Example : In the development of a new receiver for the transmission of digital information, each received bit is rated as *acceptable*, *suspect* or *unacceptable*, depending on the quality of the received signal, with probabilities 0.9, 0.08 and 0.02, respectively. Assume that the ratings of each bit are independent.

In the first four bits transmitted, let

x denote the number of acceptable bits

y denote the number of suspect bits

Q: What are the distributions for x and y?

A: $x \sim Bi(4, 0.9)$ and $y \sim Bi(4, 0.08)$.

Q: Are x and y independent?

A: No.

Q: What is the distribution for (x, y)?

Find the **Joint Probability Distribution** of the pair of random variables (x, y).

- For example, what is

$$p(2,1) = P(x = 2, y = 1)?$$

- $p(2,1)$ is the probability of having 2 acceptable bits and 1 suspect bit out of 4.

  Q: How many bits are unacceptable?

  A: 1.

  Q: What's the probability for simple event AASU

  A: $(0.9)^2(0.08)(0.02)$

  Q: What's the probability for simple event ASUA?

  Q: How many simple events are included in event $\{x = 2, y = 1\}$?

  A: $4!/(2!1!1!)$

- $$p(2,1) = \frac{4!}{2!1!1!}P(AASU) = \frac{4!}{2!1!1!}(0.9)^2(0.08)(0.02)$$

- $(X, Y) \sim MN(4, \{0.9, 0.08, 0.02\})$

Bivariate Distribution

- The **joint probability distribution** for two discrete random variables, x and y–called a **bivariate distribution**, is a function $p(x, y)$ for which

  – $0 \leq p(x, y) \leq 1$ for all values of x and y

  – $\sum_x \sum_y p(x, y) = 1$

  - $p(x_0, y_0)$ is equal to the probability of the event $\{x = x_0 \text{ AND } y = y_0\}$

- The **range** of the joint distribution is the collection of all possible values of both x and y.

Marginal Distribution

- We can recover the (univariate) distribution of x and y by summing the joint distribution over the **other** variable:

$$P(x = x_0) = \sum_y p(x_0, y) = p_x(x_0)$$

and

$$P(y = y_0) = \sum_x p(x, y_0) = p_y(y_0)$$

- There are called **marginal (unconditional) probability distributions** of x and y, because if we represent $p(x, y)$ as a table, then these distributions would be row and column sums, which one might put in the **margins** of the table.

- We use the subscript x (as in p_x) when we refer to the marginal probability distribution of random variable x. Same for p_y.

Conditional Distribution

- What if I knew that there were 3 acceptable bits ($x = 3$), and I was interested in $p_y(y)$ **given** this information?

- The random variables (x, y) are now known to be $(3, y)$, i.e., we are only interested in **one column** in the table of the bivariate distribution.

  Q: How to turn this column into a probability distribution?

  A: Normalize it to sum to 1.

y	$p(3, y)$	$p_{y\|x}(y \mid 3)$
0	.0583	.200
1	.2333	.800
$p_x(x)$	.2916	

- Recall that the probability of a bit being suspect is 4 times the probability of a bit being unacceptable. So this conditional probability distribution makes sense.

Conditional Distribution (Cont'd)

- We just derived the distribution of y given $x = 3$. This is a function of y, which we denote by

$$p_{y|x}(y \mid 3) = P(y \mid x = 3)$$

- If we condition on x taking on any values for which $p_x(x) > 0$, then we have a function of two variables, a **conditional probability distribution**:

$$p_{x|y}(x \mid y) = \frac{p(x, y)}{p_y(y)} = \frac{p(x, y)}{\sum_x p(x, y)}$$

- Note that for any value of y,

$$\sum_x p(x \mid y) = 1.$$

Bivariate Probability Density

A function of two variables $f(x, y)$ is a **bivariate joint probability density function** for two continuous random variables x and y, if

- $f(x, y) \geq 0$ for all values of x and y.
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$
- $P((x, y) \in R) = \iint_R f(x, y) dx dy$ where R is any region in the (x, y) plane. Specially

  $$P(a \leq x \leq b, c \leq y \leq d) = \int_c^d \int_a^b f(x, y) dx dy$$

  for all constants a, b, c and d.

Marginal and Conditional Densities

- We can define marginal and conditional densities just as we did for discrete random variables

- **Marginal Densities :**

$$
\begin{aligned}
f_x(x) &= \int_{-\infty}^{\infty} f(x,y)dy \\
f_y(y) &= \int_{-\infty}^{\infty} f(x,y)dx
\end{aligned}
$$

- **Conditional Densities :**

$$
\begin{aligned}
f_{x|y}(x \mid y) &= \frac{f(x,y)}{f_y(y)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dx} \\
f_{y|x}(y \mid x) &= \frac{f(x,y)}{f_x(x)} = \frac{f(x,y)}{\int_{-\infty}^{\infty} f(x,y)dy}
\end{aligned}
$$

Independence of Events (Revisit)

- The **conditional probability** that event A occurs given that event B occurs is defined to be

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)},$$

where $P(B) > 0$.

- Events A and B are **independent**, if

$$P(A \mid B) = P(A)$$

which is equivalent to

$$P(A \cap B) = P(A)P(B)$$

- Note that if $P(A), P(B) > 0$, A and B are **mutually exclusive** (i.e $A \cap B = \emptyset$), then A and B are NOT independent.

Independence of Random Variables

- The random variables x and y are **independent** if and only if for all values of x and y

  $p(x,y) = p_x(x)p_y(y)$ x and y are discrete

  $f(x,y) = f_x(x)f_y(y)$ x and y are continuous

- That is, the following are equivalent to the random variables x and y being independent:

$$\begin{aligned} f_{y|x}(y \mid x) &= f_y(y) \\ f_{x|y}(x \mid y) &= f_x(x) \end{aligned}$$

  Same for discrete probability functions.

- If x and y are independent, then

$$\mathbb{E}(xy) = \mathbb{E}(x)\mathbb{E}(y)$$

Expectation

The expected value of a function of random variable x and y, $g(x, y)$, is defined to be

$$\mathbb{E}[g(x,y)] = \sum_y \sum_x g(x,y)p(x,y)$$

$$\mathbb{E}[g(x,y)] = \iint g(x,y)f(x,y)\ dxdy$$

- $\mathbb{E}(c) = c$
- $\mathbb{E}[cg(x,y)] = c\mathbb{E}[g(x,y)]$
- $\mathbb{E}[g_1(x,y)+g_2(x,y)] = \mathbb{E}[g_1(x,y)]+\mathbb{E}[g_2(x,y)]$

Q: What's the expectation of a linear combination of random variables, $a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$, where x_i are random variables and a_i are constants?

A:

$$\begin{aligned} \mu &= a_1 \mathbb{E}(x_1) + \cdots + c_n \mathbb{E}(x_n) \\ &= a_1 \mu_1 + \cdots + a_n \mu_n \end{aligned}$$

Q: What's the variance?

A:

$$\begin{aligned} \sigma^2 &= \mathbb{E}(a_1 x_1 + \cdots + a_n x_n - \mu)^2 \\ &= \mathbb{E}[a_1(x_1 - \mu_1) + \cdots + a_n(x_n - \mu_n)]^2 \\ &= \sum_{i,j=1}^{n} a_i a_j \mathbb{E}(x_i - \mu_i)(x_j - \mu_j) \\ &= \sum_i^n a_i^2 \sigma_i^2 + \sum_{i \neq j} a_i a_j \mathbb{E}(x_i - \mu_i)(x_j - \mu_j) \\ &= \sum_i^n a_i^2 \sigma_i^2 + \sum_{i \neq j} a_i a_j \mathrm{Cov}(x_i, x_j) \end{aligned}$$

Covariance

- It is sometimes useful to say something about how two random variables vary together.

- The **covariance** is defined as

$$\text{Cov}(x, y) = \mathbb{E}[(x - \mu_x)(y - \mu_y)]$$

- Note that if $x = y$, this reduces to the variance σ_x^2

- An equivalent expression for covariance

$$\begin{aligned} \text{Cov}(x, y) &= \mathbb{E}[xy - \mu_x y - \mu_y x + \mu_x \mu_y] \\ &= \mathbb{E}(xy) - \mu_x \mu_y - \mu_y \mu_x + \mu_x \mu_y \\ &= \mathbb{E}(xy) - \mu_x \mu_y \end{aligned}$$

- If x and y are independent, then

$$\text{Cov}(x, y) = 0.$$

Correlation

- It turns out that when the covariance is normalized by dividing by the product of σ_x and σ_y, then its value will always be in the interval $[-1, 1]$,

$$-1 \leq \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y} \leq 1$$

- This "normalized covariance" is called the **coefficient of correlation**, and usually denoted by ρ:

$$\rho \equiv \frac{\text{Cov}(x, y)}{\sigma_x \sigma_y}$$

- $\rho = 1$ and $\rho = -1$ imply deterministic linear relationships between x and y, the former with a positive slope and the latter with a negative slope. $\rho = 0$ implies no linear relationship between x and y.

Independent vs Uncorrelated

- Independence implies $\rho = 0$: if x and y are independent, then

$$\begin{aligned} \text{Cov}(x, y) &= \mathbb{E}[(x - \mu_x)(y - \mu_y)] \\ &= \mathbb{E}(x - \mu_x)\mathbb{E}(y - \mu_y) \\ &= 0 \end{aligned}$$

- Uncorrelated random variables are NOT necessarily independent. But if they are **uncorrelated** and **normal**, then they are independent.

Example: Suppose $x, z \sim N(0,1)$. Define

$$y = \rho x + \sqrt{1-\rho^2} z$$

What is the joint distribution for (x, y)?

- Q: What's the expectation and variance of y?

  A:

$$\begin{aligned}
\mathbb{E}y &= \rho\mathbb{E}(x) + \sqrt{1-\rho^2}\mathbb{E}(z) \\
&= 0 \\
\text{Var}(y) &= \mathbb{E}(\rho x + \sqrt{1-\rho^2} z)^2 \\
&= \rho^2\mathbb{E}(x^2) + (1-\rho^2)\mathbb{E}(z^2) \\
&= 1
\end{aligned}$$

- Q: What's the conditional distribution $f_{y|x}(y \mid x)$?

  A:

$$\begin{aligned}
P(y \le b \mid x) &= P(z \le \frac{b-\rho x}{\sqrt{1-\rho^2}}) \\
&= \int_{-\infty}^{(b-\rho x)/\sqrt{1-\rho^2}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz
\end{aligned}$$

$$f_{y|x}(y \mid x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{(y-\rho x)^2}{2(1-\rho^2)}\} \frac{1}{\sqrt{1-\rho^2}}$$

- Q: What's the bivariate density function $f(x, y)$?

A:

$$
\begin{aligned}
f(x, y) &= f_{y|x}(y \mid x) f_x(x) \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\{\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\}
\end{aligned}
$$

where $-\infty < x, y < \infty$ with parameter $-1 < \rho < 1$.

- Q: What's the marginal density for y?

  A:

$$
\begin{aligned}
f_y(y) &= \int \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\{\frac{-1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\} dx \\
&= \int \frac{1}{2\pi\sqrt{1-\rho^2}} \times \\
&\quad \exp\{\frac{-1}{2(1-\rho^2)}[(x-\rho y)^2 + (1-\rho^2)y^2]\} dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \times \\
&\quad \int \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\{\frac{-1}{2(1-\rho^2)}(x-\rho y)^2\} dx \\
&= \frac{1}{\sqrt{2\pi}} e^{-y^2/2}
\end{aligned}
$$

Bivariate Normal Distribution

Two random variables x and y have a **bivariate normal distribution** if the joint density of x and y is

$$f(x,y) = \frac{\exp\{\frac{-1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)\}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}$$

where

$$u = \frac{x-\mu_x}{\sigma_x}$$
$$v = \frac{y-\mu_y}{\sigma_y}$$

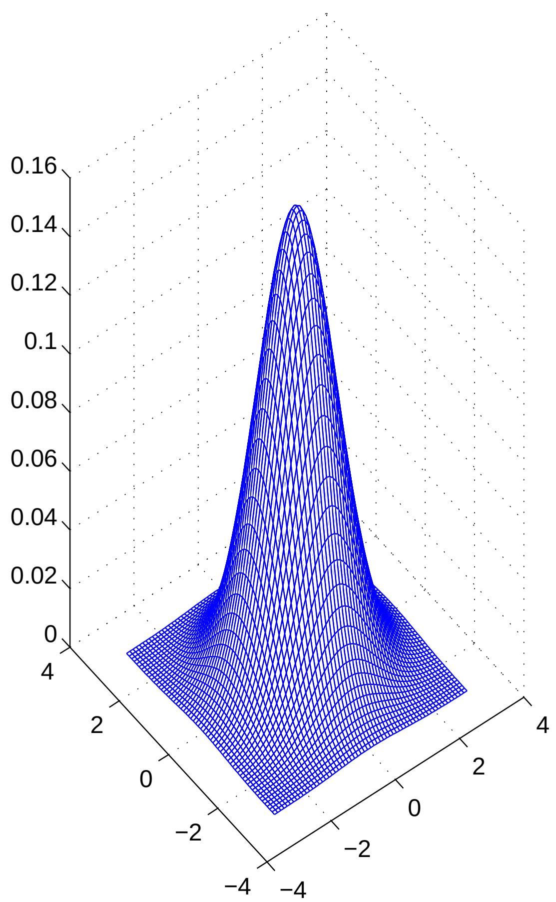
and

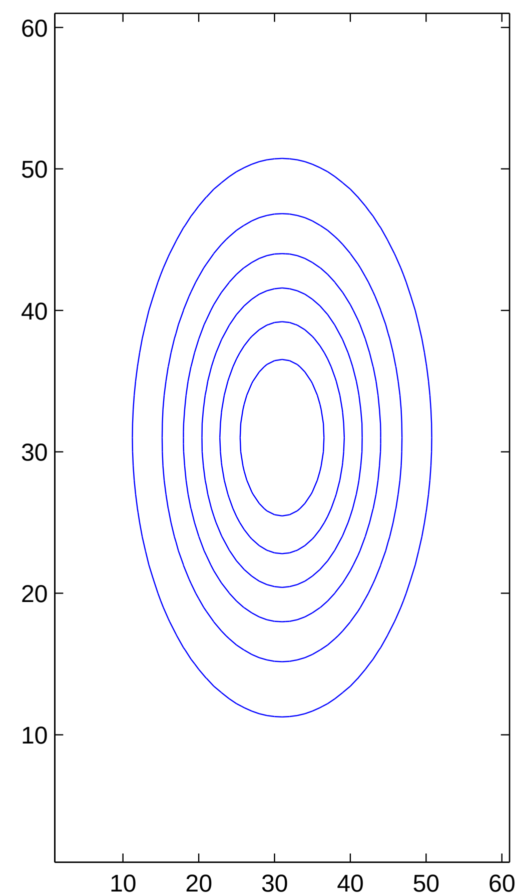
$$-1 < \rho < 1$$

Properties of Bivariate Normal

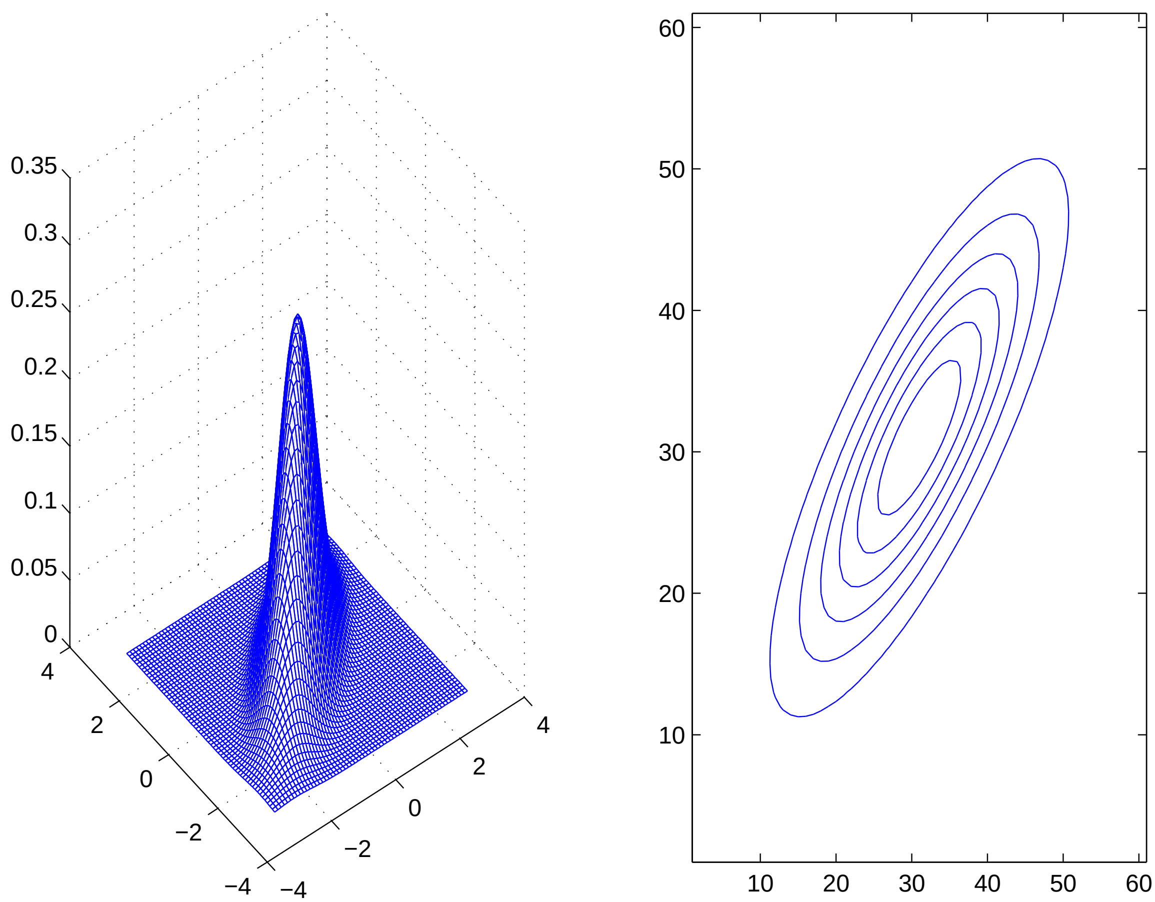
- The mean of (x, y) is (μ_x, μ_y)
- The variance of (x, y) is (σ_x^2, σ_y^2)
- The correlation of x and y is ρ.
- The marginal distribution of x is normal with mean μ_x and variance σ_x^2
- The marginal distribution of y is normal with mean μ_y and variance σ_y^2
- The conditional distributions of $y \mid x$ and $x \mid y$ are both normal
- x and y are independent if $\rho = 0$.

Bivariate normal density function and contour plot with $\rho = 0$





Bivariate normal density function and contour plot with $\rho = 0.8$



Bivariate normal density function and contour plot with $\rho = -0.9$