sigma_{\mathbf{X}}^2 > \sigma_{\mathbf{Y}}^2$ and $\sigma_{\mathbf{X}}^2 < \sigma_{\mathbf{Y}}^2$.

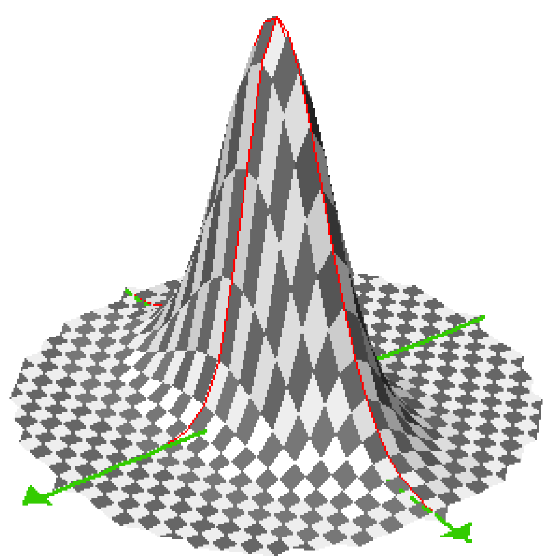
Joint pdf of two independent Gaussian random variables

$(\mu_{\mathbf{X}}, \mu_{\mathbf{Y}}) = (0, 0);\quad \sigma_{\mathbf{X}}^2 > \sigma_{\mathbf{Y}}^2$



Joint pdf of two independent Gaussian random variables

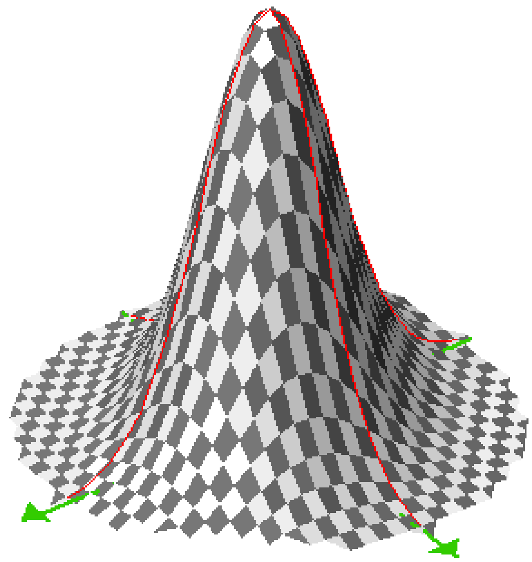
$(\mu_{\mathbf{X}}, \mu_{\mathbf{Y}}) = (0,0);\quad \sigma_{\mathbf{X}}^2 < \sigma_{\mathbf{Y}}^2$



- $\mathbf{X}$ and $\mathbf{Y}$ are said to be jointly Gaussian random variables if their joint pdf is of the form
$$f_{\mathbf{X},\mathbf{Y}}(u,v) = C\bullet\exp(-Q(u,v))$$
where $C = \dfrac{1}{2\pi\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}\sqrt{1-\rho^2}}$, and $Q(u,v) = \dfrac{1}{2(1-\rho^2)}\left[\left(\dfrac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)^2 - 2\rho\left(\dfrac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)\left(\dfrac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right) + \left(\dfrac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right)^2\right]$
- Here the parameters have their usual meanings: $\mu_{\mathbf{X}}$ is the mean of $\mathbf{X}$, $\sigma^2_{\mathbf{X}}$ is the variance of $\mathbf{X}$ and so on. In fact, If $\mathbf{X}$ and $\mathbf{Y}$ are jointly Gaussian random variables, then their marginal pdfs are also Gaussian, that is, $\mathbf{X} \sim N(\mu_{\mathbf{X}},\sigma^2_{\mathbf{X}})$, $\mathbf{Y} \sim N(\mu_{\mathbf{Y}},\sigma^2_{\mathbf{Y}})$
- $\mathbf{X}$ and $\mathbf{Y}$ jointly Gaussian      $\mathbf{X}$ and $\mathbf{Y}$ are individually Gaussian random variables
- In the above equation, $\rho$ denotes the correlation coefficient of $\mathbf{X}$ and $\mathbf{Y}$, where we do not allow $\rho$ to have value $\pm 1$ (why not?). Thus, $-1 < \rho < +1$ in the above equation.
- Recall that $\mathrm{cov}(\mathbf{X},\mathbf{Y}) = \rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}$.
- If $\rho = 0$, then C simplifies to $\dfrac{1}{2\pi\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}}$ and Q(u,v) simplifies to $\dfrac{1}{2}\left(\dfrac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)^2 + \dfrac{1}{2}\left(\dfrac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right)^2$
- If $\rho = 0$, the joint Gaussian pdf reduces to $\dfrac{1}{\sigma_{\mathbf{X}}}\varphi\left(\dfrac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)\dfrac{1}{\sigma_{\mathbf{Y}}}\varphi\left(\dfrac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right) = f_{\mathbf{X}}(u)f_{\mathbf{Y}}(v)$
- **Moral:** uncorrelated jointly Gaussian random variables are also independent Gaussian random variables. Conversely, if $\mathbf{X}$ and $\mathbf{Y}$ are independent Gaussian random variables, then their joint pdf has the jointly Gaussian form exhibited above (with $\rho = 0$)
- The jointly Gaussian pdf is a flattened bell. The contours are ellipses centered at $(\mu_{\mathbf{X}},\mu_{\mathbf{Y}})$, but the axes of the ellipses are tilted with respect to the u-v axes. This is in contrast to the case of independent Gaussian random variables for which the axes of the ellipses are parallel to the u-v axes.
- The next page shows the joint pdf of two jointly Gaussian random variables with equal variance $\sigma^2_{\mathbf{X}} = \sigma^2_{\mathbf{Y}}$ for the two cases $\rho > 0$ and $\rho < 0$. In both cases, the point $(\mu_{\mathbf{X}},\mu_{\mathbf{Y}}) = (0,0)$. It is interesting to compare these figures to those on the previous page

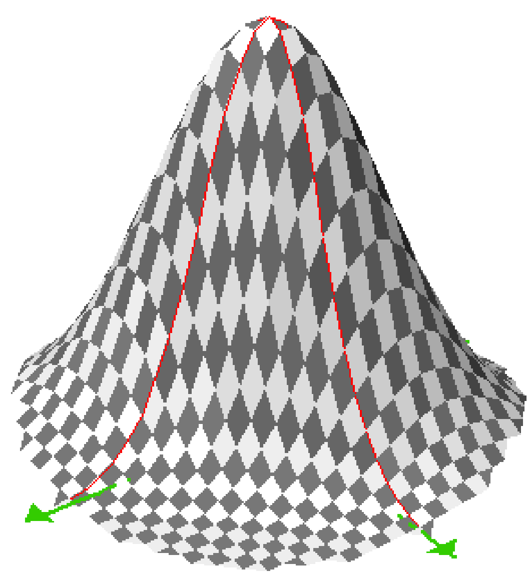
Joint pdf of two jointly Gaussian random variables

$(\mu_X, \mu_Y) = (0, 0)$; $\sigma_X^2 = \sigma_Y^2$; $\rho > 0$

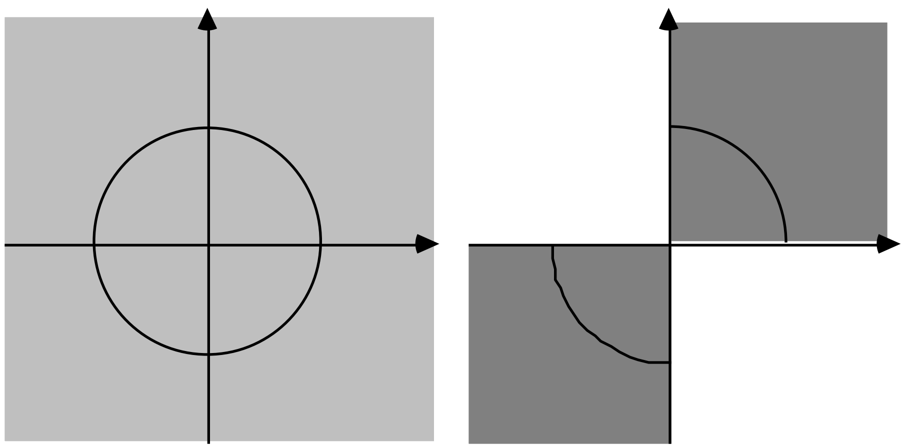


Joint pdf of two jointly Gaussian random variables

$(\mu_X, \mu_Y) = (0, 0)$; $\sigma_X^2 = \sigma_Y^2$; $\rho < 0$



- If $\mathbf{X}$ and $\mathbf{Y}$ are jointly Gaussian random variables, then their marginal pdfs are also Gaussian
- If $\rho = 0$, then C simplifies to $\frac{1}{2\pi\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}}$ and Q(u,v) simplifies to $\frac{1}{2}\left(\frac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)^2 + \frac{1}{2}\left(\frac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right)^2$
- If $\rho = 0$, the joint Gaussian pdf reduces to $\frac{1}{\sigma_{\mathbf{X}}}\phi\left(\frac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)\frac{1}{\sigma_{\mathbf{Y}}}\phi\left(\frac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right) = f_{\mathbf{X}}(u)f_{\mathbf{Y}}(v)$
- **Moral:** uncorrelated jointly Gaussian random variables are also independent Gaussian random variables. Conversely, if $\mathbf{X}$ and $\mathbf{Y}$ are independent Gaussian random variables, then their joint pdf has the jointly Gaussian form exhibited above (with $\rho = 0$)
- $\mathbf{X}$ and $\mathbf{Y}$ jointly Gaussian  $\mathbf{X}$ and $\mathbf{Y}$ are individually Gaussian random variables
- However, $\mathbf{X}$ and $\mathbf{Y}$ having Gaussian pdfs does NOT imply that their joint pdf must be the jointly Gaussian pdf C•exp(–Q(u,v))
- **Illustration:** $f_{\mathbf{X},\mathbf{Y}}(u,v) = \begin{cases} \frac{1}{\pi}\exp\left(-\frac{u^2+v^2}{2}\right), & \text{if } \operatorname{sgn}(u) = \operatorname{sgn}(v), \\ 0, & \text{if } \operatorname{sgn}(u) \neq \operatorname{sgn}(v). \end{cases}$



- $\mathbf{X}$ and $\mathbf{Y}$ jointly Gaussian  $\mathbf{X}$ and $\mathbf{Y}$ are individually Gaussian random variables
- Conditioned on $\mathbf{X} = u$, the pdf of $\mathbf{Y}$ is Gaussian with mean $\mu_{\mathbf{Y}} + \rho(\sigma_{\mathbf{Y}}/\sigma_{\mathbf{X}})(u - \mu_{\mathbf{X}})$ and variance $(\sigma_{\mathbf{Y}})^2(1-\rho^2)$
- $f_{\mathbf{X},\mathbf{Y}}(u,v) = f_{\mathbf{Y}|\mathbf{X}}(v|u)f_{\mathbf{X}}(u)$
- But, $f_{\mathbf{X},\mathbf{Y}}(u,v) = C\bullet\exp(-Q(u,v)) = \frac{1}{\sigma_{\mathbf{X}}}\phi\left(\frac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)\bullet\frac{1}{\gamma}\phi\left(\frac{v-\delta}{\gamma}\right)$

  where $\gamma = (\sigma_{\mathbf{Y}})^2(1-\rho^2)$ and $\delta = \mu_{\mathbf{Y}} + \rho(\sigma_{\mathbf{Y}}/\sigma_{\mathbf{X}})(u - \mu_{\mathbf{X}})$ after some algebraic manipulation
- $f_{\mathbf{Y}|\mathbf{X}}(v|u)$ = cross-section of $f_{\mathbf{X},\mathbf{Y}}(u,v)$ surface at u (with area normalized to 1)
- *Every* cross-section of a jointly Gaussian surface is a Gaussian shape
- Cross-section can be along *any* straight line au + bv = c
- If $\mathbf{X}$ and $\mathbf{Y}$ are jointly Gaussian, $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ is also Gaussian

- To find the pdf of **Z**, DO NOT attempt to use $f_{\mathbf{Z}}(\alpha) = \int_{u=-\infty}^{\infty} f_{\mathbf{X},\mathbf{Y}}(u,\alpha-u)du$
- To find the pdf of **Z**, note first that $\mathbf{Z} = \mathbf{X} + \mathbf{Y}$ is Gaussian
- $\mu_{\mathbf{Z}} = E[\mathbf{Z}] = \mu_{\mathbf{X}} + \mu_{\mathbf{Y}}$
  $\sigma^2_{\mathbf{Z}} = \mathrm{var}(\mathbf{Z}) = \mathrm{var}(\mathbf{X} + \mathbf{Y}) = \mathrm{var}(\mathbf{X}) + \mathrm{var}(\mathbf{Y}) + 2\mathrm{cov}(\mathbf{X},\mathbf{Y}) = \sigma^2_{\mathbf{X}} + \sigma^2_{\mathbf{Y}} + 2\rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}$
- Now, we can *write* down the pdf of **Z** as $f_{\mathbf{Z}}(\alpha) = \frac{1}{\sigma_{\mathbf{Z}}}\varphi\left(\frac{\alpha-\mu_{\mathbf{Z}}}{\sigma_{\mathbf{Z}}}\right)$
- Similarly, $\mathbf{W} = a\mathbf{X} + b\mathbf{Y}$ is Gaussian for any a and b
- $\mu_{\mathbf{W}} = a\mu_{\mathbf{X}} + b\mu_{\mathbf{Y}}$; $\sigma^2_{\mathbf{W}} = \mathrm{var}(\mathbf{W}) = (a\sigma_{\mathbf{X}})^2 + (b\sigma_{\mathbf{Y}})^2 + 2ab\rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}$ and we can write down the pdf
- **X** and **Y** are said to be jointly Gaussian if their joint pdf is $f_{\mathbf{X},\mathbf{Y}}(u,v) = C\bullet\exp(-Q(u,v))$ where
  $C = \frac{1}{2\pi\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}\sqrt{1-\rho^2}}$, and $Q(u,v) = \frac{1}{2(1-\rho^2)}\left[\left(\frac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)^2 - 2\rho\left(\frac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)\left(\frac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right) + \left(\frac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right)^2\right]$
- As $\rho \to \pm 1$, the pdf $\to \infty$
- This is a degenerate case
- If $\rho = \pm 1$, all the probability mass lies along a straight line
- **X** and **Y** are *not* jointly continuous, ($\mathbf{Y} = \alpha\mathbf{X} + \beta$) and we can solve all problems in terms of **X** alone
- Updated definition: **X** and **Y** are said to be jointly Gaussian if their joint pdf is of the form $f_{\mathbf{X},\mathbf{Y}}(u,v) = C\bullet\exp(-Q(u,v))$ or if **X** is a Gaussian random variable and $\mathbf{Y} = \alpha\mathbf{X} + \beta$
- Note that $\alpha\mathbf{X} + \beta$ is Gaussian
- An alternative definition that avoids the special case:
  **X** and **Y** are said to be jointly Gaussian random variables if $a\mathbf{X} + b\mathbf{Y}$ is Gaussian for all choices of a and b (except the trivial case a = b = 0)
- Puts fundamental fact up front but what's the joint pdf?
- If **X** and **Y** are jointly Gaussian, then so are $\mathbf{W} = a\mathbf{X} + b\mathbf{Y}$ and $\mathbf{Z} = c\mathbf{X} + d\mathbf{Y}$ for all nontrivial choices of a, b, c, d
- We can *write* down the pdf by finding the means, variances and correlation coefficient
- $E[\mathbf{W}] = aE[\mathbf{X}] + bE[\mathbf{Y}]$
- $E[\mathbf{Z}] = cE[\mathbf{X}] + dE[\mathbf{Y}]$
- $\mathrm{var}(\mathbf{W}) = (a\sigma_{\mathbf{X}})^2 + (b\sigma_{\mathbf{Y}})^2 + 2ab\rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}$
- $\mathrm{var}(\mathbf{Z}) = (c\sigma_{\mathbf{X}})^2 + (d\sigma_{\mathbf{Y}})^2 + 2cd\rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}$
- $\mathrm{cov}(\mathbf{W},\mathbf{Z}) = \mathrm{cov}(a\mathbf{X} + b\mathbf{Y}, c\mathbf{X} + d\mathbf{Y}) = ac\,\mathrm{var}(\mathbf{X}) + bd\,\mathrm{var}(\mathbf{Y}) + (ad + bc)\,\mathrm{cov}(\mathbf{X},\mathbf{Y})$
  $= ac\sigma^2_{\mathbf{X}} + bd\sigma^2_{\mathbf{Y}} + 2(ad+bc)\rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}$
  $\rho_{\mathbf{W},\mathbf{Z}} = \mathrm{cov}(\mathbf{W},\mathbf{Z})/\sqrt{\mathrm{var}(\mathbf{W})\mathrm{var}(\mathbf{Z})}$
- The contours of the joint Gaussian pdf are ellipses centered at $(\mu_{\mathbf{X}},\mu_{\mathbf{Y}})$, but the axes of the ellipses are tilted with respect to the u-v axes

- The transformation $\mathbf{W} = \mathbf{X}\cos\theta + \mathbf{Y}\sin\theta$ $\quad \mathbf{Z} = -\mathbf{X}\sin\theta + \mathbf{Y}\cos\theta$ rotates the axes by angle $\theta$
- $f_{\mathbf{W},\mathbf{Z}}(\alpha,\beta)$ is the same surface as $f_{\mathbf{X},\mathbf{Y}}(u,v)$ except that the $\alpha$-$\beta$ axes are rotated by angle $\theta$ with respect to the u-v axes
- **Exercise:** For appropriate choice of $\theta$, the $\alpha$-$\beta$ axes coincide with the axes of the ellipses, and thus **W** and **Z** are independent Gaussian random variables. Find $\theta$ in terms of the means, the variances, and the covariance of **X** and **Y**
- **Multivariate jointly Gaussian distributions**
- We begin by formulating the two-dimensional Gaussian distribution in terms of matrices.
- **X** and **Y** are said to be jointly Gaussian if their joint pdf is $f_{\mathbf{X},\mathbf{Y}}(u,v) = C\bullet\exp(-Q(u,v))$ where $C = \dfrac{1}{2\pi\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}}\sqrt{1-\rho^2}}$, and $Q(u,v) = \dfrac{1}{2(1-\rho^2)}\left[\left(\dfrac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)^2 - 2\rho\left(\dfrac{u-\mu_{\mathbf{X}}}{\sigma_{\mathbf{X}}}\right)\left(\dfrac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right) + \left(\dfrac{v-\mu_{\mathbf{Y}}}{\sigma_{\mathbf{Y}}}\right)^2\right]$
- The *covariance matrix* of $(\mathbf{X},\mathbf{Y})$ is defined to be $R = \begin{bmatrix} \sigma_{\mathbf{X}}^2 & \rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}} \\ \rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}} & \sigma_{\mathbf{Y}}^2 \end{bmatrix}$

  with inverse given by $R^{-1} = |R|^{-1}\begin{bmatrix} \sigma_{\mathbf{Y}}^2 & -\rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}} \\ -\rho\sigma_{\mathbf{X}}\sigma_{\mathbf{Y}} & \sigma_{\mathbf{X}}^2 \end{bmatrix}$
- Note that if we define $\underline{u} = (u,v)$, then $f_{\mathbf{X},\mathbf{Y}}(u,v) = C\bullet\exp(-Q(u,v)) = C\bullet\exp(-Q(\underline{u}))$
- Let $\underline{\mu} = (E[\mathbf{X}], E[\mathbf{Y}])$. We can thus write $f_{\mathbf{X},\mathbf{Y}}(u,v) = C\bullet\exp(-Q(\underline{u}))$ where $C = \dfrac{1}{2\pi|R|^{1/2}}$ and $Q(\underline{u}) = \frac{1}{2}(\underline{u}-\underline{\mu})R^{-1}(\underline{u}-\underline{\mu})^T$ is called a *quadratic form.*
- $\underline{\mathbf{X}} = (\mathbf{X}_1,\mathbf{X}_2,\ldots,\mathbf{X}_n)$ is said to be a Gaussian vector if the $\mathbf{X}_i$'s are jointly Gaussian random variables with pdf given by $f_{\underline{\mathbf{X}}}(\underline{u}) = \dfrac{1}{(2\pi)^{n/2}|R|^{1/2}} \times \exp\left(-\frac{1}{2}(\underline{u}-\underline{\mu})R^{-1}(\underline{u}-\underline{\mu})^T\right)$
- The marginal pdfs of the $\mathbf{X}_i$'s are Gaussian with mean $\mu_i$ and variance $R_{ii}$
- $\mathrm{cov}(\mathbf{X}_i, \mathbf{X}_j) = R_{ij}$
- **Fundamental fact:** Linear transformations of Gaussian vectors yield Gaussian vectors
- If $\underline{\mathbf{Y}} = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m) = \underline{\mathbf{X}}A$, then $E[\underline{\mathbf{Y}}] = E[\underline{\mathbf{X}}]A$ and the covariance matrix of $\underline{\mathbf{Y}}$ is $A^TRA$
- Once again, we can *write* down the pdf of the Gaussian vector $\underline{\mathbf{Y}}$ by means of simple calculations
- Special cases arise when $R$ is singular. In this case, a joint pdf cannot be defined (just as in the case of two variables and $\rho = \pm 1$)
- What can be done if $R$, the covariance matrix of $\underline{\mathbf{X}} = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n)$, is singular?
- Suppose that $R$ is singular
- Then, there exists a Gaussian vector $\underline{\mathbf{Z}} = (\mathbf{Z}_1,\mathbf{Z}_2,\ldots,\mathbf{Z}_m)$ with $m < n$ and nonsingular covariance matrix $S$ such that $\underline{\mathbf{X}} = \underline{\mathbf{Z}}B$ where $B$ is $m\times n$ and $R = B^TSB$
- Do everything in terms of $\underline{\mathbf{Z}}$