MTH 464/564

Lectures 1-6

Yevgeniy Kovchegov
Oregon State University

- Joint probability mass function.
- Joint probability density function.
- Independent random variables.
- Covariance and correlation.
- Sums of random variables.
- Gamma and beta random variables.
- Poisson process.

- Marginal distributions from joint distribution.
- Functions of random variables.
- Indicator variables.
- Multivariate normal distribution.

Joint probability distribution.

Definition. Function $p(x,y)$ is said to be a joint probability mass function of discrete random variables X and Y if

$$p(x,y) = P\Big(\{X = x\} \cap \{Y = y\}\Big)$$

for all value x and y.

Joint probability mass function is similarly defined for n discrete random variables.

Properties: (i) $p(x,y) \geq 0$ (ii) $\sum_{x,y} p(x,y) = 1$.

Example. Consider a joint probability mass function

$$p(1,2) = \frac{1}{2}, \quad p(2,0) = \frac{1}{6}, \quad p(3,3) = \frac{1}{3}$$

Joint probability distribution.

Definition. Function $f(x,y)$ is said to be a joint probability density function of continuous random variables X and Y if

$$P\big((X,Y)\in A\big)=\iint_A f(x,y)\,dxdy$$

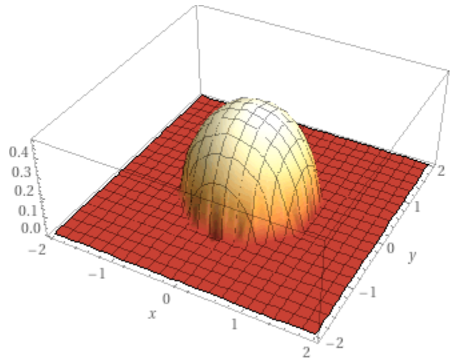
for each domain $A\subseteq\mathbb{R}^2$.

Joint probability density function is similarly defined for n continuous random variables.

Properties: (i) $f(x,y)\ge 0$ (ii) $\iint_{\mathbb{R}^2} f(x,y)\,dxdy=1$.

Example. Consider the joint probability density function

$$f(x,y)=\begin{cases}\frac{3}{2\pi}\sqrt{1-x^2-y^2} & \text{if } x^2+y^2\le 1,\\ 0 & \text{if } x^2+y^2>1.\end{cases}$$



Independent random variables.

Definition. Random variables X and Y are said to be independent if the events

$$\{a \le X \le b\} \quad \text{and} \quad \{c \le Y \le d\}$$

are independent, for any a, b, c, and d. Namely,

$$P\Big(\{a \le X \le b\} \cap \{c \le Y \le d\}\Big) = P(a \le X \le b)\, P(c \le Y \le d)$$

Similarly for n independent random variables.

• If X and Y are discrete random variables with probability mass functions p_{x} and p_{y} respectively. They are independent if and only if

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

for all x and y.

Thus, X and Y are independent if and only if $p(x, y) = p_{\mathsf{x}}(x) p_{\mathsf{y}}(y)$

Independent random variables.

Definition. Random variables X and Y are said to be independent if the events

$$\{a \le X \le b\} \quad \text{and} \quad \{c \le Y \le d\}$$

are independent, for any a, b, c, and d. Namely,

$$P\Big(\{a \le X \le b\} \cap \{c \le Y \le d\}\Big) = P(a \le X \le b)\, P(c \le Y \le d)$$

Similarly for n independent random variables.

• If X and Y are continuous random variables with probability density functions f_{x} and f_{y} respectively. They are independent if and only if

$$\int_c^d \int_a^b f(x,y)\, dx dy = \int_a^b f_{\mathsf{x}}(x) dx \int_c^d f_{\mathsf{y}}(y) dy$$

Thus, X and Y are independent if and only if $f(x,y) = f_{\mathsf{x}}(x) f_{\mathsf{y}}(y)$.

Sums of random variables.

• If X and Y are discrete random variables with joint probability mass function $p(x,y)$, then their sum, $Z = X + Y$ is also a discrete random variable with probability mass function

$$p_{\mathsf{z}}(a) \;=\; \sum_{x,y:\; x+y=a} p(x,y) = \sum_{x} p(x, a-x)$$

• If X and Y are continuous random variables with joint probability density function $f(x,y)$, then $Z = X + Y$ is also a continuous random variable with its density f_Z given as

$$f_{\mathsf{z}}(a) = \int_{-\infty}^{\infty} f(x, a-x)\, dx$$

Indeed, $F_{\mathsf{z}}(a) = P(X+Y \le a) = \iint\limits_{(x,y):\, x+y\le a} f(x,y)\, dydx = \int_{-\infty}^{\infty}\int_{-\infty}^{a-x} f(x,y)\, dydx$

$$f_{\mathsf{z}}(a) = \frac{d}{da} F_{\mathsf{z}}(a) = \int_{-\infty}^{\infty} \frac{d}{da}\left(\int_{-\infty}^{a-x} f(x,y) dy \right) dx = \int_{-\infty}^{\infty} f(x, a-x)\, dx$$

Sums of independent random variables.

• If X and Y are independent discrete random variables with probability mass functions $p_{\mathsf{x}}(x) = P(X = x)$ and $p_{\mathsf{y}}(y) = P(Y = y)$, then their sum, $Z = X + Y$ is also a discrete random variable with probability mass function

$$p_{\mathsf{z}}(a) \;=\; \sum_{x,y:\; x+y=a} p_{\mathsf{x}}(x)\, p_{\mathsf{y}}(y)$$

which can be rewritten as a **convolution sum**:

$$p_{\mathsf{z}}(a) \;=\; \sum_{x} p_{\mathsf{x}}(x)\, p_{\mathsf{y}}(a-x)$$

• If X and Y are independent continuous random variables with density functions f_{x} and f_{y}, then $Z = X + Y$ is also a continuous random variable with its density f_Z given as a **convolution integral**,

$$f_{\mathsf{z}}(a) = \int_{-\infty}^{\infty} f_{\mathsf{x}}(x)\, f_{\mathsf{y}}(a-x)\, dx$$

Sums of independent random variables.

$$p_{\mathsf{x+y}}(a) \;=\; \sum_x p_{\mathsf{x}}(x)\, p_{\mathsf{y}}(a-x)$$

• **Example.** Let X be binomial with parameters (n_1, p) and Y be binomial with parameters (n_2, p). Then their probability mass functions are

$$p_{\mathsf{x}}(k) = \binom{n_1}{k} p^k (1-p)^{n_1-k} \qquad \text{for } k = 0, 1, \ldots, n_1$$

and

$$p_{\mathsf{y}}(k) = \binom{n_2}{k} p^k (1-p)^{n_2-k} \qquad \text{for } k = 0, 1, \ldots, n_2$$

If X and Y are independent, then their sum will have the following distribution: for $j = 0, 1, \ldots, n_1 + n_2$,

$$p_{\mathsf{x+y}}(j) \;= \sum_k p_{\mathsf{x}}(k) p_{\mathsf{y}}(j-k) = \binom{n_1+n_2}{j} p^j\,(1-p)^{n_1+n_2-j} \qquad \text{for } j = 0, 1, \ldots, n_1+n_2$$

Thus $X + Y$ is binomial with parameters $(n_1 + n_2, p)$.

Sums of independent random variables.

$$p_{X+Y}(a) = \sum_x p_X(x)\, p_Y(a-x)$$

• **Example.** Let X be Poisson with parameter λ_1 and Y be Poisson with parameter λ_2. Then their probability mass functions are

$$p_X(k) = e^{-\lambda_1}\frac{\lambda_1^k}{k!} \quad \text{for } k = 0, 1, \ldots$$

and

$$p_Y(k) = e^{-\lambda_2}\frac{\lambda_2^k}{k!} \quad \text{for } k = 0, 1, \ldots$$

If X and Y are independent, then their sum will have the following distribution: for $n = 0, 1, \ldots,$

$$p_{X+Y}(n) = \sum_{k=0}^{\infty} p_X(k)\, p_Y(n-k) = \sum_{k=0}^{n} e^{-\lambda_1}\frac{\lambda_1^k}{k!}\, e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}$$

$$= e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{1}{k!(n-k)!}\,\lambda_1^k\,\lambda_2^{n-k}$$

Sums of independent random variables.

$$p_{x+y}(a) = \sum_x p_x(x)\, p_y(a-x)$$

• **Example (continued).** Let X be Poisson with parameter λ_1 and Y be Poisson with parameter λ_2. If X and Y are independent, then their sum will have the following distribution: for $n = 0, 1, \ldots,$

$$p_{x+y}(n) = e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \frac{1}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} = \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k}$$

$$= \frac{e^{-(\lambda_1+\lambda_2)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \lambda_1^k \lambda_2^{n-k} = e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}$$

by the Binomial Theorem.

Hence, $X+Y$ is Poisson with parameter $\lambda_1+\lambda_2$.

Sums of independent random variables.

$$f_{\mathsf{x+y}}(a) = \int_{-\infty}^{\infty} f_{\mathsf{x}}(x)\, f_{\mathsf{y}}(a-x)\ dx$$

• **Example.** Let X and Y each be uniform over $[0, 1]$. Then each will be distributed according to the following probability density function:

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

If X and Y are independent, then their sum will have the following continuous distribution:

$$f_{\mathsf{x+y}}(a) = \int_{-\infty}^{\infty} f(x)\, f(a-x)\ dx = \int_{0}^{1} f(a-x)dx$$

Observe that $\int_0^1 f(a-x)dx = 0$ whenever $a < 0$ or $a > 2$.

There are two more cases: $0 \leq a \leq 1$ and $1 \leq a \leq 2$.

Sums of independent random variables.

• **Example (continued).** $f(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

If $0 \le a \le 1$,

$$f_{\mathsf{x+y}}(a) = \int_0^1 f(a-x)dx = \int_0^a dx = a$$

Now, if $1 \le a \le 2$,

$$f_{\mathsf{x+y}}(a) = \int_0^1 f(a-x)dx = \int_{a-1}^1 dx = 2-a$$

Therefore,

$$f_{\mathsf{x+y}}(a) = \begin{cases} a & \text{if } 0 \le a \le 1 \\ 2-a & \text{if } 1 \le a \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Properties of convolutions.

Convolution of functions f and g is defined

$$f \circ g(a) = \int_{-\infty}^{\infty} f(x)\, g(a-x)\ dx$$

Another notation: $f * g$.

- Convolution is commutative: $f \circ g(a) = g \circ f(a)$

- Convolution is associative: $(f \circ g) \circ h = f \circ (g \circ h)$

- Convolution is distributive: $(f + g) \circ h = f \circ h + g \circ h$

- $f \circ (cg) = c(f \circ g)$ for all $c \in \mathbb{R}$

- If you are familiar with Fourier transforms: $\widehat{f \circ g}(\xi) = \widehat{f}(\xi)\widehat{g}(\xi)$

Sums of random variables.

• **Theorem.** Expectation of a sum of random variables is equal to the sum of expectations.

$$E[X+Y] = E[X] + E[Y]$$

Notice: we don't require X and Y to be independent here. The above is true even if they are dependent random variables.

• **Theorem.** If X and Y are independent random variables then,

$$Var(X+Y) = Var(X) + Var(Y)$$

Here we need independence.

• **Example.** Let X be Poisson with parameter λ_1 and Y be Poisson with parameter λ_2. For the case when X and Y are independent, we proved that $X+Y$ is Poisson with parameter $\lambda_1 + \lambda_2$. Then

$$E[X] + E[Y] = \lambda_1 + \lambda_2 = E[X+Y]$$

$$Var(X) + Var(Y) = \lambda_1 + \lambda_2 = Var(X+Y)$$

Sums of random variables.

• **Example.** Let X and Y each be uniform over $[0,1]$. For the case when X and Y are independent, we proved that $X+Y$ is distributed according to

$$f_{x+y}(a) = \begin{cases} a & \text{if } 0 \le a \le 1 \\ 2-a & \text{if } 1 \le a \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Check that

$$E[X+Y] = E[X] + E[Y]$$

and

$$Var(X) + Var(Y) = Var(X+Y)$$

Sums of random variables.

• **Example (continued).** Let X and Y each be uniform over $[0,1]$. For the case when X and Y are independent, we proved that $X+Y$ is distributed according to

$$f_{\mathsf{x+y}}(a) = \begin{cases} a & \text{if } 0 \le a \le 1 \\ 2-a & \text{if } 1 \le a \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Here $E[X] = E[Y] = \int_0^1 x\,dx = \frac{1}{2}$ while

$$E[X+Y] = \int_{-\infty}^{\infty} x f_{\mathsf{x+y}}(x)\,dx = \int_0^1 x^2\,dx + \int_1^2 x(2-x)\,dx$$

$$= \frac{1}{3} + \left[x^2 - \frac{x^3}{3}\right]_1^2 = \frac{1}{3} + 4 - \frac{8}{3} - 1 + \frac{1}{3} = 1 = E[X] + E[Y]$$

Sums of random variables.

• **Example (continued).** Let X and Y each be uniform over $[0,1]$. For the case when X and Y are independent, we proved that $X+Y$ is distributed according to

$$f_{x+y}(a) = \begin{cases} a & \text{if } 0 \le a \le 1 \\ 2-a & \text{if } 1 \le a \le 2 \\ 0 & \text{otherwise} \end{cases}$$

Here $Var(X) = E[X^2] - \Big(E[X]\Big)^2 = \int_0^1 x^2 dx - \frac{1}{4} = \frac{1}{12} = Var(Y)$

while $E[X+Y] = 1$, and therefore

$$Var(X+Y) = E[(X+Y)^2]-1 = \int_{-\infty}^{\infty} x^2 f_{x+y}(x)dx - 1 = \int_0^1 x^3 dx + \int_1^2 x^2(2-x)dx - 1$$

$$= \frac{1}{4} + \left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_1^2 - 1 = \frac{1}{4} + \frac{16}{3} - 4 - \frac{2}{3} + \frac{1}{4} - 1 = \frac{1}{6} = Var(X) + Var(Y)$$

Gamma function.

The gamma function $\Gamma(\alpha)$ is defined as

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx \quad \text{for all real } \alpha > 0.$$

Integration by parts yields $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$. Indeed,

$$\Gamma(\alpha+1) = \int_0^\infty e^{-y} y^\alpha dy = \int_0^\infty \left(-e^{-y}\right)' y^\alpha dy = \left(-e^{-y} y^\alpha\right)\Big|_0^\infty - \int_0^\infty (-e^{-y}) \, (y^\alpha)' \, dy$$

$$= 0 + \int_0^\infty e^{-y} \alpha y^{\alpha-1} dy = \alpha\Gamma(\alpha).$$

Consequently, $\Gamma(k) = (k-1)!$ for all integer $k > 0$ as

$\Gamma(1) = \int_0^\infty e^{-y} dy = 1 = 0!$ and recursively,

$$\Gamma(k) = (k-1) \cdot \Gamma(k-1) = (k-1)(k-2) \cdot \Gamma(k-2) = \ldots = (k-1)(k-2) \cdot \ldots \cdot 1 \cdot \Gamma(1)$$

$$= (k-1)! \cdot \Gamma(1) = (k-1)!$$

Gamma function and gamma distribution.

The gamma function $\Gamma(\alpha)$ is defined as

$$\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1} dx$$

for all real $\alpha > 0$. Integration by parts yields

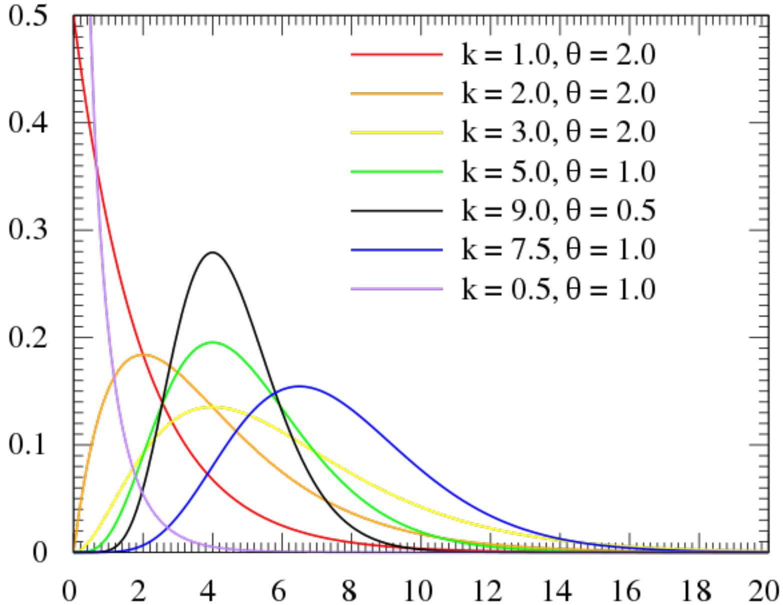
$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha).$$

In particular, $\Gamma(k) = (k-1)!$ for all integer $k > 0$.

For given $\alpha, \lambda > 0$, a gamma distributed random variable (aka **gamma random variable**) with parameters (α, λ) is defined by its probability density function

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} \lambda e^{-\lambda x} (\lambda x)^{\alpha-1} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0. \end{cases}$$

Gamma distribution.



Source: Wikipedia.org

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} \lambda e^{-\lambda x} (\lambda x)^{\alpha - 1} & \text{for } x \geq 0 \\ 0 & \text{for } x < 0. \end{cases}$$

Beta function and beta random variable.

The beta function $\mathcal{B}(a, b)$ is defined as

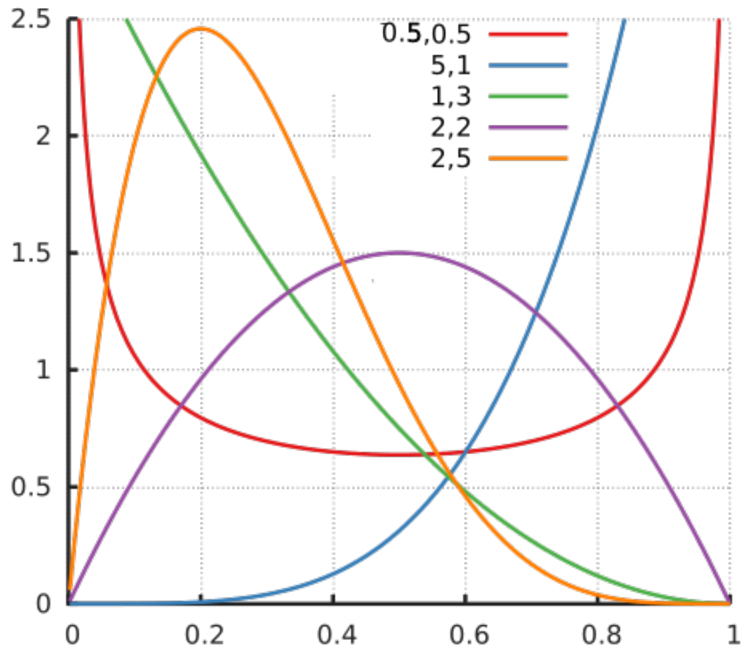
$$\mathcal{B}(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1}dx$$

for all $a > 0$ and $b > 0$.

For given $a, b > 0$, a beta distributed random variable (aka **beta random variable**) with parameters (a, b) is defined by its probability density function

$$f(x) = \begin{cases} \frac{1}{\mathcal{B}(a,b)}\, x^{a-1}(1-x)^{b-1} & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Beta distribution.



Source: Wikipedia.org

$$f(x) = \begin{cases} \frac{1}{\mathcal{B}(a,b)}\, x^{a-1}(1-x)^{b-1} & \text{for } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Sums of independent random variables.

• **Example.** Let X and Y be independent gamma random variables with the respective parameters (α, λ) and (β, λ). Their sum will have the following continuous distribution:

$$f_{\mathsf{x+y}}(a) = \int_{-\infty}^{\infty} f_{\mathsf{x}}(x)\, f_{\mathsf{y}}(a-x)\, dx = \int_0^a \frac{1}{\Gamma(\alpha)} \lambda e^{-\lambda x} (\lambda x)^{\alpha-1} \frac{1}{\Gamma(\beta)} \lambda e^{-\lambda(a-x)} \big(\lambda(a-x)\big)^{\beta-1} dx$$

$$= \frac{1}{\Gamma(\alpha+\beta)} \lambda e^{-\lambda a} (\lambda a)^{\alpha+\beta-1} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^a \left(\frac{x}{a}\right)^{\alpha-1} \left(\frac{a-x}{a}\right)^{\beta-1} \frac{dx}{a}$$

$$= \frac{1}{\Gamma(\alpha+\beta)} \lambda e^{-\lambda a} (\lambda a)^{\alpha+\beta-1} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 y^{\alpha-1} (1-y)^{\beta-1}\, dy = \frac{1}{\Gamma(\alpha+\beta)} \lambda e^{-\lambda a} (\lambda a)^{\alpha+\beta-1}.$$

Hence, $X + Y$ is a gamma random variable with parameters $(\alpha+\beta, \lambda)$.

Note: The above computations yield $\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 x^{a-1}(1-x)^{b-1} dx$.

Sums of independent random variables.

• **Example.** Let $X_1, X_2, \ldots$ be independent exponential random variables with parameter $\lambda > 0$. Then

$$T_n = \sum_{k=1}^{n} X_k \quad \text{for} \quad n = 1, 2, \ldots$$

is a gamma random variable with parameters (n, λ).

Indeed, each X_j is a gamma random variable with parameters $(1, \lambda)$.

Thus, $X_1 + X_2$ is a gamma random variable with parameters $(2, \lambda)$.

Hence, $X_1 + X_2 + X_3$ is a gamma random variable with parameters $(3, \lambda)$.

Next, $X_1 + X_2 + X_3 + X_4$ is a gamma random variable with parameters $(4, \lambda)$.

And so on.

Recursively (via induction), $T_n = X_1 + X_2 + \ldots + X_n$ is a gamma random variable with parameters (n, λ) with density function

$$f(x) = \frac{1}{(n-1)!} \lambda e^{-\lambda x} (\lambda x)^{n-1} \quad \text{for } x \geq 0.$$

Poisson process.

Let $X_1, X_2, \ldots$ be independent exponential random variables with parameter $\lambda > 0$.

- **Arrival times:** Let $T_0 = 0$ and $T_n = \sum_{k=1}^{n} X_k$ for $n = 1, 2, \ldots$.

- **Interarrival times:** $X_n = T_n - T_{n-1}$

- Poisson process with intensity λ is defined as

$$N(t) = \max\{n \geq 0 : T_n \leq t\} \quad (t \geq 0).$$

Here, $N(t)$ counts the number of arrivals between 0 and t.

- The increment $N(t_0+L)-N(t_0)$ counts the number of arrivals between t_0 and $t_0 + L$.

- Because of memorylessness property of exponential random variables X_j, the increment $N(t_0+L)-N(t_0)$ is distributed as $N(L)$.

- $N(t_0+L)-N(t_0)$ is a Poisson random variable with parameter λL:

$$P\Big(N(t_0 + L) - N(t_0) = k\Big) = e^{-\lambda L} \frac{(\lambda L)^k}{k!} \quad (k = 0, 1, \ldots).$$

Poisson process.

$$P\Big(N(t_0+L)-N(t_0)=k\Big)=e^{-\lambda L}\frac{(\lambda L)^k}{k!}\quad (k=0,1,\ldots).$$

Proof: Recall $N(t_0+L)-N(t_0)$ is distributed as $N(L)$.

Now, since

$$\{T_k \le L\} = \{T_k \le L < T_{k+1}\} \cup \{T_{k+1} \le L\},$$

$$P(N(L)=k) = P(T_k \le L < T_{k+1}) = P(T_k \le L) - P(T_{k+1} \le L)$$

Since $T_k = X_1 + \ldots + X_k$ is a gamma random variable with parameters (k,λ),

$$P(N(L)=k) = \int_0^L \frac{\lambda^k x^{k-1}}{(k-1)!}e^{-\lambda x}dx - \int_0^L \frac{\lambda^{k+1}x^k}{k!}e^{-\lambda x}dx$$

Integration by parts:

$$\int_0^L \frac{x^{k-1}}{(k-1)!}e^{-\lambda x}dx = e^{-\lambda x}\frac{x^k}{k!}\Big|_{x=0}^{L} + \int_0^L \frac{x^k}{k!}\lambda e^{-\lambda x}dx$$

Plugging in: $\quad P(N(L)=k) = e^{-\lambda L}\dfrac{(\lambda L)^k}{k!}$

Marginal distributions from joint distribution.

• Suppose X and Y are discrete random variables with joint probability mass function $p(x, y)$ and marginal probability mass functions p_{x} and p_{y} respectively. Then

$$p_{\mathsf{x}}(x) = \sum_{y:\, p(x,y)>0} p(x, y) \quad \text{and} \quad p_{\mathsf{y}}(y) = \sum_{x:\, p(x,y)>0} p(x, y)$$

Indeed, for all x,

$$p_{\mathsf{x}}(x) = P(X = x) = \sum_y P\Big(\{X = x\} \cap \{Y = y\}\Big) = \sum_y p(x, y).$$

Similarly, for all y,

$$p_{\mathsf{y}}(y) = P(Y = y) = \sum_x P\Big(\{X = x\} \cap \{Y = y\}\Big) = \sum_x p(x, y).$$

Marginal distributions from joint distribution.

• Suppose X and Y are continuous random variables with joint probability density function $f(x,y)$ and marginal probability density functions f_{x} and f_{y} respectively. Then

$$f_{\mathsf{x}}(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad f_{\mathsf{y}}(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

Indeed, for all a, the cumulative distribution function

$$F_{\mathsf{x}}(a) = P(X \le a) = P\Big(\{X \le a\} \cap \{-\infty < Y < \infty\}\Big) = \int_{-\infty}^{a} \int_{-\infty}^{\infty} f(x,y)\,dy\,dx.$$

Thus, by the Fundamental Theorem of Calculus, we have

$$f_{\mathsf{x}}(a) = \frac{d}{da} F_{\mathsf{x}}(a) = \frac{d}{da} \int_{-\infty}^{a} \int_{-\infty}^{\infty} f(x,y)\,dy\,dx = \int_{-\infty}^{\infty} f(a,y)\,dy.$$

Similarly, for all a,

$$f_{\mathsf{y}}(a) = \frac{d}{da} F_{\mathsf{y}}(a) = \frac{d}{da} \int_{-\infty}^{a} \int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_{-\infty}^{\infty} f(x,a)\,dx.$$

Marginal distributions from joint distribution.

• Suppose X and Y are discrete random variables with joint probability mass function $p(x, y)$ and marginal probability mass functions p_{x} and p_{y} respectively. Then

$$p_{\mathsf{x}}(x) = \sum_{y:\, p(x,y)>0} p(x, y) \quad \text{and} \quad p_{\mathsf{y}}(y) = \sum_{x:\, p(x,y)>0} p(x, y)$$

• **Example.** Consider a joint probability mass function

$$p(1, 2) = \frac{1}{4}, \quad p(2, 0) = \frac{1}{6}, \quad p(2, 3) = \frac{1}{4}, \quad p(3, 3) = \frac{1}{3}$$

Then,

$$p_{\mathsf{x}}(1) = p(1, 2) = \frac{1}{4}, \qquad p_{\mathsf{x}}(2) = p(2, 0) + p(2, 3) = \frac{5}{12}, \qquad p_{\mathsf{x}}(3) = p(3, 3) = \frac{1}{3}$$

and

$$p_{\mathsf{y}}(0) = p(2, 0) = \frac{1}{6}, \qquad p_{\mathsf{y}}(2) = p(1, 2) = \frac{1}{4}, \qquad p_{\mathsf{y}}(3) = p(2, 3) + p(3, 3) = \frac{7}{12}$$

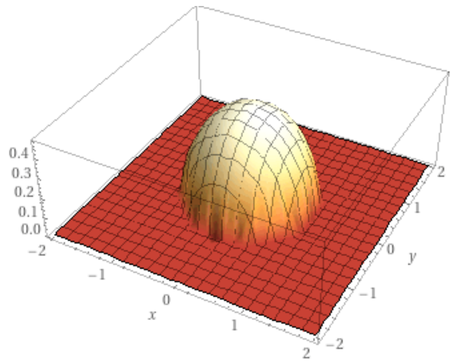
Marginal distributions from joint distribution.

• Suppose X and Y are continuous random variables with joint probability density function $f(x,y)$ and marginal probability density functions f_{x} and f_{y} respectively. Then

$$f_{\mathsf{x}}(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad f_{\mathsf{y}}(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

• **Example.** Consider the joint probability density function

$$f(x,y) = \begin{cases} \frac{3}{2\pi}\sqrt{1-x^2-y^2} & \text{if } x^2+y^2 \le 1, \\ 0 & \text{if } x^2+y^2 > 1. \end{cases}$$

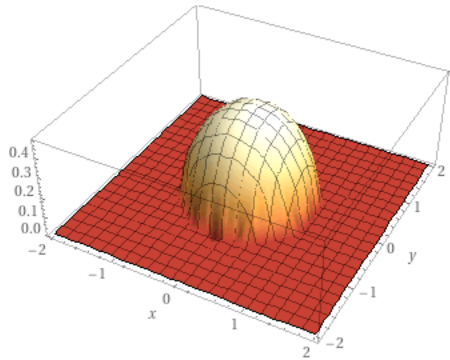


Then, for a given $x \in (-1,1)$,

$$f_{\mathsf{x}}(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \frac{3}{2\pi}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2}\,dy$$

• **Example (Cont.).**

$$f(x,y) = \begin{cases} \frac{3}{2\pi}\sqrt{1-x^2-y^2} & \text{if } x^2+y^2 \le 1, \\ 0 & \text{if } x^2+y^2 > 1. \end{cases}$$



Then, for a given $x \in (-1,1)$,

$$f_{\mathsf{x}}(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \frac{3}{2\pi}\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} \sqrt{1-x^2-y^2}\,dy = \frac{3}{2\pi}(1-x^2)\int_{-1}^{1}\sqrt{1-z^2}\,dz$$

after substituting $z = y/\sqrt{1-x^2}$. Thus, since $\int_{-1}^{1}\sqrt{1-z^2}\,dz = \frac{\pi}{2}$ is the area of the unit semicircle, we have

$$f_{\mathsf{x}}(x) = \frac{3}{4}(1-x^2) \quad \text{for} \quad x \in (-1,1).$$

Similarly, $f_{\mathsf{y}}(y) = \frac{3}{4}(1-y^2)$ for $y \in (-1,1)$. Are X and Y independent?

Functions of random variables.

Recall:

• **Theorem.** Let X be a discrete random variable characterized by its probability mass function $p_{\mathsf{x}}(x)$. Then, for any real valued function g, $g(X)$ will also be a **random variable**, and

$$E[g(X)] = \sum_{x:\, p_X(x)>0} g(x)\, p_{\mathsf{x}}(x)$$

• **Theorem.** Let X be a continuous random variable characterized by its probability density function $f(x)$. Then, for any real valued function g, $g(X)$ will also be a **random variable**, and

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\, dx$$

Functions of random variables.

Similarly to 1D random variables, we have the following results for random vectors.

• **Theorem.** Let X and Y be discrete random variables with joint probability mass function $p(x,y)$. Then, for any real valued function $g : \mathbb{R}^2 \to \mathbb{R}$, $g(X,Y)$ will also be a **random variable**, and

$$E[g(X,Y)] = \sum_{(x,y):\, p(x,y)>0} g(x,y)\, p(x,y)$$

• **Theorem.** Let X and Y be continuous random variables with joint probability density function $f(x,y)$. Then, for any real valued function $g : \mathbb{R}^2 \to \mathbb{R}$, $g(X,Y)$ will also be a **random variable**, and

$$E[g(X,Y)] = \iint_{\mathbb{R}^2} g(x,y) f(x,y)\, dx\, dy$$

Similar results hold for n discrete or continuous random variables.

Functions of random variables.

• **Theorem.** Let X and Y be discrete random variables with joint probability mass function $p(x,y)$. Then, for any real valued function $g : \mathbb{R}^2 \to \mathbb{R}$, $g(X,Y)$ will also be a **random variable**, and

$$E[g(X,Y)] = \sum_{(x,y):\, p(x,y)>0} g(x,y)\, p(x,y)$$

Proof: Let $Z = g(X,Y)$. We find the probability mass function $p_{\mathsf{z}}(z)$ of Z:

$$p_{\mathsf{z}}(z) = P\Big(g(X,Y) = z\Big) = \sum_{(x,y):\, g(x,y)=z} P\Big(\{X = x\} \cap \{Y = y\}\Big) = \sum_{(x,y):\, g(x,y)=z} p(x,y)$$

as $\Big\{g(X,Y) = z\Big\} = \bigcup_{(x,y):\, g(x,y)=z} \Big(\{X = x\} \cap \{Y = y\}\Big)$ is a union of disjoint events.

Thus, $$E[Z] = \sum_z z\, p_{\mathsf{z}}(z) = \sum_z \left(z \sum_{(x,y):\, g(x,y)=z} p(x,y) \right) = \sum_z \left(\sum_{(x,y):\, g(x,y)=z} z\, p(x,y) \right)$$

$$= \sum_z \left(\sum_{(x,y):\, g(x,y)=z} g(x,y)\, p(x,y) \right) = \sum_{(x,y):\, p(x,y)>0} g(x,y)\, p(x,y)$$

Sums of random variables.

• **Theorem.** Expectation of a sum of random variables is equal to the sum of expectations.

$$E[X+Y] = E[X] + E[Y]$$

Notice: we don't require X and Y to be independent here. The above is true even if they are dependent random variables.

Proof: Suppose X and Y are discrete random variables. Let $g(x,y) = x+y$. Then,

$$E[g(X,Y)] = \sum_{(x,y):\, p(x,y)>0} g(x,y)\,p(x,y) = \sum_{(x,y):\, p(x,y)>0} (x+y)\,p(x,y)$$

$$= \sum_{(x,y):p(x,y)>0} x\,p(x,y) + \sum_{(x,y):\, p(x,y)>0} y\,p(x,y) = \sum_{x:\, p_{\mathsf{x}}(x)>0} x \left(\sum_{y:\, p(x,y)>0} p(x,y) \right) + \sum_{y:\, p_{\mathsf{y}}(y)>0} y \left(\sum_{x:\, p(x,y)>0} p(x,y) \right)$$

$$= \sum_{x:\, p_{\mathsf{x}}(x)>0} x\,p_{\mathsf{x}}(x) + \sum_{y:\, p_{\mathsf{y}}(y)>0} y\,p_{\mathsf{y}}(y) = E[X] + E[Y].$$

Sums of random variables.

• **Theorem.** Expectation of a sum of random variables is equal to the sum of expectations.

$$E[X+Y] = E[X] + E[Y]$$

Notice: we don't require X and Y to be independent here. The above is true even if they are dependent random variables.

Proof: Suppose X and Y are continuous random variables. Let $g(x,y) = x + y$. Then,

$$E[g(X,Y)] = \iint\limits_{\mathbb{R}^2} g(x,y) f(x,y)\,dx\,dy \;=\iint\limits_{\mathbb{R}^2} (x+y)\,f(x,y)\,dx\,dy$$

$$=\iint\limits_{\mathbb{R}^2} x\,f(x,y)\,dy\,dx + \iint\limits_{\mathbb{R}^2} y\,f(x,y)\,dx\,dy = \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x,y)dy\right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f(x,y)dx\right) dy$$

$$= \int_{-\infty}^{\infty} x\,f_{\mathsf{x}}(x)\,dx + \int_{-\infty}^{\infty} y\,f_{\mathsf{y}}(y)\,dy \;= E[X] + E[Y].$$

Sums of random variables.

• **Theorem.** Expectation of a sum of random variables is equal to the sum of expectations.

$$E[X+Y] = E[X] + E[Y]$$

Consequently, for n random variables, we have

$$E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n]$$

• Given $0 \leq p \leq 1$. We say that X is a Bernoulli random variable with parameter p if

$$P(X=1) = p \quad \text{and} \quad P(X=0) = 1-p$$

There $E[X] = p$ and $Var(X) = p(1-p)$.

Let $X_1,\ X_2, \ldots,\ X_n$ be independent Bernoulli random variables with the same parameter p. Then

$$S_n = X_1 + X_2 + \ldots + X_n$$

is a Binomial random variable with parameters (n, p). Then

$$E[S_n] = E[X_1 + \ldots + X_n] = E[X_1] + \ldots + E[X_n] = np.$$

Covariance and correlation.

Consider random variables X and Y with finite means μ_x and μ_y and finite variances $\sigma_x^2 > 0$ and $\sigma_y^2 > 0$ respectively.

- **Definition.** The covariance of X and Y is

$$Cov(X,Y) = E\Big[(X-\mu_x)(Y-\mu_y)\Big].$$

The correlation of X and Y is

$$\text{corr}(X,Y) = \frac{Cov(X,Y)}{\sigma_x\,\sigma_y}.$$

Properties:

- $Cov(X,X) = E\Big[(X-\mu_x)^2\Big] = Var(X).$

- **Lemma.** The covariance can be expressed as follows:

$$Cov(X,Y) = E[XY] - E[X]E[Y] \;= E[XY] - \mu_x\,\mu_y.$$

Proof:

$$Cov(X,Y) = E\Big[(X-\mu_x)(Y-\mu_y)\Big] = E\Big[XY - \mu_x Y - \mu_y X + \mu_x\mu_y\Big]$$

$$= E[XY] - \mu_x E[Y] - \mu_y E[X] + \mu_x\mu_y = E[XY] - \mu_x\,\mu_y$$

Covariance and correlation.

- **Example.** Consider a joint probability mass function

$$p(1,2) = \frac{1}{4}, \quad p(2,0) = \frac{1}{6}, \quad p(2,3) = \frac{1}{4}, \quad p(3,3) = \frac{1}{3}$$

Then,

$$p_{\mathsf{x}}(1) = p(1,2) = \frac{1}{4}, \qquad p_{\mathsf{x}}(2) = p(2,0) + p(2,3) = \frac{5}{12}, \qquad p_{\mathsf{x}}(3) = p(3,3) = \frac{1}{3}$$

and

$$p_{\mathsf{y}}(0) = p(2,0) = \frac{1}{6}, \qquad p_{\mathsf{y}}(2) = p(1,2) = \frac{1}{4}, \qquad p_{\mathsf{y}}(3) = p(2,3) + p(3,3) = \frac{7}{12}$$

Hence, $E[X] = 1\cdot\frac{1}{4}+2\cdot\frac{5}{12}+3\cdot\frac{1}{3} = \frac{25}{12}$, $\qquad E[Y] = 0\cdot\frac{1}{6}+2\cdot\frac{1}{4}+3\cdot\frac{7}{12} = \frac{9}{4}$, $\qquad$ and

$$E[XY] = 1\cdot 2\cdot p(1,2)+2\cdot 0\cdot p(2,0)+2\cdot 3\cdot p(2,3)+3\cdot 3\cdot p(3,3) = \frac{1}{2}+0+\frac{3}{2}+3 = 5.$$

Thus, $Cov(X,Y) = E[XY] - E[X]\,E[Y] = 5-\frac{25}{12}\cdot\frac{9}{4} = 5-\frac{75}{16} = \frac{5}{16} = 0.3125.$

Covariance and correlation.

- If X and Y are independent then

$$E[h(X)g(Y)] = E[h(X)]\,E[g(Y)]$$

for any pair of functions, h and g for which the above expectations exist and are finite.

Proof: Suppose X and Y are independent continuous random variables with density functions f_{x} and f_{y}. Then, the joint probability density function equals $f(x,y) = f_{\mathsf{x}}(x)f_{\mathsf{y}}(y)$, and

$$E[h(X)g(Y)] = \iint_{\mathbb{R}^2} h(x)g(y)f(x,y)\,dx\,dy = \iint_{\mathbb{R}^2} h(x)g(y)f_{\mathsf{x}}(x)f_{\mathsf{y}}(y)\,dx\,dy$$

$$= \int_{-\infty}^{\infty} g(y)\left(\int_{-\infty}^{\infty} h(x)f_{\mathsf{x}}(x)\,dx\right) f_{\mathsf{y}}(y)\,dy = \int_{-\infty}^{\infty} h(x)f_{\mathsf{x}}(x)\,dx \cdot \int_{-\infty}^{\infty} g(y)f_{\mathsf{y}}(y)\,dy$$

$$= E[h(X)]\,E[g(Y)]$$

The case of independent discrete random variables is proved similarly.

Covariance and correlation.

• **Definition.** The covariance of X and Y is

$$Cov(X,Y) = E\Big[(X - \mu_{\mathsf{x}})(Y - \mu_{\mathsf{y}})\Big].$$

The correlation of X and Y is

$$\text{corr}(X,Y) = \frac{Cov(X,Y)}{\sigma_{\mathsf{x}}\,\sigma_{\mathsf{y}}}.$$

Properties:

• $Cov(X,X) = Var(X)$.

• $Cov(X,Y) = E[XY] - E[X]E[Y] \;= E[XY] - \mu_{\mathsf{x}}\,\mu_{\mathsf{y}}$.

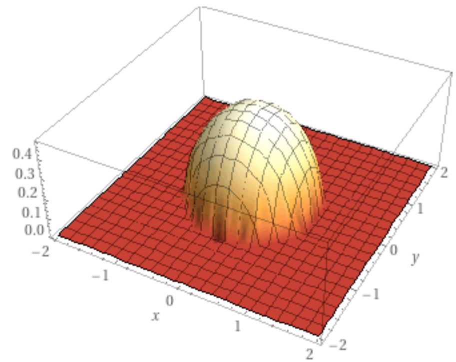
• If X and Y are independent, then $Cov(X,Y) = 0$.

Proof: $Cov(X,Y) = E[XY] - E[X]E[Y] = E[X]E[Y] - E[X]E[Y] = 0$.

So, X and Y independent $\Rightarrow$ X and Y uncorrelated.

Covariance and correlation.

- **Example.**



$$f(x,y) = \begin{cases} \frac{3}{2\pi}\sqrt{1-x^2-y^2} & \text{if } x^2+y^2 \le 1, \\ 0 & \text{if } x^2+y^2 > 1. \end{cases}$$

We obtained

$$f_X(x) = f_Y(x) = \frac{3}{4}(1-x^2) \quad \text{for } x \in (-1,1) \quad \text{implying} \quad E[X] = E[Y] = 0.$$

$$E[XY] = \frac{3}{2\pi} \iint_{x^2+y^2\le 1} xy\sqrt{1-x^2-y^2}\,dx\,dy = \frac{3}{2\pi}\int_{-1}^{1}\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} xy\sqrt{1-x^2-y^2}\,dx\,dy$$

$$= \frac{3}{2\pi}\int_{-1}^{1} y \int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} x\sqrt{1-x^2-y^2}\,dx\,dy = \frac{3}{2\pi}\int_{-1}^{1} y\cdot 0\,dy = 0.$$

Thus, $Cov(X,Y) = E[XY] - E[X]\,E[Y] = 0$. So, X and Y are dependent and uncorrelated at the same time.

Covariance and correlation.

Other simple properties: • $Cov(X, Y) = Cov(Y, X)$.

- $Cov(aX, Y) = a\,Cov(X, Y)$ for all $a \in \mathbb{R}$.
- $Cov\left(\sum_{i=1}^{n} X_i\,,\ \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$.

Proof:

$$Cov\left(\sum_{i=1}^{n} X_i\,,\ \sum_{j=1}^{m} Y_j\right) = E\left[\sum_{i=1}^{n} X_i \cdot \sum_{j=1}^{m} Y_j\right] - E\left[\sum_{i=1}^{n} X_i\right] \cdot E\left[\sum_{j=1}^{m} Y_j\right]$$

$$= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m} X_i Y_j\right] - \sum_{i=1}^{n} E[X_i] \cdot \sum_{j=1}^{m} E[Y_j] = \sum_{i=1}^{n}\sum_{j=1}^{m} E[X_i Y_j] - \sum_{i=1}^{n}\sum_{j=1}^{m} E[X_i]E[Y_j]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} \big(E[X_i Y_j] - E[X_i]E[Y_j]\big) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j).$$

Covariance and correlation.

The covariance is bilinear:

- $Cov(X, Y) = Cov(Y, X)$.
- $Cov(aX, Y) = a\,Cov(X, Y)$ for all $a \in \mathbb{R}$.
- $Cov\left(\sum_{i=1}^{n} X_i\,,\, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$.

In linear algebra, the inner (dot) product $\langle \mathbf{x}, \mathbf{y} \rangle$ is also bilinear:

- $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$.
- $\langle a\mathbf{x}, \mathbf{y} \rangle = a\,\langle \mathbf{x}, \mathbf{y} \rangle$ for all $a \in \mathbb{R}$.
- $\left\langle \sum_{i=1}^{n} \mathbf{x}_i\,,\, \sum_{j=1}^{m} \mathbf{y}_j \right\rangle = \sum_{i=1}^{n}\sum_{j=1}^{m} \langle \mathbf{x}_i, \mathbf{y}_j \rangle$.

Compare $\langle \mathbf{x}, \mathbf{x} \rangle = |\mathbf{x}|^2$ to $Cov(X, X) = Var(X) = \sigma_{\mathsf{x}}^2$.

Covariance and correlation.

Other simple properties: • $Cov(X,Y) = Cov(Y,X)$.

- $Cov(aX,Y) = a\,Cov(X,Y)$ for all $a \in \mathbb{R}$.
- $Cov\left(\sum_{i=1}^{n} X_i\,,\ \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i,Y_j)$.
- $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i,X_j) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i,j:\, i<j} Cov(X_i,X_j)$.

Proof:

$$Var\left(\sum_{i=1}^{n} X_i\right) = Cov\left(\sum_{i=1}^{n} X_i\,,\ \sum_{j=1}^{n} X_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i,X_j).$$

In particular, $Var(X+Y) = Var(X) + Var(Y) + 2\,Cov(X,Y)$.

Covariance and correlation.

Other simple properties: • $Cov(X, Y) = Cov(Y, X)$.

• $Cov(aX, Y) = a\,Cov(X, Y)$ for all $a \in \mathbb{R}$.

• $Cov\left(\sum_{i=1}^{n} X_i\,, \sum_{j=1}^{m} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} Cov(X_i, Y_j)$.

• $Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) = \sum_{i=1}^{n} Var(X_i) + 2 \sum_{i,j:\, i<j} Cov(X_i, X_j)$.

In particular, $Var(X + Y) = Var(X) + Var(Y) + 2\,Cov(X, Y)$.

• If $X_1, X_2, \ldots, X_n$ are independent, then

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

as $Cov(X_i, X_j) = 0$ for all $i \neq j$.

Covariance and correlation.

Observe that

$$0 \leq Var(X-\alpha Y) = Var(X)+\alpha^2 Var(Y)-2\alpha Cov(X,Y) \quad \text{for all } \alpha > 0.$$

Thus,

$$Cov(X,Y) \leq \frac{1}{2}\alpha^{-1} Var(X) + \frac{1}{2}\alpha Var(Y).$$

Recall that $Var(X) = \sigma_x^2$ and $Var(Y) = \sigma_y^2$.

Next, we substitute $\alpha = \frac{\sigma_x}{\sigma_y}$, obtaining

$$Cov(X,Y) \leq \frac{1}{2}\frac{\sigma_y}{\sigma_x} Var(X) + \frac{1}{2}\frac{\sigma_x}{\sigma_y} Var(Y) = \sigma_x\sigma_y.$$

Therefore, $-Cov(X,Y) = Cov(-X,Y) \leq \sigma_x\sigma_y$ as $Var(-X) = Var(X) = \sigma_x^2$.

Thus, we proved the following inequality

$$\left|Cov(X,Y)\right| \leq \sigma_x\sigma_y = \sqrt{Var(X)\,Var(Y)}$$

which is a probabilistic version of the **Cauchy-Bunyakovsky-Schwarz inequality**.

Covariance and correlation.

We proved the following inequality

$$\Big|Cov(X,Y)\Big| \le \sigma_x\sigma_y = \sqrt{Var(X)\,Var(Y)}$$

which is a probabilistic version of the **Cauchy-Bunyakovsky-Schwarz inequality**.

Recall $\Big|\langle \mathbf{x},\mathbf{y}\rangle\Big| \le |\mathbf{x}|\,|\mathbf{y}|$ in linear algebra.

• The correlation of X and Y is measuring statistical association of X and Y on the scale from -1 to 1:

$$-1 \le \mathrm{corr}(X,Y) = \frac{Cov(X,Y)}{\sigma_x\,\sigma_y} \le 1.$$

Proof:

$$\Big|\mathrm{corr}(X,Y)\Big| = \frac{\Big|Cov(X,Y)\Big|}{\sigma_x\,\sigma_y} \le 1.$$

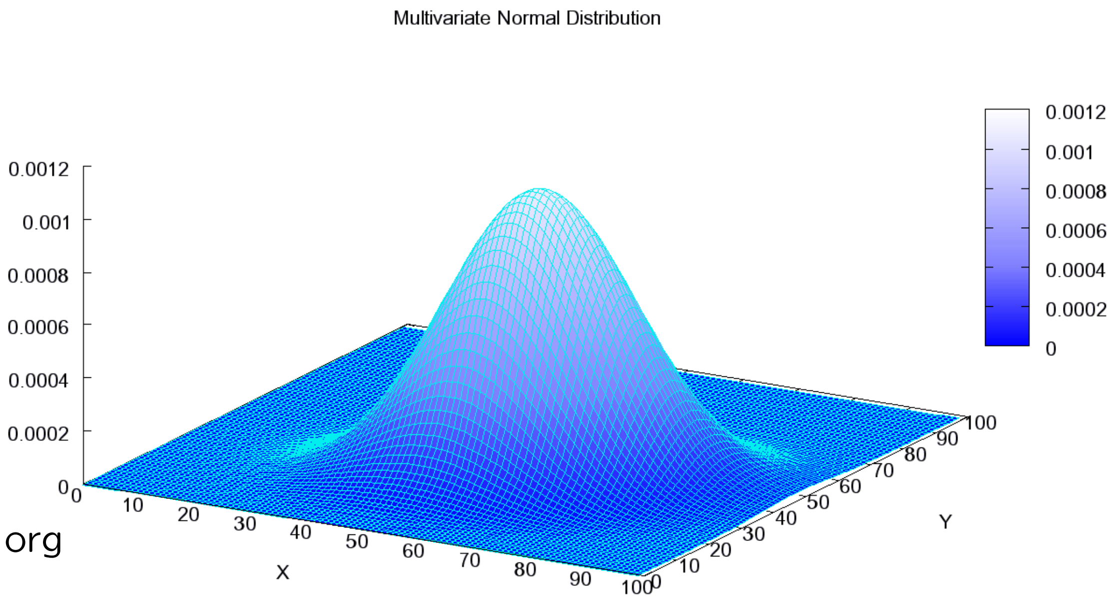
Compare to linear algebra:

$$-1 \le \cos\theta = \frac{\langle \mathbf{x},\mathbf{y}\rangle}{|\mathbf{x}|\,|\mathbf{y}|} \le 1,$$

where θ is the angle between $\mathbf{x}$ and $\mathbf{y}$.

Multivariate normal distribution.

$$\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$$



Source: Wikipedia.org

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \text{ where } \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \ \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

and $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A satisfying $\det(A) \neq 0$.

Multivariate normal distribution.

• **Example.** Consider independent standard normal random variables X and Y. Their marginal probability density functions are the same:

$$f_{\mathsf{x}}(x) = f_{\mathsf{y}}(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}$$

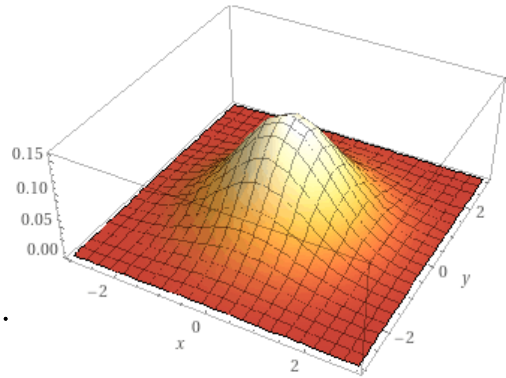
By independence, the joint probability density function

$$f(x,y) = f_{\mathsf{x}}(x) f_{\mathsf{y}}(y) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(x^2+y^2)\right\}$$

• Let $\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$, then $\Sigma^{-1} = I$, and

$$x^2 + y^2 = [x \;\; y] \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{x}^{\mathsf{T}} \Sigma^{-1} \mathbf{x},$$

where $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$.



Hence, $f(\mathbf{x}) = f(x,y) = \dfrac{1}{2\pi} \exp\left\{-\dfrac{1}{2}\mathbf{x}^{\mathsf{T}} \Sigma^{-1} \mathbf{x}\right\}$.

Multivariate normal distribution.

• **Example.** Consider independent normal random variables X and Y with respective parameters (μ_x, σ_x^2) and (μ_y, σ_y^2). Their marginal probability density functions are

$$f_x(x) = \frac{1}{\sqrt{2\pi\,\sigma_x^2}} \exp\left\{-\frac{1}{2\sigma_x^2}(x-\mu_x)^2\right\} \text{ and } f_y(x) = \frac{1}{\sqrt{2\pi\,\sigma_y^2}} \exp\left\{-\frac{1}{2\sigma_y^2}(x-\mu_y)^2\right\}$$

By independence, the joint probability density function

$$f(x,y) = f_x(x) f_y(y) = \frac{1}{2\pi\,\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right\}$$

• Let $\Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix}$, then $\Sigma^{-1} = \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix}$, and

$$\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} = [x-\mu_x, y-\mu_y] \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix} \begin{bmatrix} x-\mu_x \\ y-\mu_y \end{bmatrix} = (\mathbf{x}-\mu)^{\mathsf{T}}\, \Sigma^{-1}\, (\mathbf{x}-\mu),$$

where $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$. Hence, the joint probability density function is

$$f(\mathbf{x}) = f(x,y) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\, \Sigma^{-1}\, (\mathbf{x}-\mu)\right\}.$$

Multivariate normal distribution.

The independence assumption is not necessary. For parameters

$-\infty < \mu_x < \infty, \quad -\infty < \mu_y < \infty, \quad 0 < \sigma_x, \quad 0 < \sigma_y,$ and $-1 < \rho < 1,$

consider a symmetric matrix

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\,\sigma_x\sigma_y \\ \rho\,\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$$

• **Definition.** Let $\begin{bmatrix} X \\ Y \end{bmatrix}$ be a vector of random variables distributed according to the joint probability density function

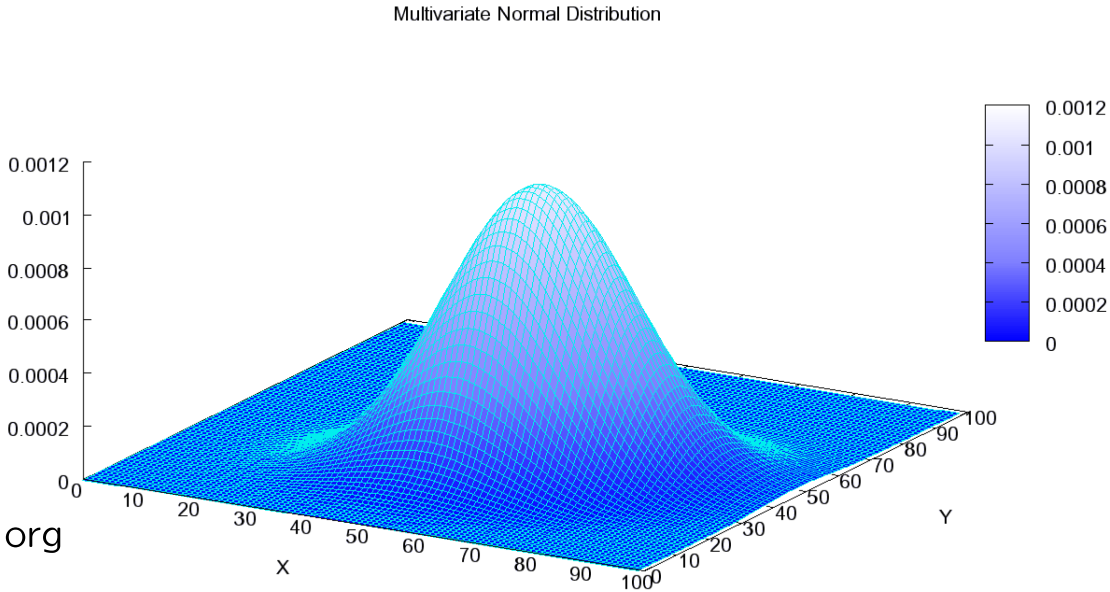
$$f(\mathbf{x}) = f(x, y) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\},$$

where $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ and $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$.

Then, $\begin{bmatrix} X \\ Y \end{bmatrix}$ is a bivariate normal. Notation: $\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$.

Multivariate normal distribution.

$$\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$$



Source: Wikipedia.org

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \text{ where } \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \ \mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

and $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A satisfying $\det(A) \neq 0$.

Multivariate normal distribution.

The random vector $\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$ is multivariate normal if the joint density function

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}} \Sigma^{-1} (\mathbf{x}-\mu)\right\}, \qquad \mathbf{x} \in \mathbb{R}^n,$$

where $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \in \mathbb{R}^n$ is the vector of mean values ($E[Z_i] = \mu_i$), and Σ is a positive definite real matrix:

(i.) Σ is symmetric: $\Sigma^{\mathsf{T}} = \Sigma$; **(ii.)** $\mathbf{x}^{\mathsf{T}}\Sigma\mathbf{x} > 0$ for all $\mathbf{x} \neq 0$.

As a positive definite matrix, Σ can be represented as follows: $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A satisfying $\det(A) \neq 0$.

Next, we check that

$$\int \cdots \int_{\mathbb{R}^n} f(\mathbf{x})\, dx_1 \ldots dx_n = 1.$$

Multivariate normal distribution.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \quad \mathbf{x} \in \mathbb{R}^n,$$

where $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \in \mathbb{R}^n$, and $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A satisfying $\det(A) \neq 0$. Observe that if we let $\mathbf{y} = A^{-1}(\mathbf{x}-\mu)$ then

$$\frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} = \frac{1}{(2\pi)^{n/2}|\det(A)|} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right\}$$

and

$$\int_{\mathbb{R}^n}\!\!\cdots\!\int f(\mathbf{x})dx_1\ldots dx_n = \int_{\mathbb{R}^n}\!\!\cdots\!\int \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} dx_1\ldots dx_n$$

$$= \int_{\mathbb{R}^n}\!\!\cdots\!\int \frac{1}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right\} dy_1\ldots dy_n = \prod_{i=1}^{n}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y_i^2\right\} dy_i = 1$$

as $dy_1\ldots dy_n = |J|\,dx_1\ldots dx_n = |\det(A^{-1})|dx_1\ldots dx_n = \frac{1}{|\det(A)|}dx_1\ldots dx_n$.

Multivariate normal distribution.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \quad \mathbf{x} \in \mathbb{R}^n,$$

where $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \in \mathbb{R}^n$, and $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A

satisfying $\det(A) \neq 0$. Suppose $\begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$.

• **Claim:** Vector $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$ is the vector of mean values, i.e.,

$E[Z_i] = \mu_i$.

Proof. Observe that $x_i - \mu_i = \mathbf{e}_i^{\mathsf{T}}(\mathbf{x}-\mu)$ for $i = 1, \ldots, n$, and

$$E[Z_i] = \mu_i + E[Z_i - \mu_i] = \mu_i + \int_{\mathbb{R}^n} \ldots \int (x_i - \mu_i)\, f(\mathbf{x}) dx_1 \ldots dx_n$$

$$= \mu_i + \int_{\mathbb{R}^n} \ldots \int \frac{\mathbf{e}_i^{\mathsf{T}}(\mathbf{x}-\mu)}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} dx_1 \ldots dx_n$$

Multivariate normal distribution.

- **Claim:** $E[Z_i] = \mu_i$.

Proof (cont.): Let $\mathbf{y} = A^{-1}(\mathbf{x}-\mu)$ then $(\mathbf{x}-\mu) = A\mathbf{y}$ and

$$E[Z_i] = \mu_i + \int_{\mathbb{R}^n}\cdots\int \frac{\mathbf{e}_i^\mathsf{T}(\mathbf{x}-\mu)}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^\mathsf{T}\Sigma^{-1}(\mathbf{x}-\mu)\right\} dx_1\ldots dx_n$$

$$= \mu_i + \int_{\mathbb{R}^n}\cdots\int \frac{\mathbf{e}_i^\mathsf{T} A\mathbf{y}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^\mathsf{T}\mathbf{y}\right\} dy_1\ldots dy_n$$

$$= \mu_i + \mathbf{e}_i^\mathsf{T} A\left(\int_{\mathbb{R}^n}\cdots\int \frac{\mathbf{y}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^\mathsf{T}\mathbf{y}\right\} dy_1\ldots dy_n\right) = \mu_i$$

as $\int_{-\infty}^{\infty} y_j \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y_j^2\right\} dy_j = 0$ for all j, and therefore,

$$\int_{\mathbb{R}^n}\cdots\int \frac{\mathbf{y}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^\mathsf{T}\mathbf{y}\right\} dy_1\ldots dy_n = 0.$$

Multivariate normal distribution.

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}, \quad \mathbf{x}\in\mathbb{R}^n,$$

where $\mu = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \in \mathbb{R}^n$ is the vector of mean values ($E[Z_i] = \mu_i$),

and $\Sigma = AA^{\mathsf{T}}$ for some $n \times n$ matrix A satisfying $\det(A) \neq 0$.

- **Claim:** Matrix $\Sigma = \Big(\; Cov(Z_i, Z_j) \; \Big)$ is the covariance matrix.

Proof. Observe that $(x_i-\mu_i)(x_j-\mu_j) = \mathbf{e}_i^{\mathsf{T}}(\mathbf{x}-\mu)(\mathbf{x}-\mu)^{\mathsf{T}}\mathbf{e}_j$

$$Cov(Z_i,Z_j) = E\Big[(Z_i-\mu_i)(Z_j-\mu_j)\Big] = \int_{\mathbb{R}^n}\!\!\cdots\!\int (x_i-\mu_i)(x_j-\mu_j) f(\mathbf{x})\,dx_1\ldots dx_n$$

$$= \int_{\mathbb{R}^n}\!\!\cdots\!\int \frac{\mathbf{e}_i^{\mathsf{T}}(\mathbf{x}-\mu)(\mathbf{x}-\mu)^{\mathsf{T}}\mathbf{e}_j}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} dx_1\ldots dx_n$$

Multivariate normal distribution.

- **Claim:** Matrix $\Sigma = \Big(Cov(Z_i, Z_j) \Big)$ is the covariance matrix.

Proof (cont.): Let $\mathbf{y} = A^{-1}(\mathbf{x}-\mu)$ then $(\mathbf{x}-\mu) = A\mathbf{y}$ and

$$Cov(Z_i,Z_j) = \int_{\mathbb{R}^n} \cdots \int \frac{\mathbf{e}_i^{\mathsf{T}}(\mathbf{x}-\mu)(\mathbf{x}-\mu)^{\mathsf{T}}\mathbf{e}_j}{(2\pi)^{n/2}\sqrt{\det\Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\} dx_1 \dots dx_n$$

$$= \int_{\mathbb{R}^n} \cdots \int \frac{\mathbf{e}_i^{\mathsf{T}} A\mathbf{y}\mathbf{y}^{\mathsf{T}}A^{\mathsf{T}}\mathbf{e}_j}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right\} dy_1 \dots dy_n$$

$$= \mathbf{e}_i^{\mathsf{T}} A \left(\int_{\mathbb{R}^n} \cdots \int \frac{\mathbf{y}\mathbf{y}^{\mathsf{T}}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right\} dy_1 \dots dy_n \right) A^{\mathsf{T}}\mathbf{e}_j$$

$$= \mathbf{e}_i^{\mathsf{T}} AA^{\mathsf{T}}\mathbf{e}_j = \mathbf{e}_i^{\mathsf{T}}\Sigma\,\mathbf{e}_j = \Sigma_{i,j}$$

as $\int_{-\infty}^{\infty} y_i y_j \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y_i^2\right\} dy_i = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}$ and therefore,

$$\int_{\mathbb{R}^n} \cdots \int \frac{\mathbf{y}\mathbf{y}^{\mathsf{T}}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2}\mathbf{y}^{\mathsf{T}}\mathbf{y}\right\} dy_1 \dots dy_n = I.$$

Multivariate normal distribution.

- **Claim:** Matrix $\Sigma = \Big(Cov(Z_i, Z_j) \Big)$ is the covariance matrix.

- **Example.** Suppose $\begin{bmatrix} X \\ Y \end{bmatrix}$ is a bivariate normal with mean $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$ and with

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\,\sigma_x\sigma_y \\ \rho\,\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix},$$

i.e., $\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma)$. Then,

$$Var(X) = \sigma_x^2, \quad Var(Y) = \sigma_y^2, \quad \text{and} \quad Cov(X,Y) = \rho\,\sigma_x\sigma_y.$$

Therefore, the correlation

$$\text{corr}(X,Y) = \frac{Cov(X,Y)}{\sigma_x\,\sigma_y} = \rho.$$

Multivariate normal distribution.

- **Claim:** Matrix $\Sigma = \Big(Cov(Z_i, Z_j) \Big)$ is the covariance matrix.

- **Example.** Suppose $\begin{bmatrix} X \\ Y \end{bmatrix}$ is a bivariate normal with mean $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}$, $Var(X) = \sigma_x^2$, $Var(Y) = \sigma_y^2$, and with $Cov(X,Y) = 0$.

Then,

$$\Sigma = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{bmatrix},$$

$\Sigma^{-1} = \begin{bmatrix} \sigma_x^{-2} & 0 \\ 0 & \sigma_y^{-2} \end{bmatrix}$, and

$$f(x,y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left\{ -\frac{1}{2} (\mathbf{x} - \mu)^{\mathsf{T}} \Sigma^{-1} (\mathbf{x} - \mu) \right\}$$

$$= \frac{1}{2\pi\, \sigma_x \sigma_y} \exp\left\{ -\frac{1}{2} \left(\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right) \right\} = f_x(x) f_y(y)$$

We proved the following:
For **normal random variables**, X and Y uncorrelated $\Rightarrow$ X and Y independent.

Multivariate normal distribution.

- **Example.** Suppose $\begin{bmatrix} X \\ Y \end{bmatrix}$ is a bivariate normal with mean $\mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and with $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\,\sigma_x\sigma_y \\ \rho\,\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}$.
Then,

$$Var(X) = \sigma_x^2 = 1, \quad Var(Y) = \sigma_y^2 = 4, \quad \text{and} \quad Cov(X,Y) = \rho\,\sigma_x\sigma_y = -1.$$

Thus, the correlation $\text{corr}(X,Y) = \rho = -\frac{1}{2}$.

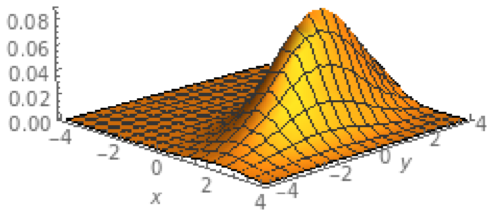
Next, $\Sigma^{-1} = \frac{1}{3}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix}$ and the joint probability density function is

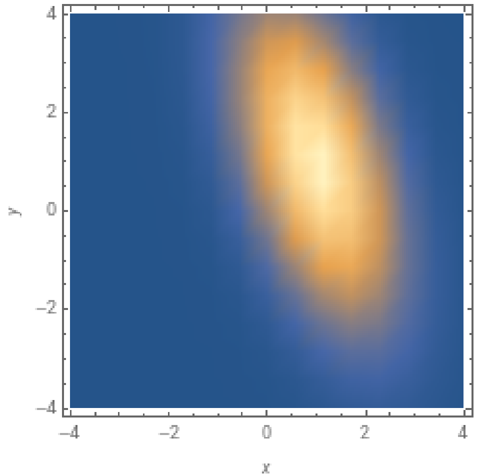
$$f(x,y) = \frac{1}{2\pi\sqrt{\det \Sigma}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\mu)^{\mathsf{T}}\Sigma^{-1}(\mathbf{x}-\mu)\right\}$$

$$= \frac{1}{2\sqrt{3}\pi} \exp\left\{-\frac{1}{6}\Big(4(x-1)^2 + 2(x-1)(y-1) + (y-1)^2\Big)\right\}.$$

Multivariate normal distribution.

- **Example (continued).**





https://demonstrations.wolfram.com/TheBivariateNormalDistribution/

$$f(x,y) = \frac{1}{2\sqrt{3}\pi} \exp\left\{-\frac{1}{6}\Big(4(x-1)^2 + 2(x-1)(y-1) + (y-1)^2\Big)\right\}.$$