4. Special Distributions

- $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B})$
- A discrete distribution is fully characterized by its probability function (pmf), $f_{\tilde{x}} : \tilde{x}(\Omega) \longrightarrow [0, 1]$, since

$$P_{\tilde{x}}(B) = P\{\tilde{x} \in B\} = \sum_{x \in B} f_{\tilde{x}}(x), \text{ for all } B \in \mathcal{B}.$$

- An absolutely continuous distribution is fully characterized by its density function (pdf), $f_{\tilde{x}} : (\mathbb{R}, \mathcal{B}) \longrightarrow (\overline{\mathbb{R}}, \mathcal{B})$, since

$$P_{\tilde{x}}(B) = P\{\tilde{x} \in B\} = \int_B f_{\tilde{x}}(x)dx, \text{ for all } B \in \mathcal{B}.$$

- Let $\tilde{x}$ be an absolutely continuous random variable with density $f_{\tilde{x}}$. We assume, without loss of generality, that the Borel set $A = \{x \in \mathbb{R}^n \,|\, f_{\tilde{x}}(x) \neq 0\}$ is open in $\mathbb{R}^n$.
- Thus, under this convention, the closure of A is the support $supp\,(P_{\tilde{x}})$ of the distribution of the absolutely continuous random variable $\tilde{x}$.

4.1. The discrete uniform distribution and the Dirac distribution

- The random variable/vector $\tilde{x} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}^n, \mathcal{B})$ has the **discrete uniform** distribution if its probability function is

$$f(x) = \frac{1}{k}, \text{ for } x = \underbrace{x_1, \ldots, x_k}_{\tilde{x}(\Omega)}, \text{ with } x_i \neq x_j \text{ for } i \neq j.$$

- If $k = 1$, then $\tilde{x}$ is a constant a.s.

- **Definition:** An extended random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \to \left(\overline{\mathbb{R}}^n, \mathcal{B}\right)$ is called "degenerate" when there is a value $x_0 \in \overline{\mathbb{R}}^n$ such that $P_{\tilde{x}}(x_0) = 1$. The distribution associated with this random variable is also called "degenerate".

- **Example:** The discrete random vector $(\tilde{x}_1, \tilde{x}_2)$, where $\tilde{x}_1$ is the number of points when rolling a dice and $\tilde{x}_2$ is the number of heads when tossing a coin has a probability function $f(x_1, x_2)$ summarized in the following table:

$x_2 \backslash x_1$	1	2	3	4	5	6
0	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$
1	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$	$1/12$

- Therefore,

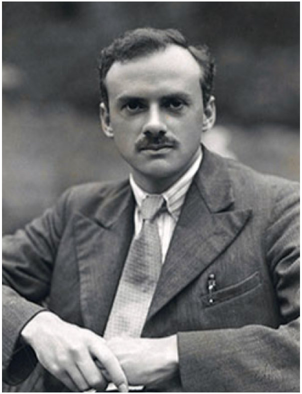
$$f(x_1, x_2) = \frac{1}{12}, \text{ for } (x_1, x_2) \in \{1, 2, 3, 4, 5, 6\} \times \{0, 1\}.$$

- The **Dirac** distribution $P_{\tilde{x}}$ of the random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ satisfies

$$P_{\tilde{x}}(A) = \begin{cases} 1 & \text{if } a \in A \\ \\ 0 & \text{if } a \notin A, \end{cases}$$

for all $A \in \mathcal{B}$.

- Note that $P_{\tilde{x}} \{a\} = 1$, i.e., the value a is taken almost surely by the random variable $\tilde{x}$.
- Some people define the density function of the previous Dirac distribution as a function (called the delta function) such that $f_{\tilde{x}}(x) = \infty$ if $x = a$, $f_{\tilde{x}}(x) = 0$ if $x \neq a$, and $\int_{\mathbb{R}} f_{\tilde{x}}(x)\, dx = 1$.
- Obviously, there is no density function with those properties, even though it can be viewed as the limit of a sequence of strictly positive density functions converging to zero for all $x \neq a$.



Paul Dirac (1902 - 1984)

4.2. The Bernoulli, binomial, Pascal, geometric, and hypergeometric distributions

Jacob Bernoulli (1654 - 1705)

- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ has the **Bernoulli** distribution if its probability function (pmf) is

$$f(x;\theta) = \theta^x (1-\theta)^{1-x}, \quad \text{for } x = 0, 1.$$

- or

$$f(x;\theta) = \begin{cases} 1-\theta & \text{for } x = 0 \\ \\ \theta & \text{for } x = 1. \end{cases}$$

- Mean and variance:

$$\mu = \theta$$

and

$$\sigma^2 = \theta(1-\theta).$$

- **The binomial distribution.**
- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ has a binomial distribution (or is binomial) if its probability function is

$$b(x; n, \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad \text{for } x = 0, 1, \ldots, n.$$

- *Motivation:* θ is the probability of a success in each trial. Then, $b(x; n, \theta)$ gives the probability of x successes in n independent trials.
- *Note:*

$$b(x; 1, \theta) = \theta^x (1-\theta)^{1-x}, \text{ for } x = 0, 1. \leftarrow \text{Bernoulli pmf}$$

- Obviously, if $\tilde{x}_1, ..., \tilde{x}_n$ are independently distributed random variables having a Bernoulli distribution with parameter θ, then its sum $\tilde{S} = \sum_{i=1}^{n} \tilde{x}_i$ has a binomial distribution with parameters n and θ.
- Therefore, if $\tilde{x}_1, ..., \tilde{x}_m$ are independently distributed random variables having a binomial distribution with parameters n_i and θ, for $i = 1, ..., m$, then its sum $\tilde{y} = \sum_{i=1}^{m} \tilde{x}_i$ has a binomial distribution with parameters $\sum_{i=1}^{m} n_i$ and θ.

- **Example:** Let $\tilde{x}$ be the number of heads when tossing 4 coins.

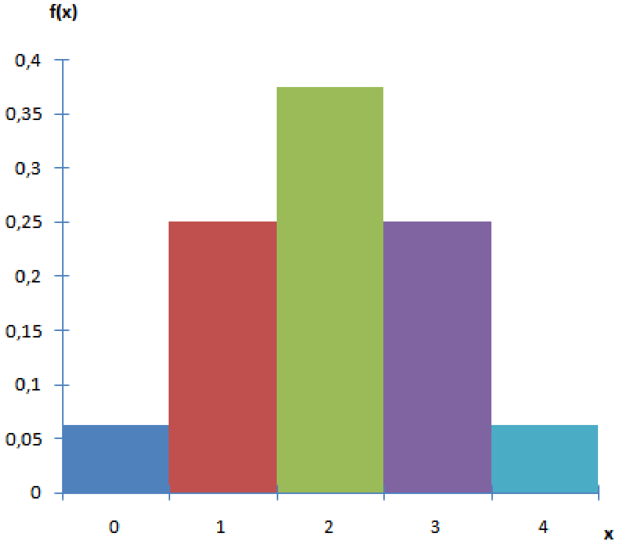
$$P\{\tilde{x}=x\}=f_{\tilde{x}}(x)=\underbrace{\binom{4}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{4-x}}_{b\left(x;4,\frac{1}{2}\right)}=\binom{4}{x}\left(\frac{1}{2}\right)^{4}=\frac{1}{16}\binom{4}{x},$$

$$\text{for } x=\underbrace{0,1,2,3,4}_{\tilde{x}(\Omega)}$$

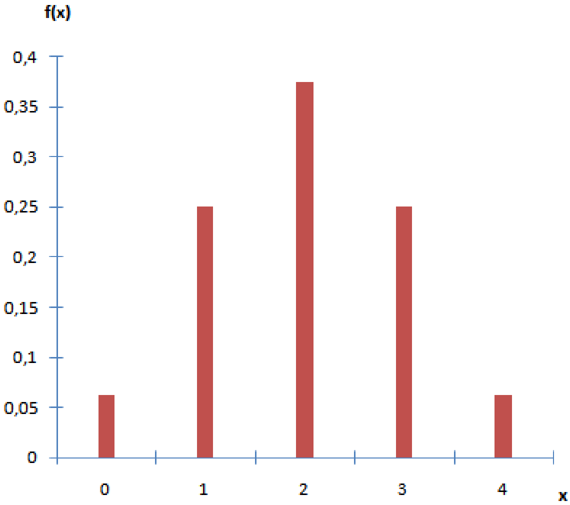
or

$$f_{\tilde{x}}(x)=b\left(x;4,\frac{1}{2}\right)=\begin{cases}1/16=0.0625 & \text{for } x=0\\ 4/16=0.25 & \text{for } x=1\\ 6/16=0.375 & \text{for } x=2\\ 4/16=0.25 & \text{for } x=3\\ 1/16=0.0625 & \text{for } x=4.\end{cases}$$

- Probability Histogram (which is symmetric iff $\theta = 1/2$):



- Probability Bar Chart (which is symmetric iff $\theta = 1/2$):



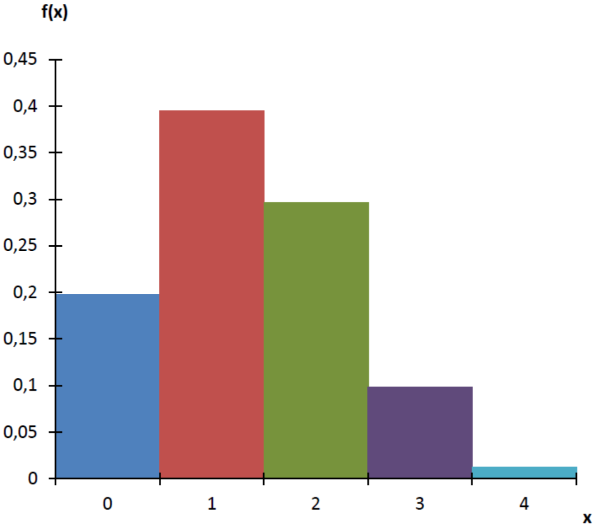
- **Example:** Let $\tilde{x}$ be the number of heads when tossing 4 coins. The coin is unbalanced and the probability of head is $\theta = 1/3$.

$$P\{\tilde{x} = x\} = f_{\tilde{x}}(x) = \underbrace{\binom{4}{x}\left(\frac{1}{3}\right)^{x}\left(\frac{2}{3}\right)^{4-x}}_{b\left(x;4,\frac{1}{3}\right)} \quad \text{for } x = \underbrace{0,1,2,3,4}_{\tilde{x}(\Omega)}$$

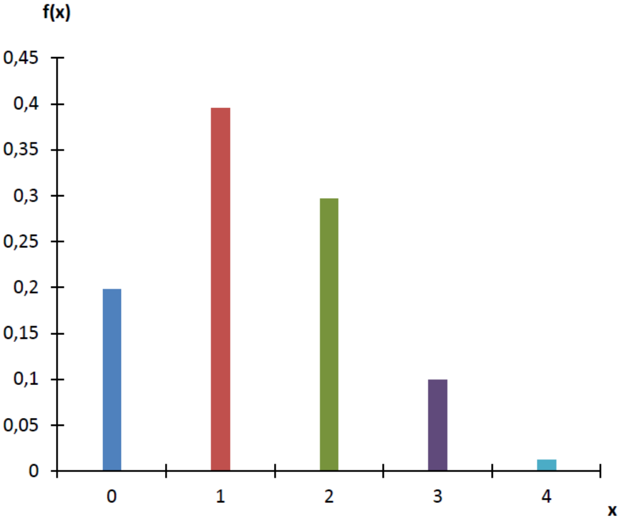
or

$$f_{\tilde{x}}(x) = b\left(x;4,\frac{1}{3}\right) = \begin{cases} 0.1975 & \text{for } x = 0 \\ 0.3951 & \text{for } x = 1 \\ 0.2963 & \text{for } x = 2 \\ 0.0988 & \text{for } x = 3 \\ 0.0123 & \text{for } x = 4. \end{cases}$$

- Probability Histogram (which is not symmetric since $\theta = 1/3$):



- Probability Bar Chart (which is not symmetric since $\theta = 1/3$):



- Moment-generating function of a binomial random variable:

$$M_{\tilde{x}}(t) = \left[1 + \theta(e^t - 1)\right]^n.$$

- **Proof:**

$$M_{\tilde{x}}(t) = \mathrm{E}\left(e^{t\tilde{x}}\right) = \sum_{x=0}^{n} e^{tx} \underbrace{\binom{n}{x} \theta^x (1-\theta)^{n-x}}_{b(x;n,\theta)}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \left(\theta e^t\right)^x (1-\theta)^{n-x} = \left[\theta e^t + (1-\theta)\right]^n = \left[1 + \theta(e^t - 1)\right]^n.$$

- Then,

$$\mu = M'_{\tilde{x}}(0) = n\left[1+\theta(e^t-1)\right]^{n-1} \theta e^t \Big|_{t=0} = n\theta$$

and

$$\sigma^2 = \mathrm{E}\left(\tilde{x}^2\right) - \left[\mathrm{E}\left(\tilde{x}\right)\right]^2 = M''_{\tilde{x}}(0) - \mu^2 =$$

$$\left\{n(n-1)\left[1+\theta(e^t-1)\right]^{n-2}\theta^2 e^{2t} + n\left[1+\theta(e^t-1)\right]^{n-1}\theta e^t\right\}\Big|_{t=0}$$

$$- n^2\theta^2 = n(n-1)\theta^2 + n\theta - n^2\theta^2 = -n\theta^2 + n\theta = n\theta(1-\theta).$$

- **The Pascal (negative binomial or binomial waiting-time) distribution.**

Blaise Pascal (1623 - 1662)

- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ has a Pascal distribution if its probability function is

$$b^*(x;\ k, \theta) = \binom{x-1}{k-1} \theta^k (1-\theta)^{x-k}, \quad \text{for } x = k, k+1, \ldots$$

- *Motivation:* θ is the probability of success in each trial. Then, $b^*(x;\ k, \theta)$ gives the probability that the k^{th} success will occur on the x^{th} independent trial.
- *Note:* $b^*(x; k, \theta) = \theta b(k-1; x-1, \theta)$.
- Mean, variance, and moment-generating function:

$$\mu = \frac{k}{\theta}\ , \quad \sigma^2 = \frac{k(1-\theta)}{\theta^2},$$

$$M_{\tilde{x}}(t) = \left[\frac{\theta e^t}{1-(1-\theta)e^t} \right]^k \quad \text{for } t < \underbrace{-\ln(1-\theta)}_{+}.$$

- **The geometric distribution.**
- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ has a geometric distribution if its probability function is

$$g(x;\theta) = \theta(1-\theta)^{x-1}, \quad \text{for } x = 1, 2, \ldots$$

- *Motivation:* θ is the probability of success in each trial. Then, $g(x;\theta)$ gives the probability that the first success will occur on the x^{th} independent trial.
- *Note:* $g(x;\theta) = b^*(x;1,\theta)$.
- Thus, the moment-generating function is

$$M_{\tilde{x}}(t) = \frac{\theta e^t}{1-(1-\theta)e^t} \quad \text{for } t < \underbrace{-\ln(1-\theta)}_{+}.$$

- Mean and variance:

$$\mu = M'_{\tilde{x}}(0) = \frac{1}{\theta} \quad \text{and} \quad \sigma^2 = M''_{\tilde{x}}(0) - \mu^2 = \frac{1-\theta}{\theta^2}.$$

- **The hypergeometric distribution.**
- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ has a hypergeometric distribution if its probability function is

$$h(x; n, N, a) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}},$$

for $x = 0, 1 \ldots, n, \; x \leq a, \; n - x \leq N - a.$

- *Motivation:* Consider a set of N elements of which a are successes and $N-a$ are failures. We choose, without replacement, n of the N elements contained in the set. Then, $h(x;\, n, N, a)$ gives the probability of getting x successes.

- Mean and variance:

$$\mu = \frac{na}{N}$$

and

$$\sigma^2 = \frac{na(N-a)(N-n)}{N^2(N-1)}.$$

- *Note:*

$$\lim_{\substack{N \to \infty \\ a/N = \theta}} h(x; n, N, a) = b(x; n, \theta)$$

4.3. The multinomial and multivariate hypergeometric distributions.

- **The multinomial distribution.**
- The random vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_k) : (\Omega, \mathcal{F}, P) \to \left(\mathbb{R}^k, \mathcal{B}\right)$ has the multinomial distribution if its probability function is

$$m\left(x_1, x_2, \ldots, x_k; n, \theta_1, \theta_2, \ldots, \theta_k\right) = \binom{n}{x_1, x_2, \ldots, x_k} \theta_1^{x_1} \theta_2^{x_2} \cdots \theta_k^{x_k},$$

for $x_i = 0, 1, \ldots, n$, $\sum_{i=1}^{k} x_i = n$ and $\sum_{i=1}^{k} \theta_i = 1$.

- Recall that

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1! \cdot x_2! \cdot \ldots \cdot x_k!}.$$

- *Motivation:* There are n independent trials permitting k exclusive outcomes, whose respective probabilities are $\theta_1, \theta_2, \ldots, \theta_k$ (with $\sum_{i=1}^{k} \theta_i = 1$). We shall refer to the outcomes as being of the first type, the second type, ... and the k^{th} type. Then, $m\left(x_1, x_2, \ldots, x_k; n, \theta_1, \theta_2, \ldots, \theta_k\right)$ gives the probability of getting x_1 outcomes of the first type, x_2 outcomes of the second type, ... and x_k outcomes of the k^{th} type (with $\sum_{i=1}^{k} x_i = n$).

- *Note:* $m(x, n-x; n, \theta, 1-\theta) = b(x;\ n, \theta)$.

- **Example:** Consider a very large population. 50% of the individuals have brown eyes, 30% have black eyes, and 20% have blue eyes. We pick randomly 8 individuals. The probability of picking 5 individuals with brown eyes, 2 with black eyes, and 1 with blue eyes is

$$m\left(5, 2, 1; 8, 0.5, 0.3, 0.2\right) = \frac{8!}{5!2!1!} \cdot 0.5^5 \cdot 0.3^2 \cdot 0.2^1 = 0.0945.$$

- **The multivariate hypergeometric distribution.**
- The random vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_k) : (\Omega, \mathcal{F}, P) \to \left(\mathbb{R}^k, \mathcal{B}\right)$ has the multivariate hypergeometric distribution if its probability function is

$$f(x_1, x_2, \ldots, x_k; n, N, a_1, a_2, \ldots, a_k) = \frac{\binom{a_1}{x_1}\binom{a_2}{x_2}\cdots\binom{a_k}{x_k}}{\binom{N}{n}},$$

for $x_i = 0, 1, \ldots, n$, and $x_i \leq a_i$, where $\sum_{i=1}^{k} x_i = n$, $\sum_{i=1}^{k} a_i = N$.

- *Motivation:* There is a set of N elements, of which a_1 are elements of the first type, a_2 are elements of the second type, ..., a_k are elements of the k^{th} type (with $\sum_{i=1}^{k} a_i = N$). We choose, without replacement, n of the N elements in the set. Then $f(x_1, x_2, \ldots, x_k; n, N, a_1, a_2, \ldots, a_k)$ gives the probability of getting x_1 outcomes of the first type, x_2 outcomes of the second type, ... and x_k outcomes of the k^{th} type (with $\sum_{i=1}^{k} x_i = n$).

- *Note:* $f\,(x, n-x;\ n, N, a, N-a) = h(x;\ n, N, a)$.
- *Note:*

$$\lim_{\substack{N \to \infty \\ a_1/N = \theta_1 \\ a_2/N = \theta_2 \\ \cdots \\ a_k/N = \theta_k}} f\,(x_1, x_2, \ldots, x_k; n, N, a_1, a_2, \ldots, a_k)$$

$$= m\,(x_1, x_2, \ldots, x_k; n, \theta_1, \theta_2, \ldots, \theta_k)\,.$$

- Example of the multinomial with $N = 2000$:

$$f\,(5, 2, 1; 8, 2000, 1000, 600, 400) = \frac{\binom{1000}{5}\binom{600}{2}\binom{400}{1}}{\binom{2000}{8}} = 0.0947$$

$$\approx m\,(5, 2, 1; 8, 0.5, 0.3, 0.2) = 0.0945.$$

4.4. Integration by parts for Lebesgue-Stieltjes integrals.

Recall that, if μ is the Lebesgue-Stieltjes measure associated with F,

$$\mu(a,b] = \int_{(a,b]} d\mu(x) \equiv \int_{(a,b]} dF(x) = F(b) - F(a),$$

$$\mu\,[a,b] = \mu\,\{a\} + \int_{(a,b]} dF(x)$$

$$= F(a) - F(a^-) + \int_{(a,b]} dF(x) = F(b) - F(a^-),$$

$$\mu\,(a,b) = -\mu\,\{b\} + \int_{(a,b]} dF(x)$$

$$= -\left[F(b) - F(b^-)\right] + \int_{(a,b]} dF(x) = F(b^-) - F(a),$$

$$\mu[a,b) = \mu\,\{a\} - \mu\,\{b\} + \int_{(a,b]} dF(x)$$

$$= F(a) - F(a^-) - \left[F(b) - F(b^-)\right] + \int_{(a,b]} dF(x) = F(b^-) - F(a^-).$$

- **Integration by parts.**
- From now on, whenever we write the integral of a function w.r.t. a measure it should be understood that the function is integrable w.r.t. that measure.
- Assume that $F : M \longrightarrow \mathbb{R}$ and $G : M \longrightarrow \mathbb{R}$ are continuously differentiable functions, where M is an open subset of $\mathbb{R}$, and $[a, b] \subset M$. Then,

$$\int_{[a,b]} G(x)F'(x)dx = \underbrace{F(b)G(b) - F(a)G(a)}_{[F(x)G(x)]_a^b} - \int_{[a,b]} G'(x)F(x)dx.$$

- *Note:* The interval of integration can be replaced by (a, b), $(a, b]$, or $[a, b)$.
- The previous Lebesgue integrals are equal to their Riemann counterparts as the functions $G \cdot F'$ and $G' \cdot F$ are continuous on $[a, b]$.

- Remember that, if F is a distribution function, then

$$\int_{(a,b]} G(x)dF(x) \equiv \int_{(a,b]} G(x)d\mu(x),$$

where μ is the Lebesgue-Stieltjes measure associated with F.

- Assume that $F : \mathbb{R} \longrightarrow \mathbb{R}$ is a distribution function, $G : M \longrightarrow \mathbb{R}$ is a continuously differentiable function, where M is an open subset of $\mathbb{R}$, and $[a, b] \subset M$. Then,

$$\int_{(a,b]} G(x)dF(x) = F(b)G(b) - F(a)G(a) - \int_{(a,b]} F(x)G'(x)dx. \quad (\star)$$

- Note that $F(a^+) = F(a)$, because of the right-continuity of F.

- Since G is continuous on $[a, b]$, then

$$\underbrace{\int_{(a,b]} G(x)dF(x)}_{\text{Lebesgue-Stieltjes integral}} = \underbrace{\int_a^b G(x)dF(x)}_{\text{Riemann-Stieltjes integral}}.$$

- The last (Lebesgue) integral in $(\star)$ obviously satisfies

$$\int_{(a,b]} F(x)G'(x)dx = \int_{(a,b)} F(x)G'(x)dx$$
$$= \int_{[a,b]} F(x)G'(x)dx = \int_{[a,b)} F(x)G'(x)dx.$$

-

$$\int_{[a,b]} G(x)dF(x) =$$
$$G(a)\left(F(a) - F(a^-)\right) + F(b)G(b) - F(a)G(a) - \int_{[a,b]} F(x)G'(x)dx$$
$$= F(b)G(b) - F(a^-)G(a) - \int_{[a,b]} F(x)G'(x)dx.$$

- $$\int_{(a,b)} G(x)dF(x) =$$

  $$F(b)G(b) - F(a)G(a) - G(b)\left(F(b) - F(b^-)\right) - \int_{(a,b)} F(x)G'(x)dx$$

  $$= F(b^-)G(b) - F(a)G(a) - \int_{(a,b)} F(x)G'(x)dx.$$

- $$\int_{[a,b)} G(x)dF(x) =$$

  $$G(a)\left(F(a) - F(a^-)\right) + \underbrace{F(b^-)G(b) - F(a)G(a) - \int_{(a,b)} F(x)G'(x)dx}_{\int_{(a,b)} G(x)dF(x)}$$

  $$= F(b^-)G(b) - F(a^-)G(a) - \int_{[a,b)} F(x)G'(x)dx.$$

4.5. Lebesgue integration by change of variable: polar coordinates.

- Assume that (i) $g : M \longrightarrow \mathbb{R}$ is a continuously differentiable function, where M is an open subset of $\mathbb{R}$, and (ii) the function g restricted to $g^{-1}([a, b])$, $g : g^{-1}([a, b]) \longrightarrow [a, b]$, where $[a, b] \subset g(M)$ (or $g^{-1}([a, b]) \subset M$), is a bijective function (or one-to-one correspondence). Let $f : ([a, b], \mathcal{B}([a, b])) \longrightarrow (\mathbb{R}, \mathcal{B})$ be a Lebesgue integrable function. Then,

$$\int_{[a,b]} f(x)dx = \int_{g^{-1}([a,b])} f(\underbrace{g(y)}_{x}) \left|g'(y)\right| dy.$$

- See the handout for an informal proof.
- Note that $g^{-1}([a, b])$ is the interval $\left[g^{-1}(a), g^{-1}(b)\right]$ if g is increasing or is the interval $\left[g^{-1}(b), g^{-1}(a)\right]$ if g is decreasing.

- Assume that (i) $g : M \longrightarrow \mathbb{R}^n$ is a continuously differentiable function, where M is an open subset of $\mathbb{R}^n$, and (ii) the function g restricted to $g^{-1}(B)$, $g : g^{-1}(B) \longrightarrow B$, where B is a Borel set in $\mathbb{R}^n$ such that $B \subset g(M)$ (or $g^{-1}(B) \subset M$), is a bijective function (or one-to-one correspondence).
- Let $J_g(y_1, y_2, ..., y_n)$ be the Jacobian matrix of the function

$$g : \underbrace{(y_1, ..., y_n)}_{y \in \mathbb{R}^n} \longmapsto \underbrace{(g_1(y_1, y_2, ..., y_n), ..., g_n(y_1, y_2, ..., y_n))}_{g(y) \in \mathbb{R}^n},$$

which is given by

$$J_g(y_1, y_2, ..., y_n) = \begin{pmatrix} \frac{\partial g_1(y)}{\partial y_1} & \frac{\partial g_1(y)}{\partial y_2} & \cdots & \cdots & \frac{\partial g_1(y)}{\partial y_n} \\ \frac{\partial g_2(y)}{\partial y_1} & \frac{\partial g_2(y)}{\partial y_2} & \cdots & \cdots & \frac{\partial g_2(y)}{\partial y_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g_n(y)}{\partial y_1} & \frac{\partial g_n(y)}{\partial y_2} & \cdots & \cdots & \frac{\partial g_n(y)}{\partial y_n} \end{pmatrix},$$

- Let $f:(B,\mathcal{B}(B))\longrightarrow(\mathbb{R}^n,\mathcal{B})$ be a Lebesgue integrable function. Then,

$$\int_B f(x_1,x_{2,}...,x_n)d(x_1,...,x_n)=$$

$$\int_{g^{-1}(B)} f(\underbrace{g(y_1,y_2,...,y_n)}_{(x_1,x_{2,}...,x_n)})\left|J_g(y_1,y_2,...,y_n)\right| d(y_1,...,y_n),$$

where $\left|J_g(y_1,y_2,...,y_n)\right|$ is the absolute value of the determinant of $J_g(y_1,y_2,...,y_n)$.

- Alternatively, we can assume that (i) $g : M \longrightarrow \mathbb{R}^n$ is a continuously differentiable function, where M is an open subset of $\mathbb{R}^n$, and (ii) the function g restricted to A, $g : A \longrightarrow g(A)$, where A is a Borel set in $\mathbb{R}^n$ such that $A \subset M$, is a bijective function (or one-to-one correspondence). Let $f : (g(A), \mathcal{B}(g(A))) \longrightarrow (\mathbb{R}^n, \mathcal{B})$ be a Lebesgue integrable function. Then,

$$\int_{g(A)} f(x_1, x_2, ..., x_n) d(x_1, ..., x_n) =$$

$$\int_A f(\underbrace{g(y_1, y_2, ..., y_n)}_{(x_1, x_2, ..., x_n)}) \left| J_g(y_1, y_2, ..., y_n) \right| d(y_1, ..., y_n).$$

- A very useful change of variable (or substitution) is the change to **polar coordinates** in $\mathbb{R}^2$.
- Consider the function

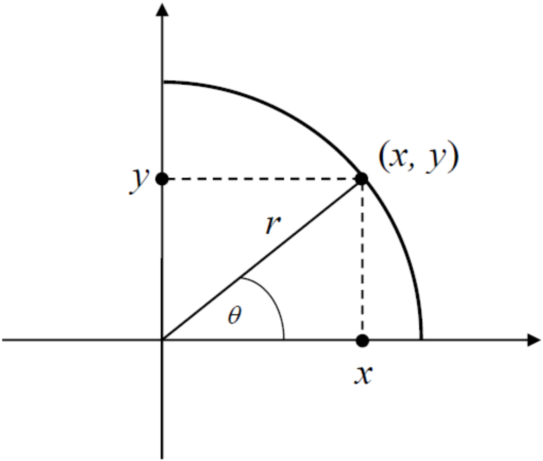
$$g : \mathbb{R}_{++} \times [0, 2\pi) \longrightarrow \mathbb{R}^2,$$

given by

$$(x, y) = g(r, \theta), \text{ with } \begin{cases} x = r \cdot \cos\theta \\ \\ y = r \cdot \sin\theta. \end{cases}$$

- Note that

$$x^2 + y^2 = r^2 \cos^2 \theta + r^2 \sin^2 \theta = r^2.$$



- Clearly, $g\left(\mathbb{R}_{++}\times[0,2\pi)\right)=\mathbb{R}^2\backslash\left(0,0\right)$, where $\mathbb{R}^2\backslash\left(0,0\right)$ is the real plane without the point $(0,0)$, which has zero Lebesgue measure.

- Moreover, the function g restricted to
$g^{-1}\left(\mathbb{R}^2\backslash\left(0,0\right)\right)=\mathbb{R}_{++}\times[0,2\pi)$,
$g:\mathbb{R}_{++}\times[0,2\pi)\longrightarrow\mathbb{R}^2\backslash\left(0,0\right)$, is bijective.

- C is a circular region in $\mathbb{R}^2\backslash\left(0,0\right)$ if $g^{-1}(C)$ is a measurable rectangle on $\mathbb{R}_{++}\times[0,2\pi)$, that is, if $g^{-1}(C)$ is the Cartesian product of two Borel sets, $g^{-1}(C)=D_1\times D_2$, with $D_1\in\mathcal{B}\left(\mathbb{R}_{++}\right)$ and $D_2\in\mathcal{B}\left([0,2\pi)\right)$.

- Obviously, we can define the inverse function g^{-1} defined on $\mathbb{R}^2 \setminus (0,0)$ as

$$(r,\theta) = g^{-1}(x,y), \text{ with } \begin{cases} r = \left(x^2 + y^2\right)^{1/2} \\ \theta = \arctan\left(\dfrac{y}{x}\right). \end{cases}$$

- Moreover,

$$J_g\left(r,\theta\right) = \begin{bmatrix} \cos\theta & -r\cdot\sin\theta \\ \sin\theta & r\cdot\cos\theta \end{bmatrix},$$

- so that

$$\left|J_g\left(r,\theta\right)\right| = \left|r\cos^2\theta + r\sin^2\theta\right| = |r| = r.$$

- Therefore, by making the change to polar coordinates, we can compute the following integral over a circular region C:

$$\int_C f(x,y)\, d(x,y) = \int_{g^{-1}(C)} f(r\cos\theta, r\sin\theta) r\ d(r,\theta)$$

$$= \int_{D_1}\int_{D_2} f(r\cos\theta, r\sin\theta) r d\theta dr,$$

where we use Fubini's theorem in the last equality since $g^{-1}(C)$ is a measurable rectangle on $\mathbb{R}_{++} \times [0, 2\pi)$, i.e., it is the Cartesian product of two Borel sets, $g^{-1}(C) = D_1 \times D_2$, with $D_1 \in \mathcal{B}(\mathbb{R}_{++})$ and $D_2 \in \mathcal{B}([0, 2\pi))$.

- Moreover, we can easily compute the following integral over a circular region C :

$$\int_C h\left(x^2 + y^2\right) d(x,y) = \int_{g^{-1}(C)} h(r^2) r\ d(r,\theta)$$

$$= \int_{D_1}\int_{D_2} h(r^2) r d\theta dr = \left(\int_{D_1} h(r^2) r dr\right)\left(\int_{D_2} d\theta\right).$$

- **Example:**

$$\int_0^\infty \int_0^\infty e^{-\left(x^2+y^2\right)} dxdy = \int_{\mathbb{R}_+ \times \mathbb{R}_+} e^{-\left(x^2+y^2\right)} d(x,y) =$$

$$= \left(\int_{\mathbb{R}_{++}} e^{-r^2} r dr\right) \left(\int_{[0,\pi/2)} d\theta\right) = \left[\frac{-e^{-r^2}}{2}\right]_0^\infty \cdot [\theta]_0^{\pi/2} = \frac{1}{2} \cdot \frac{\pi}{2} = \frac{\pi}{4}.$$

- Note that

$$\int_0^\infty \int_0^\infty e^{-\left(x^2+y^2\right)} dxdy = \left(\int_0^\infty e^{-x^2} dx\right) \left(\int_0^\infty e^{-y^2} dy\right) = M^2,$$

where $M = \int_0^\infty e^{-x^2} dx = \int_0^\infty e^{-y^2} dy.$

- Therefore,

$$M = \int_0^\infty e^{-x^2} dx = \left(\frac{\pi}{4}\right)^{1/2} = \frac{\sqrt{\pi}}{2} \Longrightarrow \int_{-\infty}^\infty e^{-x^2} dx = \sqrt{\pi}.$$

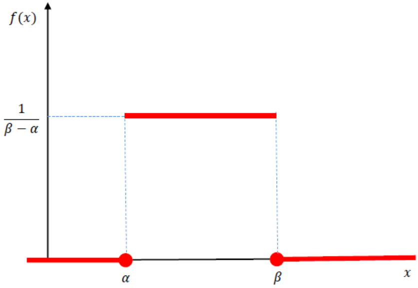
4.6. The uniform density

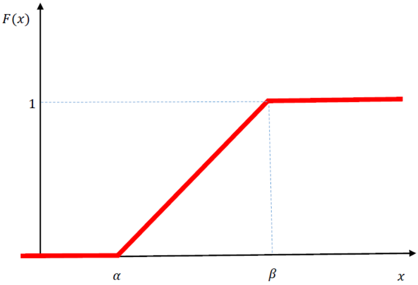
- The absolutely continuous random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ is uniform (or has a uniform distribution) on (α, β) if its density function is

$$f(x) = \begin{cases} \dfrac{1}{\beta - \alpha} & \text{for } x \in (\alpha, \beta) \\ 0 & \text{otherwise.} \end{cases}$$

- The distribution function is

$$F(x) = \begin{cases} 0 & \text{for } x \leq \alpha \\ \dfrac{x - \alpha}{\beta - \alpha} & \text{for } x \in (\alpha, \beta) \\ 1 & \text{for } x \geq \beta. \end{cases}$$


$f(x)$
$\frac{1}{\beta-\alpha}$
α
β
x


$F(x)$
1
α
β
x

- Quantile function:

$$Q(p) = \alpha + (\beta - \alpha)p \;\text{ for } p \in (0,1).$$

- If $\alpha = 0$ and $\beta = 1$, then $\tilde{x}$ has the uniform distribution on $(0,1)$ (also called the standard uniform distribution). In this case, $Q(p) = p = F(p)$ for $p \in (0,1)$.
- **Proposition.** Let F be a distribution function and Q is its associated quantile function (i.e., $Q(p) = \inf\{x \,|\, F(x) \geq p\}$ for $p \in (0,1)$). If the random variable $\tilde{p}$ has the standard uniform distribution, then the random variable $\tilde{x} = Q(\tilde{p})$ has the distribution function F.
- **Proof.** Recall that the quantile function satisfies the following:

$$Q(p) \leq x \;\text{ if and only if }\; F(x) \geq p.$$

Since the random variable $\tilde{p}$ has the standard uniform distribution, it holds that $F_{\tilde{p}}(p) = P\{\tilde{p} \leq p\} = p$ for $p \in (0,1)$. Therefore,

$$P\{\tilde{x} \leq x\} = P\{Q(\tilde{p}) \leq x\} = P\{\tilde{p} \leq F(x)\} = F(x). \qquad Q.E.D.$$

- The previous proposition allows us to generate random values from a distribution having an arbitrarily given distribution function F.
- To do so, generate random values p_i, $i = 1, 2, ..., N$, from the standard uniform distribution and apply to these values the quantile function, $x_i = Q(p_i)$, then the values x_i, $i = 1, 2, ..., N$, are random values from a distribution having the distribution function F.

- Let $\alpha \leq c < d \leq \beta$, then,

$$P\left\{c \underset{(<)}{\leq} \widetilde{x} \underset{(<)}{\leq} d\right\} = P_{\tilde{x}}\left[c, d\right] = \frac{d-c}{\beta-\alpha}.$$

- Mean:

$$\mu = \mathrm{E}\left(\tilde{x}\right) = \int_{-\infty}^{\infty} xf\left(x\right)dx = \int_{\alpha}^{\beta} x\left(\frac{1}{\beta-\alpha}\right)dx = \frac{1}{\beta-\alpha}\left[\frac{x^2}{2}\right]_{\alpha}^{\beta}$$

$$= \frac{1}{\beta-\alpha}\left[\frac{\beta^2-\alpha^2}{2}\right] = \frac{\beta+\alpha}{2},$$

where the last inequality follows since $(\beta+\alpha)(\beta-\alpha) = \beta^2-\alpha^2$.

- Moreover,

$$\mathrm{E}\left(\tilde{x}^2\right) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_{\alpha}^{\beta} x^2 \left(\frac{1}{\beta-\alpha}\right) dx$$

$$= \frac{1}{\beta-\alpha}\left[\frac{x^3}{3}\right]_{\alpha}^{\beta} = \frac{\beta^3-\alpha^3}{3(\beta-\alpha)},$$

so that

$$\mathrm{Var}\left(\tilde{x}\right) = \mathrm{E}\left(\tilde{x}^2\right) - \left[\mathrm{E}\left(\tilde{x}\right)\right]^2 = \frac{\beta^3-\alpha^3}{3(\beta-\alpha)} - \left[\frac{\beta+\alpha}{2}\right]^2 = \frac{(\beta-\alpha)^2}{12},$$

where the last equality is obtained after some tedious algebra.

- Moment-generating function:

$$M_{\tilde{x}}(t) = \begin{cases} \dfrac{e^{t\beta}-e^{t\alpha}}{t\left(\beta-\alpha\right)} & \text{for } t \neq 0 \\ 1 & \text{for } t = 0. \end{cases}$$

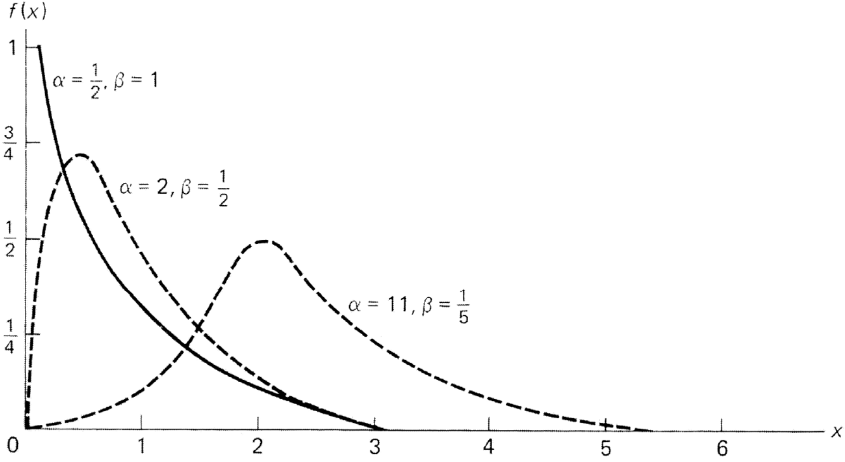
- **The gamma distribution.**
- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ has the gamma distribution if its density is

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{1}{\beta^{\alpha}\Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

with $\alpha > 0$, $\beta > 0$ and where

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy, \text{ for } \alpha > 0. \qquad \text{(the gamma function)}$$

Densities of the gamma distribution:



- Making the change of variable

$$x = g(y) = \beta y \Leftrightarrow y = g^{-1}(x) = \frac{x}{\beta}, \text{ so that } \frac{dx}{dy} = g'(y) = \beta > 0,$$

we can check that

$$\int_0^{\infty} f(x;\alpha,\beta)dx = \int_0^{\infty} kx^{\alpha-1}e^{-x/\beta}dx = k\int_0^{\infty} (\beta y)^{\alpha-1} e^{-y}\beta dy$$

$$= k\beta^{\alpha} \underbrace{\int_0^{\infty} y^{\alpha-1}e^{-y}dy}_{\Gamma(\alpha)} = 1.$$

Therefore,

$$k = \frac{1}{\beta^{\alpha}\Gamma(\alpha)}.$$

Properties of the gamma function $\Gamma(\alpha)$.

- **(a)** $\Gamma(1) = \int_0^\infty e^{-y}\, dy = \left[-e^{-y}\right]_0^\infty = \lim_{y \to \infty} \left(-e^{-y}\right) + e^0 = 1.$
- **(b)** $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ for $\alpha > 1$.
- **Proof.** Integrating by parts:

$$\Gamma(\alpha) = \int_0^\infty \underbrace{y^{\alpha-1}}_{F(y)} \underbrace{e^{-y}}_{G'(y)} dy = \left[-y^{\alpha-1}e^{-y}\right]_0^\infty - \int_0^\infty \underbrace{(\alpha-1)y^{\alpha-2}}_{F'(y)} \underbrace{\left(-e^{-y}\right)}_{G(y)} dy$$

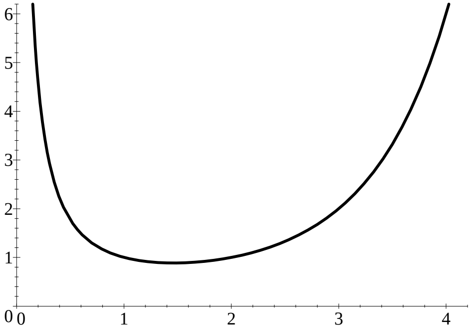
$$= 0 + (\alpha - 1)\int_0^\infty y^{\alpha-2}e^{-y}\, dy = (\alpha-1)\Gamma(\alpha-1).$$

- Note that (a), (b), and the continuity of the gamma function imply that

$$1 = \lim_{\alpha \to 1^+} \Gamma(\alpha) = \lim_{x \equiv \alpha - 1 \to 0^+} \left[x \cdot \Gamma(x)\right] \Longrightarrow \lim_{x \to 0^+} \Gamma(x) = \infty.$$

- **(c)** $\Gamma(\alpha) = (\alpha - 1)!$ when α is a strictly positive integer.

- **(d)** $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
  - **Proof:** See handout 1.A.
- **Corollary.** $\int_0^\infty e^{-\frac{1}{2}z^2} dz = \sqrt{\frac{\pi}{2}}$.
  - **Proof:** See handout 1.B.



The gamma function $\Gamma(\alpha)$

- Let $\tilde{x}$ be a random variable that has a gamma distribution with parameters α and β. Then,

- (a) $\mu'_r = \dfrac{\beta^r \Gamma(\alpha + r)}{\Gamma(\alpha)}$.

    - **Proof.** See handout 1.C.

- (b) $M_{\tilde{x}}(t) = (1 - \beta t)^{-\alpha}$ for $t < 1/\beta$.

    - **Proof.** See handout 1.D.

- **Corollary.**

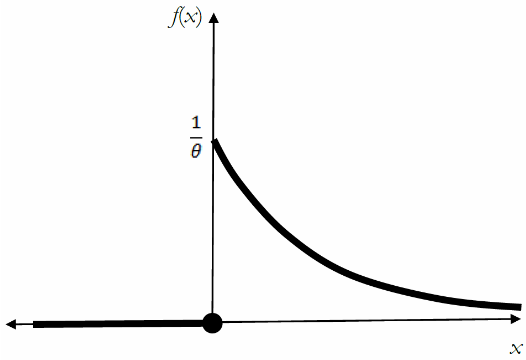
  (i) $\mu = \mu'_1 = \alpha\beta$,

  (ii) $\mu'_2 = \alpha(\alpha + 1)\beta^2$,

  (iii) $\sigma^2 = \mu'_2 - \mu^2 = \alpha\beta^2$.

- **The exponential distribution.**
- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ has the exponential distribution if its density is the gamma density with $\alpha = 1$ and $\beta = \theta$,

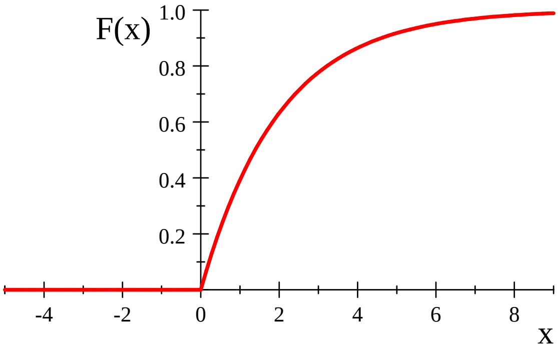
$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-x/\theta} & \text{for } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$



- This distribution is used to model waiting time.
- The distribution function is

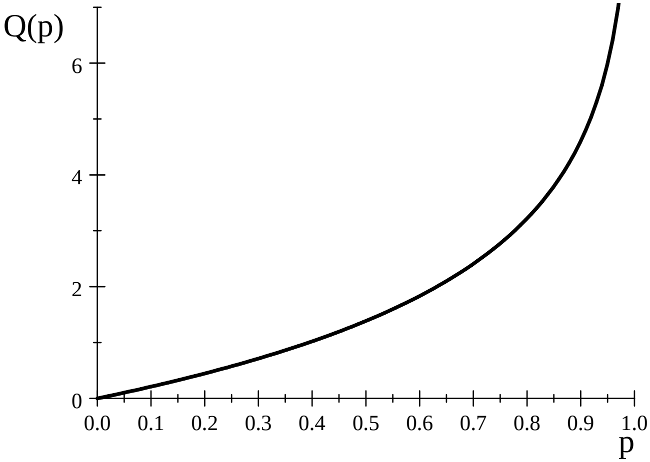
$$P\left\{\tilde{x} \leq x\right\} = F(x) = \begin{cases} 1 - e^{-x/\theta} & \text{for } x > 0 \\ \\ 0 & \text{otherwise,} \end{cases}$$

which gives the probability of waiting less than x units of time.



- Quantile function:

$$Q(p) = -\theta \ln(1-p) \quad \text{for } p \in (0,1).$$



- Mean and variance:

$$\mu = \alpha\beta = \theta,$$

$$\sigma^2 = \alpha\beta^2 = \theta^2.$$

- Moment-generating function:

$$M_{\tilde{x}}(t) = (1 - \beta t)^{-\alpha} = \frac{1}{1 - \theta t} \text{ for } t < 1/\theta.$$

- **The chi-square (χ^2) distribution.**

Karl Pearson (1857 – 1936)

- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ has the χ^2 (chi-square) distribution if its density is the gamma density with $\alpha = \nu/2$ and $\beta = 2$,

$$f(x; \nu) = \begin{cases} \dfrac{1}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)} x^{\frac{\nu-2}{2}} e^{-x/2} & \text{for } x > 0 \\ \\ 0 & \text{otherwise.} \end{cases}$$

- Mean and variance:

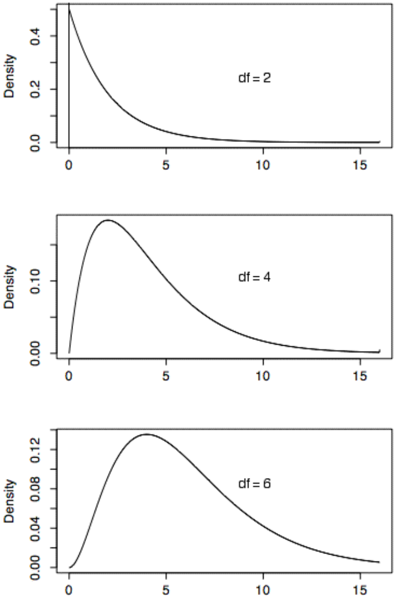
$$\mu = \alpha\beta = \nu,$$
$$\sigma^2 = \alpha\beta^2 = 2\nu.$$

- Moment-generating function:

$$M_{\tilde{x}}(t) = (1 - \beta t)^{-\alpha} = (1 - 2t)^{-\nu/2} \quad \text{for } t < 1/2.$$

- *Notation:*
  - $\tilde{x} \sim \mathrm{B}(n, \theta) \longleftarrow \tilde{x}$ has the binomial distribution
  - $\tilde{x} \sim \mathrm{U}(\alpha, \beta) \longleftarrow \tilde{x}$ has the uniform distribution on (α, β)
  - $\tilde{x} \sim \Gamma(\alpha, \beta) \longleftarrow \tilde{x}$ has the gamma distribution
  - $\tilde{x} \sim \chi^2_\nu \longleftarrow \tilde{x}$ has the chi-square distribution with ν degrees of freedom.

Densities of the chi-square distributions with 2, 4, and 6 degrees of freedom:



4.8. The beta distribution.

- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ has the beta distribution if its density is

$$f(x; \alpha, \beta) = \begin{cases} \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} & \text{for } x \in (0,1) \\ \\ 0 & \text{otherwise,} \end{cases}$$

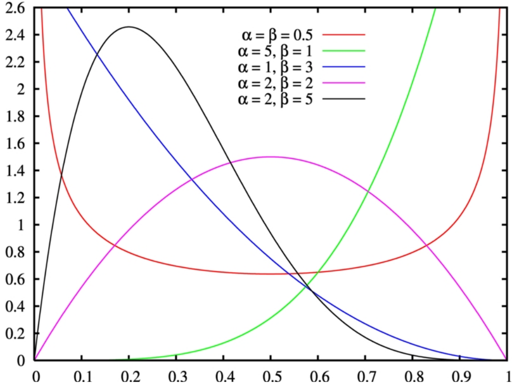
  with $\alpha > 0$, $\beta > 0$.

- If $\alpha = 1$ and $\beta = 1$, then the beta distribution becomes the uniform distribution on $(0, 1)$.

- Obviously, the following holds:

$$\int_0^1 f(x; \alpha, \beta)dx = 1 \qquad \text{(see the handout)}$$

- Densities of the beta distribution:



- Mean and variance:

$$\mu = \frac{\alpha}{\alpha + \beta},$$

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}.$$

4.9. The normal distribution

Carl Friedrich Gauss (1777 – 1855)

- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ has the normal distribution (or is normal) if its density is

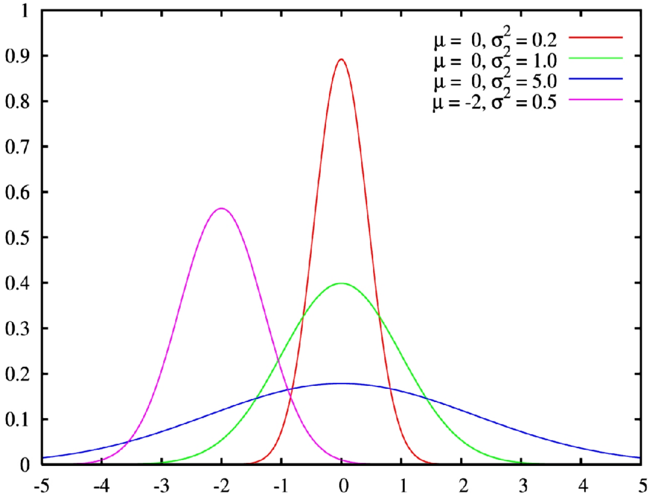
$$n(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad \text{with}\, \sigma > 0, \quad \text{for all } \ x \in \mathbb{R}.$$

- *Notation:* $\tilde{x} \sim \mathrm{N}(\mu, \sigma^2)$.

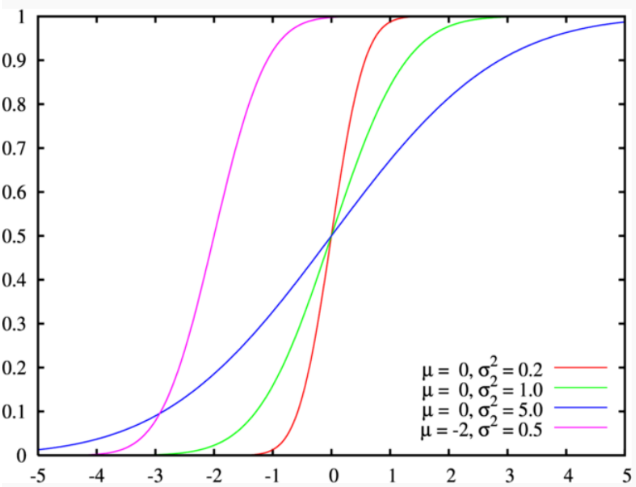
- A random variable $\tilde{z}$ has the standard normal distribution (or is standard normal) if $\tilde{z} \sim \mathrm{N}(0, 1)$. Thus, the density of a standard normal random variable is

$$n(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad \text{for all } \ z \in \mathbb{R}.$$

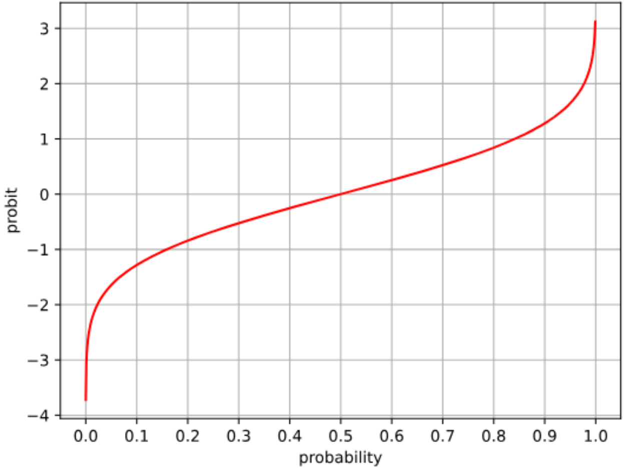
- Normal densities (i.e., densities of normal distributions) are symmetric around μ. The graph of the normal density is the Gauss bell.



Distribution functions (they do not have a closed-form expression):



- The quantile function of the standard normal distribution is the **probit** function used in probit regressions/models:



Properties of the normal distribution $\mathrm{N}\left(\mu, \sigma^2\right)$.

- **(1)** $\int_{-\infty}^{\infty} n(x; \mu, \sigma) = 1$.
  - **Proof:** See handout 3.A.
- **(2)** $M_{\tilde{x}}(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$. In particular, if $\tilde{z} \sim \mathrm{N}(0, 1)$, then $M_{\tilde{z}}(t) = e^{t^2/2}$.
  - **Proof:** See handout 3.B.
- **(3)**

$$M'_{\tilde{x}}(t) = (\mu + \sigma^2 t) M_{\tilde{x}}(t) \Rightarrow M'_{\tilde{x}}(0) = \mu = \mathrm{E}(\tilde{x}).$$

$$M''_{\tilde{x}}(t) = \sigma^2 M_{\tilde{x}}(t) + (\mu + \sigma^2 t)^2 M_{\tilde{x}}(t) \Rightarrow M''_{\tilde{x}}(0) = \sigma^2 + \mu^2 = \mathrm{E}(\tilde{x}^2).$$

$$\mathrm{Var}(\tilde{x}) = \mathrm{E}(\tilde{x}^2) - [\mathrm{E}(\tilde{x})]^2 = \sigma^2.$$

- *Note:* If $\mu_{\tilde{x}}$ and $\sigma_{\tilde{x}} > 0$ are the mean and the standard deviation, respectively, of the random variable $\tilde{x}$, then the "standardized" random variable $\tilde{z} = \dfrac{\tilde{x} - \mu_{\tilde{x}}}{\sigma_{\tilde{x}}}$ has $\mu_{\tilde{z}} = 0$ and $\sigma^2_{\tilde{z}} = \sigma_{\tilde{z}} = 1$.

- **(4)** If $\tilde{x} \sim \mathrm{N}(\mu, \sigma^2)$ and $\tilde{z} = \dfrac{\tilde{x} - \mu}{\sigma}$, then $\tilde{z} \sim \mathrm{N}(0, 1)$.

- **Proof:** Let $x = g(z) = \mu + \sigma z$ so that $z = g^{-1}(x) = \dfrac{x - \mu}{\sigma}$ and $g'(z) = \sigma > 0$.
  For all $A \in \mathcal{B}$, we have

$$P_{\tilde{z}}(A) = P\left\{\tilde{z} \in A\right\} = P\left\{g^{-1}(\tilde{x}) \in A\right\} = P\left\{\tilde{x} \in g\left(A\right)\right\}$$

$$= P_{\tilde{x}}(g\left(A\right)) = \int_{g(A)} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_A \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \sigma dz$$

$$= \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz = \int_A n(z; 0, 1) dz.$$

  Therefore, $\tilde{z} \sim \mathrm{N}(0, 1)$. *Q.E.D.*

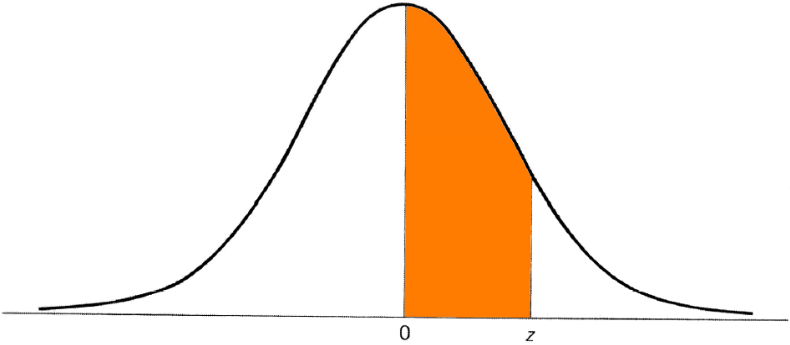
The distribution function of the standard normal distribution is tabulated:

The Standard Normal Distribution Function N(0,1)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Also, for z = 4.0, 5.0, and 6.0 the probabilities are 0.49997, 0.4999997, 0.499999999.

The previous table gives the area of the shaded region.



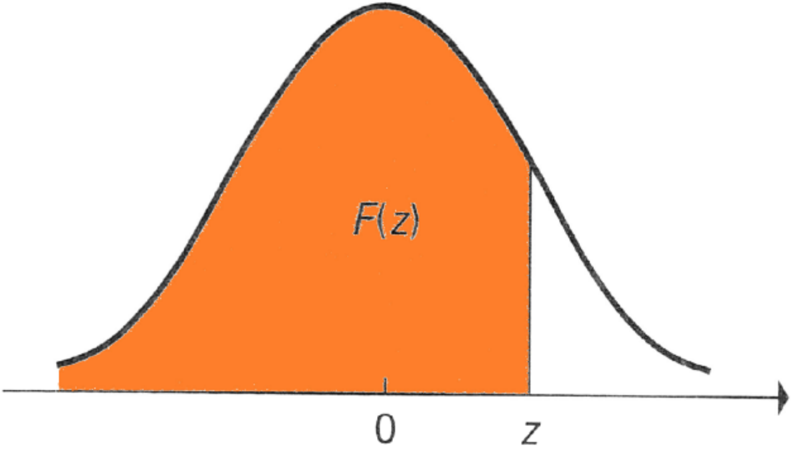
or

The Standard Normal Distribution Function N(0,1)

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0000	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1000	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2000	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3000	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4000	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5000	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6000	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7000	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7793	0.7823	0.7852
0.8000	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9000	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0000	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1000	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2000	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3000	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4000	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5000	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6000	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7000	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8000	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9000	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0000	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1000	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2000	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3000	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4000	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5000	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6000	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7000	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8000	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9000	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0000	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1000	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2000	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3000	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4000	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5000	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6000	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999

For z =4.0, 5.0 and 6.0 the probabilities are 0.99997, 0.9999997 and 0.999999999.

The previous table gives the area of the shaded region.



- If $a < 0$, then $\mathrm{N}(a) = 1 - \mathrm{N}(-a)$.
- If $a < b < 0$, then

$$\mathrm{N}(b) - \mathrm{N}(a) = 1 - \mathrm{N}(-b) - [1 - \mathrm{N}(-a)] = \mathrm{N}(-a) - \mathrm{N}(-b).$$

- if $a < 0$ and $b > 0$, then

$$\mathrm{N}(b) - \mathrm{N}(a) = \mathrm{N}(b) - [1 - \mathrm{N}(-a)] = \mathrm{N}(b) + \mathrm{N}(-a) - 1.$$

- Then, if $\tilde{x} \sim \mathrm{N}(\mu, \sigma^2)$,

$$P\{a \leq \tilde{x} \leq b\} = P\left\{\frac{a-\mu}{\sigma} \leq \frac{\tilde{x}-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right\} =$$

$$P\left\{\frac{a-\mu}{\sigma} \leq \tilde{z} \leq \frac{b-\mu}{\sigma}\right\} = \mathrm{N}\left(\frac{b-\mu}{\sigma}\right) - \mathrm{N}\left(\frac{a-\mu}{\sigma}\right),$$

where $\tilde{z} \sim \mathrm{N}(0,1)$ and $\mathrm{N}(\cdot)$ is the distribution function of the standard normal distribution.

- **Example.** If $\tilde{x} \sim \mathrm{N}(\mu, \sigma^2)$ with $\mu = 4$ and $\sigma^2 = 49$, then

$$P\{-2 \leq \tilde{x} \leq 5\} = P\left\{\frac{-2-4}{7} \leq \frac{\tilde{x}-\mu}{\sigma} \leq \frac{5-4}{7}\right\}$$

$$= P\{-0.8571 \leq \tilde{z} \leq 0.1429\} = \mathrm{N}(0.1429) - \mathrm{N}(-0.8571)$$

$$= \mathrm{N}(0.1429) + \mathrm{N}(0.8571) - 1 = 0.5568 + 0.8043 - 1 = 0.3611.$$

4.10. The multivariate normal distribution and its properties

The random vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^\intercal \colon\ (\Omega, \mathcal{F}, P) \to (\mathbb{R}^n, \mathcal{B})$ has the multivariate normal distribution ($\tilde{x} \sim \mathrm{MN}(\mu, \Sigma)$) if its density function is

$$f(x;\, \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^\intercal \Sigma^{-1} (x-\mu)\right], \ \forall\, x \in \mathbb{R}^n,$$

where

$$x = (x_1, x_2, \ldots, x_n)^\intercal,$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \vdots \\ \mu_n \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix},$$

Σ is a symmetric positive definite matrix, and $|\Sigma| > 0$ is the (absolute value of the) determinant of Σ.

- If $n = 1$, then $f(x;\, \mu, \Sigma) = n(x;\, \mu, \sigma)$, with $\sigma = \sqrt{\Sigma}$ as $\Sigma \in \mathbb{R}$ in this case.

- **Properties of the multivariate normal distribution.**

- **(1)** The marginal distribution of any sub-vector of the multivariate normally distributed random vector $(\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n)^\mathsf{T}$ will be multivariate normally distributed and the corresponding sub-vector of μ and the corresponding sub-matrix of Σ will be the mean vector and the variance-covariance matrix of that random sub-vector. In particular,

$$\tilde{x}_i \sim \mathrm{N}(\mu_i, \sigma_i^2), \text{ where } \sigma_i^2 \equiv \sigma_{ii},$$

and

$$\mathrm{Cov}(\tilde{x}_i, \tilde{x}_j) = \sigma_{ij}.$$

- However, it is not true that the joint distribution of (multivariate) normal random vectors/variables is multivariate normal.

- **(2)** Moment-generating function of the multivariate normal distribution:

$$M_{\tilde{x}}\underbrace{(t_1, t_2, ..., t_n)}_{t\in\mathbb{R}^n} = e^{t^\intercal \mu + \frac{1}{2} t^\intercal \Sigma t}.$$

- **(3)** The random vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^\intercal$ is multivariate normally distributed, $\tilde{x} \sim \mathrm{MN}(\mu, \Sigma)$, with Σ diagonal (i.e., $\sigma_{ij} = 0$, for all $i \neq j$) if and only if the random variables $\tilde{x}_1, \tilde{x}_2, ..., \tilde{x}_n$ are normally distributed and independent.
- Observe that, for Σ diagonal,

$$f(x_1, x_2, \ldots, x_n;\ \mu, \Sigma) = \prod_{i=1}^{n} n(x_i; \mu_i, \sigma_i), \text{ where } \sigma_i \equiv (\sigma_{ii})^{1/2}.$$

- Under multivariate normality, zero covariance implies independency! However, it is not true that the joint distribution of uncorrelated normally distributed random variables is multivariate normal.
- *Note:* When Σ is diagonal and σ_i^2 is the same for all i, we say that $\tilde{x}$ has a multivariate <u>circular</u> or <u>spherical</u> normal distribution.

- **(4)** Consider the following partitioned vectors/matrices of x, μ, Σ and $\tilde{x}$:

$$x = \begin{pmatrix} x_1 \\ \cdots \\ x_2 \end{pmatrix}, \; x_1 \in \mathbb{R}^{n_1}, \;\; x_2 \in \mathbb{R}^{n_2}, \;\; n_1 + n_2 = n,$$

$$\mu = \begin{pmatrix} \mu_1 \\ \cdots \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \vdots & \Sigma_{12} \\ \cdots & \cdots & \cdots \\ \Sigma_{21} & \vdots & \Sigma_{22} \end{pmatrix}, \qquad \Sigma_{12}^{\intercal} = \Sigma_{21},$$

$$\tilde{x} = \begin{pmatrix} \tilde{x}_1 \\ \cdots \\ \tilde{x}_2 \end{pmatrix}, \quad \text{where } \tilde{x}_1 \sim \mathrm{MN}(\mu_1, \Sigma_{11}) \text{ and } \tilde{x}_2 \sim \mathrm{MN}(\mu_2, \Sigma_{22}),$$

and $\Sigma_{ij} = \mathrm{Cov}(\tilde{x}_i, \tilde{x}_j) = \mathrm{E}\left[(\tilde{x}_i - \mu_i)(\tilde{x}_j - \mu_j)^{\intercal}\right]$ is a $n_i \times n_j$ matrix, for $i = 1, 2, \; j = 1, 2$.

- Then, the conditional density of the random vector $\tilde{x}_1$ given $\tilde{x}_2 = x_2$,

$$f_{\tilde{x}_1|\tilde{x}_2}\left(x_1|x_2; \mu_{\tilde{x}_1|\tilde{x}_2=x_2}, \Sigma_{\tilde{x}_1|\tilde{x}_2=x_2}\right) = \frac{f(x;\ \mu, \Sigma)}{f_{\tilde{x}_2}(x_2;\ \mu_2, \Sigma_{22})}, \quad \forall x_1 \in \mathbb{R}^{n_1},$$

  is the density of a multivariate normal random vector with mean $\mu_{\tilde{x}_1|\tilde{x}_2=x_2}$ and covariance matrix $\Sigma_{\tilde{x}_1|\tilde{x}_2=x_2}$.

- Moreover, the conditional mean of the random vector $\tilde{x}_1$ given $\tilde{x}_2 = x_2$ is the following:

$$\mu_{\tilde{x}_1|\tilde{x}_2=x_2} \equiv \mathrm{E}(\tilde{x}_1|\tilde{x}_2 = x_2) = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2) \in \mathbb{R}^{n_1}.$$

- *Note:* If $n_1 = n_2 = 1$, then

$$\mathrm{E}(\tilde{x}_1|\tilde{x}_2 = x_2) = \mu_1 + \frac{\sigma_{12}}{\sigma_2^2}(x_2 - \mu_2).$$

- Finally, let

$$\Sigma_{\tilde{x}_1|\tilde{x}_2=x_2} \equiv \mathrm{Var}(\tilde{x}_1|\tilde{x}_2 = x_2) =$$

$$\mathrm{E}([\tilde{x}_1 - \mathrm{E}(\tilde{x}_1|\tilde{x}_2 = x_2)][\tilde{x}_1 - \mathrm{E}(\tilde{x}_1|\tilde{x}_2 = x_2)]^{\intercal}|\tilde{x}_2 = x_2)$$

be the $n_1 \times n_1$ conditional covariance matrix of the random vector $\tilde{x}_1$ given $\tilde{x}_2 = x_2$. Then,

$$\Sigma_{\tilde{x}_1|\tilde{x}_2=x_2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21},$$

which does not depend on the value x_2 taken by the random vector $\tilde{x}_2$.

- *Note:* If $n_1 = n_2 = 1$, then

$$\mathrm{Var}(\tilde{x}_1|\tilde{x}_2 = x_2) = \sigma_1^2 - \frac{\sigma_{12}^2}{\sigma_2^2} = \sigma_1^2(1 - \rho^2),$$

so that the random variable $\mathrm{Var}(\tilde{x}_1|\tilde{x}_2) : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$ is a constant.

- **(5)** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^\intercal \sim \mathrm{MN}(\mu, \Sigma)$ and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)^\intercal$ is a non-zero vector of scalars, then

$$\alpha^\intercal \tilde{x} = \sum_{i=1}^{n} \alpha_i \tilde{x}_i \sim \mathrm{N}(\alpha^\intercal \mu, \alpha^\intercal \Sigma \alpha).$$

- Note that in the previous result the vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^\intercal$ has to be multivariate normal. It is not enough that each component of that vector be normal.

- **General Proposition.** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^\intercal \sim \mathrm{MN}(\mu, \Sigma)$ and $\tilde{y} = c + A\tilde{x}$ be an affine transformation of $\tilde{x}$, where $c \in \mathbb{R}^m$ is a column vector and A is $m \times n$ matrix with $\operatorname{rank}(A) = m \leq n$. Then, $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_m)^\intercal \sim \mathrm{MN}(c + A\mu, A\Sigma A^\intercal)$.

- The previous Proposition implies Property 5 (when $m = 1$, $A = \alpha^{\mathsf{T}}$, and $c = \underline{0}$) and Property 1. For the later, consider the following example: to extract the sub-vector $(\tilde{x}_1, \tilde{x}_2, \tilde{x}_4)^{\mathsf{T}}$ from the random vector $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^{\mathsf{T}}$ use the vector $c = \underline{0}$ and the $3 \times n$ matrix

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix},$$

which extracts the desired sub-vector directly.

- **From linearity to multivariate normality:**
- Consider the random variable $\tilde{y}$,

$$\tilde{y} = \alpha + \underbrace{\beta^\intercal \tilde{x}}_{=\tilde{x}^\intercal \beta} + \tilde{\varepsilon} = \alpha + \sum_{i=1}^{n} \beta_i \tilde{x}_i + \tilde{\varepsilon},$$

  where $\alpha \in \mathbb{R}$, $\beta = (\beta_1, \beta_2, \ldots, \beta_n)^\intercal \in \mathbb{R}^n$ is a column vector, $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^\intercal \sim \mathrm{MN}(\mu_x, \Sigma_x)$, $\tilde{\varepsilon} \sim \mathrm{N}(0, \sigma_\varepsilon^2)$, and $\tilde{\varepsilon}$ and $\tilde{x}$ are independent.
- We know from Property 3 above that the vector $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n, \tilde{\varepsilon})^\intercal$ is multivariate normal. Thus, from the General Proposition above, the random variable $\tilde{y}$ is normal since it is an affine transformation of the random variables appearing in the multivariate normal random vector $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n, \tilde{\varepsilon})^\intercal$. Finally, also from the General Proposition above, the random vector $(\tilde{y}, \tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^\intercal$ is multivariate normal.

- The mean μ_y and the variance σ_y^2 of $\tilde{y}$ are computed as follows:
- Mean:

$$\mu_y = \mathrm{E}(\tilde{y}) = \mathrm{E}\left(\alpha + \beta^{\mathsf{T}}\tilde{x} + \tilde{\varepsilon}\right) = \alpha + \beta^{\mathsf{T}}\mu_x + \mathrm{E}\left(\tilde{\varepsilon}\right).$$

$$= \alpha + \beta^{\mathsf{T}}\mu_x + 0 = \alpha + \underbrace{\beta^{\mathsf{T}}\mu_x}_{=\mu_x^{\mathsf{T}}\beta}. \tag{1}$$

- Variance:

$$\sigma_y^2 = \mathrm{Var}(\tilde{y}) = \mathrm{Var}(\alpha + \beta^{\mathsf{T}}\tilde{x} + \tilde{\varepsilon}) = \beta^{\mathsf{T}}\Sigma_x\beta + \sigma_\varepsilon^2. \tag{2}$$

- Therefore, $\tilde{y} \sim \mathrm{N}(\underbrace{\alpha + \beta^{\mathsf{T}}\mu_x}_{\mu_y}, \underbrace{\beta^{\mathsf{T}}\Sigma_x\beta + \sigma_\varepsilon^2}_{\sigma_y^2})$.

- Let us compute the conditional expectation of $\tilde{y}$ given $\tilde{x} = x$,

$$\mathrm{E}(\tilde{y}|\tilde{x} = x) = \mathrm{E}\left(\alpha + \beta^{\mathsf{T}}\tilde{x} + \tilde{\varepsilon}|\, \tilde{x} = x\right)$$

$$= \alpha + \mathrm{E}\left(\beta^{\mathsf{T}}\tilde{x}|\, \tilde{x} = x\right) + \mathrm{E}\left(\tilde{\varepsilon}|\tilde{x} = x\right)$$

$$= \alpha + \mathrm{E}\left(\beta^{\mathsf{T}}x\right) + \mathrm{E}\left(\tilde{\varepsilon}\right) = \alpha + \underbrace{\beta^{\mathsf{T}}x}_{=x^{\mathsf{T}}\beta},$$

where the third equality holds because $\tilde{\varepsilon}$ and $\tilde{x}$ are independent.

- Therefore, the conditional expectation $\mathrm{E}(\tilde{y}|\tilde{x} = x)$ is an affine transformation of x. This agrees with our previous Property 4, according to which,

$$\mathrm{E}(\tilde{y}|\tilde{x} = x) = \mu_y + \Sigma_{y,x}\Sigma_x^{-1}(x - \mu_x)$$

$$= \left(\mu_y - \Sigma_{y,x}\Sigma_{\tilde{x}}^{-1}\mu_x\right) + \Sigma_{y,x}\Sigma_x^{-1}x.$$

- Thus, the following equalities should be true:

$$\alpha = \mu_y - \Sigma_{y,x}\Sigma_x^{-1}\mu_x \tag{3}$$

and

$$\beta^\intercal = \Sigma_{y,x}\Sigma_x^{-1} \quad \text{or} \quad \beta = \Sigma_x^{-1}\Sigma_{y,x}^\intercal = \Sigma_x^{-1}\Sigma_{x,y}. \tag{4}$$

- We can check that (4) holds since the $1 \times n$ matrix $\Sigma_{y,x}$ satisfies

$$\Sigma_{y,x} = \mathrm{Cov}\left(\tilde{y}, \tilde{x}\right) = \mathrm{Cov}\left(\alpha + \beta^\intercal \tilde{x} + \tilde{\varepsilon}, \tilde{x}\right) = \beta^\intercal \Sigma_x$$

so that

$$\Sigma_{y,x}\Sigma_{\tilde{x}}^{-1} = \beta^\intercal \Sigma_x \Sigma_x^{-1} = \beta^\intercal.$$

- To check that (3) holds note that, from (1) and (4), we get

$$\alpha = \mu_y - \beta^\intercal \mu_x = \mu_y - \Sigma_{y,x}\Sigma_x^{-1}\mu_x.$$

- We can compute now the conditional variance of $\tilde{y}$ given $\tilde{x} = x$,

$$\mathrm{Var}(\tilde{y}|\tilde{x} = x) = \mathrm{Var}\left(\ \alpha + \beta^\intercal \tilde{x} + \tilde{\varepsilon} \middle|\ \tilde{x} = x\right)$$

$$= \mathrm{Var}\left(\alpha + \beta^\intercal x + \tilde{\varepsilon}\right) = \mathrm{Var}\left(\tilde{\varepsilon}\right) = \sigma_\varepsilon^2,$$

which agrees with our previous Property 4, according to which,

$$\mathrm{Var}(\tilde{y}|\tilde{x} = x) = \sigma_y^2 - \Sigma_{y,x}\Sigma_x^{-1}\Sigma_{x,y}$$

and, thus, $\mathrm{Var}(\tilde{y}|\tilde{x}= x)$ does not depend on the value x taken by the random vector $\tilde{x}$, i.e., the random variable $\mathrm{Var}(\tilde{y}|\tilde{x})$ is a constant.

- Therefore, we should have

$$\sigma_\varepsilon^2 = \sigma^2{}_y - \Sigma_{y,x}\Sigma_x^{-1}\Sigma_{x,y}.$$

- We can check that the previous equality holds indeed since, from (2) and (4), we get

$$\begin{aligned}\sigma_\varepsilon^2 = \sigma_y^2 - \beta^\intercal \Sigma_x \beta = \sigma_y^2 - \Sigma_{y,x}\Sigma_x^{-1}\Sigma_x\Sigma_x^{-1}\Sigma_{x,y} \\ = \sigma_y^2 - \Sigma_{y,x}\Sigma_x^{-1}\Sigma_{x,y}.\end{aligned}$$

- **From multivariate normality to linearity:**
- Assume that the random vector $(\tilde{y}, \underbrace{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n}_{\tilde{x}^\intercal})^\intercal$ is $\mathrm{MN}(\mu, \Sigma)$, with

$$\mu = (\mu_y, \underbrace{\mu_1, \mu_2, ..., \mu_n}_{\mu_x^\intercal})^\intercal$$

and

$$\Sigma = \begin{pmatrix} \sigma_y^2 & \Sigma_{y,x} \\ \Sigma_{x,y} & \Sigma_x \end{pmatrix}.$$

- Then, we know from Property 4 above that

$$\mathrm{E}(\tilde{y}|\tilde{x}=x)=\mu_y+\Sigma_{y,x}\Sigma_x^{-1}(x-\mu_x)=\left(\mu_y-\Sigma_{y,x}\Sigma_{\tilde{x}}^{-1}\mu_x\right)+\Sigma_{y,x}\Sigma_x^{-1}x.$$

or equivalently,

$$\mathrm{E}(\tilde{y}|\tilde{x}) = \left(\mu_y - \Sigma_{y,x}\Sigma_{\tilde{x}}^{-1}\mu_x\right) + \Sigma_{y,x}\Sigma_x^{-1}\tilde{x},$$

so that $\mathrm{E}(\tilde{y}|\tilde{x})$ is an affine transformation of $\tilde{x}$. Thus, since $\tilde{x} \sim \mathrm{MN}(\mu_x, \Sigma_x)$, the random variable $\mathrm{E}(\tilde{y}|\tilde{x})$ is normal as dictated by the General Proposition above.

- Define the random variable

$$\tilde{\varepsilon} = \tilde{y} - \mathrm{E}(\tilde{y}|\tilde{x}) = \tilde{y} - \left(\mu_y - \Sigma_{y,x}\Sigma_{\tilde{x}}^{-1}\mu_x\right) - \Sigma_{y,x}\Sigma_x^{-1}\tilde{x}.$$

- According to the General Proposition above, the random variable $\tilde{\varepsilon}$ is normal since it is an affine transformation of the random variables appearing in the multivariate normal random vector $(\tilde{y}, \tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)^{\mathsf{T}}$. Moreover, also from the General Proposition above, the random vector $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n, \tilde{\varepsilon})^{\mathsf{T}}$ is multivariate normal.

- Thanks to the theorem of total expectation, we can find the expectation for $\tilde{\varepsilon}$:

$$\mathrm{E}(\tilde{\varepsilon}) = \mathrm{E}\left[\tilde{y} - \mathrm{E}(\tilde{y}|\tilde{x})\right] = \mathrm{E}\left(\tilde{y}\right) - \mathrm{E}\left[\mathrm{E}(\tilde{y}|\tilde{x})\right] = \mathrm{E}\left(\tilde{y}\right) - \mathrm{E}\left(\tilde{y}\right) = 0.$$

- Moreover, the variance of $\tilde{\varepsilon}$ is

$$\sigma_\varepsilon^2 = \mathrm{Var}(\tilde{\varepsilon}) = \mathrm{Var}\left[\tilde{y} - \mathrm{E}(\tilde{y}|\tilde{x})\right]$$

$$= \mathrm{Var}\left[\tilde{y} - \left(\mu_y - \Sigma_{y,x}\Sigma_{\tilde{x}}^{-1}\mu_x\right) - \Sigma_{y,x}\Sigma_x^{-1}\tilde{x}\right]$$

$$= \mathrm{Var}\left(\tilde{y}\right) + \mathrm{Var}\left(\Sigma_{y,x}\Sigma_x^{-1}\tilde{x}\right) - 2\mathrm{Cov}\left(\tilde{y}, \Sigma_{y,x}\Sigma_x^{-1}\tilde{x}\right)$$

$$= \sigma_y^2 + \Sigma_{y,x}\Sigma_x^{-1}\Sigma_x\Sigma_x^{-1}\Sigma_{x,y} - 2\Sigma_{y,x}\Sigma_x^{-1}\Sigma_{x,y}$$

$$= \sigma_y^2 + \Sigma_{y,x}\Sigma_x^{-1}\Sigma_{x,y} - 2\Sigma_{y,x}\Sigma_x^{-1}\Sigma_{x,y} = \sigma_y^2 - \Sigma_{y,x}\Sigma_x^{-1}\Sigma_{x,y}.$$

- Therefore $\tilde{\varepsilon} \sim \mathrm{N}(0, \underbrace{\sigma_y^2 - \Sigma_{y,x}\Sigma_x^{-1}\Sigma_{x,y}}_{\sigma_\varepsilon^2})$.

- Moreover, from generalizing Exercise 30, part (c), of List 3, we know that $\tilde{\varepsilon} = \tilde{y} - \mathrm{E}(\tilde{y}|\tilde{x})$ has zero covariance with $\tilde{x}$, $\mathrm{Cov}\left(\tilde{\varepsilon}, \tilde{x}\right) = \underline{0}^\intercal \in \mathbb{R}^n$ and, thus, from Property 3 above, $\tilde{\varepsilon}$ and $\tilde{x}$ are independent.

- Note that, from the definition of $\tilde{\varepsilon}$, we have

$$\tilde{y} = \mathrm{E}(\tilde{y}|\tilde{x}) + \tilde{\varepsilon} = \left(\mu_y - \Sigma_{y,x}\Sigma_{\tilde{x}}^{-1}\mu_x\right) + \Sigma_{y,x}\Sigma_x^{-1}\tilde{x} + \tilde{\varepsilon}.$$

- Then, we can define the scalar $\alpha = \mu_y - \Sigma_{y,x}\Sigma_{\tilde{x}}^{-1}\mu_x$ and the column vector $\beta = \Sigma_x^{-1}\Sigma_{x,y} \in \mathbb{R}^n$ so that the previous equation becomes

$$\tilde{y} = \alpha + \underbrace{\beta^{\mathsf{T}}\tilde{x}}_{=\tilde{x}^{\mathsf{T}}\beta} + \tilde{\varepsilon}.$$

  where the random vector $\tilde{x}$ and the random variable $\tilde{\varepsilon}$ are independent, and $\tilde{\varepsilon} \sim \mathrm{N}(0, \sigma_y^2 - \Sigma_{y,x}\Sigma_x^{-1}\Sigma_{x,y})$.

- We can summarize all this section with the following proposition, which is related with regression analysis:
- **Proposition.** The random vector $(\tilde{y}, \underbrace{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n}_{\tilde{x}^\mathsf{T}})^\mathsf{T}$ has a multivariate normal distribution $\mathrm{MN}(\mu, \Sigma)$ with

$$\mu = (\mu_y, \underbrace{\mu_1, \mu_2, ..., \mu_n}_{\mu_x^\mathsf{T}})^\mathsf{T} \quad \text{and} \quad \Sigma = \begin{pmatrix} \sigma_y^2 & \Sigma_{y,x} \\ \Sigma_{x,y} & \Sigma_x \end{pmatrix}$$

if and only if there exist a scalar α, a vector of scalars $\beta = (\beta_1, \beta_2, \ldots, \beta_n)^\mathsf{T}$, and a random variable $\tilde{\varepsilon} \sim \mathrm{N}(0, \sigma_\varepsilon^2)$ such that

$$\tilde{y} = \alpha + \beta^\mathsf{T} \tilde{x} + \tilde{\varepsilon} \quad (\text{or} \;\; \tilde{y} = \alpha + \tilde{x}^\mathsf{T} \beta + \tilde{\varepsilon}),$$

where the random vector $\tilde{x} \sim \mathrm{MN}(\mu_x, \Sigma_x)$ and the random variable $\tilde{\varepsilon}$ are independent.
Moreover, the following equalities hold:

$$\alpha = \mu_y - \Sigma_{y,x} \Sigma_x^{-1} \mu_x \,, \quad \beta = \Sigma_x^{-1} \Sigma_{x,y} \,, \quad \text{and} \quad \sigma_\varepsilon^2 = \sigma_y^2 - \Sigma_{y,x} \Sigma_x^{-1} \Sigma_{x,y}.$$

- It is straightforward to generalize the previous proposition when $\tilde{y} = (\tilde{y}_1, \tilde{y}_2, ..., \tilde{y}_m)^\intercal$ is a random vector as follows:
- **Proposition.** The random vector $(\underbrace{\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_m}_{\tilde{y}^\intercal}, \underbrace{\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n}_{\tilde{x}^\intercal})^\intercal$ has a multivariate normal distribution $\mathrm{MN}(\mu, \Sigma)$ with

$$\mu = (\underbrace{\mu_{y_1}, \mu_{y_2}, ..., \mu_{y_m}}_{\mu_y^\intercal}, \underbrace{\mu_{x_1}, \mu_{x_2}, ..., \mu_{x_n}}_{\mu_x^\intercal})^\intercal \quad \text{and} \quad \Sigma = \begin{pmatrix} \Sigma_y & \Sigma_{y,x} \\ \Sigma_{x,y} & \Sigma_x \end{pmatrix}$$

if and only if there exist a vector $\alpha \in \mathbb{R}^m$, a $n \times m$ matrix B, and a random vector $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, ..., \tilde{\varepsilon}_m)^\intercal \sim \mathrm{MN}(\underline{0}, \Sigma_\varepsilon)$ such that

$$\tilde{y} = \alpha + \mathrm{B}^\intercal \tilde{x} + \tilde{\varepsilon} \quad (\text{or} \;\; \tilde{y}^\intercal = \alpha^\intercal + \tilde{x}^\intercal \mathrm{B} + \tilde{\varepsilon}^\intercal),$$

where the random vectors $\tilde{x} \sim \mathrm{MN}(\mu_x, \Sigma_x)$ and $\tilde{\varepsilon}$ are independent. Moreover, the following equalities hold:

$$\alpha = \mu_y - \Sigma_{y,x}\Sigma_x^{-1}\mu_x \,, \quad \mathrm{B} = \Sigma_x^{-1}\Sigma_{x,y} \,, \quad \text{and} \quad \Sigma_\varepsilon = \Sigma_y - \Sigma_{y,x}\Sigma_x^{-1}\Sigma_{x,y} \,.$$

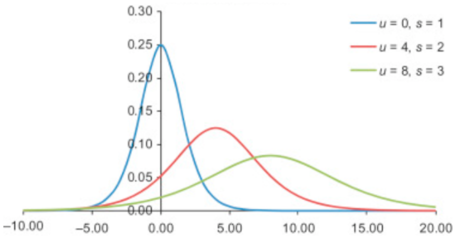
4.11. The logistic distribution

- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ has the logistic distribution (or is logistic) if its density is

$$l(x; m, s) = \frac{e^{-\left(\frac{x-m}{s}\right)}}{s\left[1 + e^{-\left(\frac{x-m}{s}\right)}\right]^2} = \frac{e^{\frac{x-m}{s}}}{s\left[1 + e^{\frac{x-m}{s}}\right]^2}, \quad \text{for all } x \in \mathbb{R},$$

where m is a location parameter (which is equal to the mean μ) and $s > 0$ is the scale parameter.

- Density:



- The logistic density is symmetric around m (Exercise).
- $$\mu = \mathrm{E}(\tilde{x}) = m \quad \text{and} \quad \mathrm{Var}(\tilde{x}) = \frac{s^2 \pi^2}{3}.$$
- Distribution function:

$$F(x; m.s) = \frac{1}{1 + e^{-\left(\frac{x-m}{s}\right)}} = \frac{e^{\frac{x-m}{s}}}{1 + e^{\frac{x-m}{s}}},$$

as $F'(x; m.s) = l(x; m.s)$.

- Quantile function:

$$Q(p; m, s) = m + s \ln \left(\frac{p}{1-p} \right).$$

- Standard logistic distribution when $m = 0$ and $s = 1$,

$$l(x; 0, 1) = \frac{e^{-x}}{(1 + e^{-x})^2} = \frac{e^x}{(1 + e^x)^2}, \quad \text{for all } x \in \mathbb{R}.$$

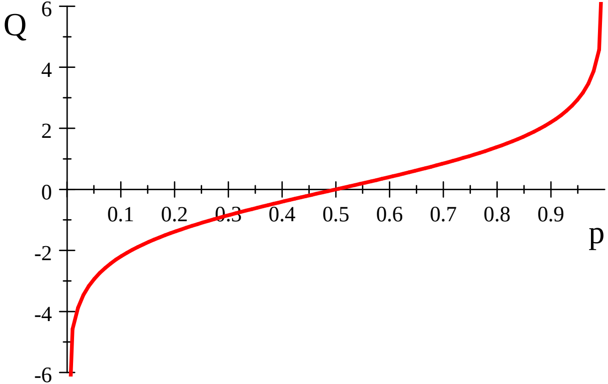
- Distribution function:

$$F(x; 0, 1) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x}.$$

- Quantile function:

$$Q(p; 0, 1) = \ln \left(\frac{p}{1 - p} \right).$$

- The quantile function of the standard logistic distribution is the **logit** function used in logit regressions/models:



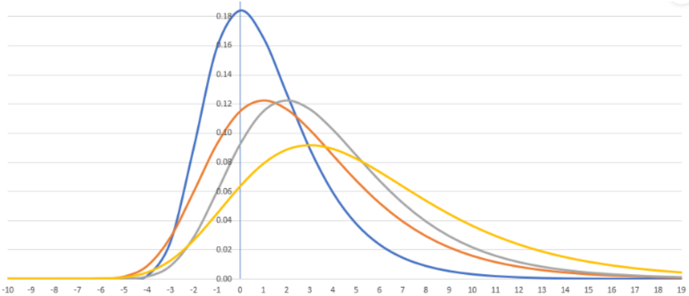
4.11. Generalized extreme value distributions

- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ has the generalized **extreme value type I** (or Gumbel) distribution if its density is

$$g(x; m, s) = \frac{1}{s} e^{-\left(\frac{x-m}{s}\right)} e^{-e^{-\left(\frac{x-m}{s}\right)}}, \text{ for all } x \in \mathbb{R},$$

where m is the location parameter and $s > 0$ is the scale parameter.

- Density:



- $$\mathrm{E}(\tilde{x}) = m + \gamma s \quad \text{and} \quad \mathrm{Var}(\tilde{x}) = \frac{s^2\pi^2}{6},$$

  where $\gamma = \lim_{n\to\infty}\left(-\ln n + \sum_{k=1}^{n}\frac{1}{k}\right) \approx 0.57722$ is the Euler(-Mascheroni) constant.

- (Standard) extreme value type I distribution when $m = 0$ and $s = 1$,

$$g(x;0,1) = e^{-x}e^{-e^{-x}}, \quad \text{for all } x \in \mathbb{R}.$$

- Distribution function:

$$G(x;m,s) = e^{-e^{-\left(\frac{x-m}{s}\right)}}, \quad \text{for all } x \in \mathbb{R},$$

$$G(x;0,1) = e^{-e^{-x}}, \quad \text{for all } x \in \mathbb{R}.$$

- Quantile function:

$$Q(p; m, s) = m - s \ln(-\ln p),$$

$$Q(p; 0, 1) = -\ln(-\ln p).$$

- The following proposition is very important for (logit) multinomial discrete choice models:

- **Proposition.** If $\tilde{x}_1 \sim G(m_1, s)$ and $\tilde{x}_2 \sim G(m_2, s)$ are independent, then $\tilde{x}_1 - \tilde{x}_2$ has the logistic distribution with density $l(m_1 - m_2, s)$.

- **Proof.** Exercise 27 in List 5.

- *Note* 1: If $\tilde{x}_1 \sim G(m_1, s)$ and $\tilde{x}_2 \sim G(m_2, s)$ are independent, then $\mathrm{E}(\tilde{x}_1 - \tilde{x}_2) = m_1 + \gamma s - (m_2 + \gamma s) = m_1 - m_2$.

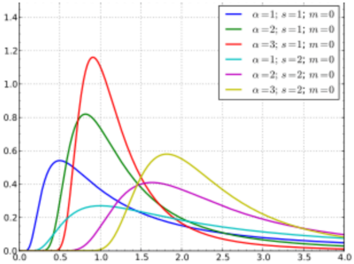
- *Note* 2: If $\tilde{x}_1 \sim G(m_1, s)$ and $\tilde{x}_2 \sim G(m_2, s)$ are independent, then $\tilde{x}_1 + \tilde{x}_2 \nsim l(m_1 + m_2, s)$ since $\mathrm{E}(\tilde{x}_1 + \tilde{x}_2) = m_1 + m_2 + 2\gamma s \neq m_1 + m_2$.

- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ has the generalized **extreme value type II** (or Fréchet) distribution if its density is

$$f(x; m, s, \alpha) = \begin{cases} \dfrac{\alpha}{s} \left(\dfrac{x-m}{s} \right)^{-1-\alpha} e^{-\left(\frac{x-m}{s}\right)^{-\alpha}} & \text{for } x > m \\ 0 & \text{otherwise,} \end{cases}$$

where m is the location parameter, $s > 0$ is the scale parameter, and $\alpha > 0$ is the shape parameter.

- Density:



- (Standard) extreme value type II distribution when $m = 0$, $s = 1$, and $\alpha > 0$,

$$f(x; 0, 1) = \alpha x^{-1-\alpha} e^{-x^{-\alpha}}, \quad \text{for all } x \in \mathbb{R}.$$

- Distribution function:

$$F(x; m, s) = \begin{cases} 0 & \text{for } x < m \\ e^{-\left(\frac{x-m}{s}\right)^{-\alpha}} & \text{for } x \geq m, \end{cases}$$

$$F(x; 0, 1) = \begin{cases} 0 & \text{for } x < 0 \\ e^{-x^{-\alpha}} & \text{for } x \geq 0. \end{cases}$$

- $$\mathrm{E}(\tilde{x}) = m + s\Gamma\left(1 - \frac{1}{\alpha}\right) \quad \text{if } \alpha > 1,$$
$$\mathrm{Var}(\tilde{x}) = s^2\left[\Gamma\left(1 - \frac{2}{\alpha}\right) + \left(\Gamma\left(1 - \frac{1}{\alpha}\right)\right)^2\right] \quad \text{if } \alpha > 2.$$

- Moments of order k are finite if $\alpha > k$.

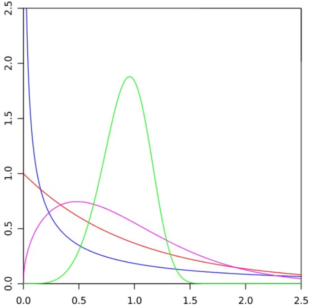
- Quantile function:
$$Q(p; m, s, \alpha) = m + s\left(-\ln p\right)^{-1/\alpha},$$
$$Q(p; 0, 1, \alpha) = \left(-\ln p\right)^{-1/\alpha}.$$

- The random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ has the generalized **extreme value type III** (or Weibull) distribution if its density is

$$f(x; m, s, \alpha) = \begin{cases} \dfrac{\alpha}{s}\left(\dfrac{x-m}{s}\right)^{\alpha-1} e^{-\left(\frac{x-m}{s}\right)^{\alpha}} & \text{for } x > m \\ 0 & \text{otherwise,} \end{cases}$$

where m is the location parameter, $s > 0$ is the scale parameter, and $\alpha > 0$ is the shape parameter.

- Density (with $m = 0$):



- (Standard) extreme value type III distribution when $m = 0$, $s = 1$, and $\alpha > 0$,

$$f(x; 0, 1, \alpha) = \begin{cases} \alpha x^{\alpha - 1} e^{-x^{\alpha}} & \text{for } x > 0 \\ \\ 0 & \text{otherwise,} \end{cases}$$

- For $m = 0$ and $\alpha = 1$, the Weibull distribution becomes exponential with the parameter s,

$$f(x; 0, s, 1) = \begin{cases} \dfrac{1}{s} e^{-x/s} & \text{for } x > 0 \\ \\ 0 & \text{otherwise.} \end{cases}$$

- Distribution function:

$$F(x; m, s, \alpha) = \begin{cases} 0 & \text{for } x < m \\ 1 - e^{-\left(\frac{x-m}{s}\right)^{\alpha}} & \text{for } x \geq m, \end{cases}$$

$$F(x; 0, 1, \alpha) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-x^{\alpha}} & \text{for } x \geq 0. \end{cases}$$

- $$\mathrm{E}(\tilde{x}) = m + s\Gamma\left(1 + \frac{1}{\alpha}\right),$$

$$\mathrm{Var}(\tilde{x}) = s^2 \left[\Gamma\left(1 + \frac{2}{\alpha}\right) + \left(\Gamma\left(1 + \frac{1}{\alpha}\right)\right)^2\right].$$

- Quantile function:

$$Q(p; m, s, \alpha) = m + s\left[-\ln(1-p)\right]^{1/\alpha},$$

$$Q(p; 0, 1, \alpha) = \left[-\ln(1-p)\right]^{1/\alpha}.$$

- Other famous absolutely continuous distributions appearing in the exercises are the **log-normal**, the **Cauchy**, the **Pareto**, the **Rayleigh**, the **Laplace**, and the **triangular**. In Chaper 7, we will also see the **Student's t** and the **F** distributions.

Informal proof of the formula of integration by change of variable

We want to compute the following Lebesgue integral:

$$\int_{[a,b]} f(x)dx.$$

Note that when we consider an interval $[a,b]$, we must have $a < b$. From now on, whenever we write the integral of a function w.r.t. the Lebesgue measure it should be understood that the function is not only integrable w.r.t. that measure, but also that it is Riemann integrable so that

$$\int_{[a,b]} f(x)dx = \int_a^b f(x)dx. \tag{1}$$

Let $x = g(y)$ and assume that g is differentiable on $g^{-1}([a,b])$. This requirement is fulfilled if we assume that the function $g : M \longrightarrow \mathbb{R}$ is differentiable and M is an open subset of $\mathbb{R}$ with $g^{-1}([a,b]) \subset M$ (or equivalently with $[a,b] \subset g(M)$).

Consider the inverse function $g^{-1} : [a,b] \longrightarrow \mathbb{R}$ so that $y = g^{-1}(x)$. This inverse function g^{-1} exists if and only if the function g restricted to $g^{-1}([a,b])$, i.e., $g : g^{-1}([a,b]) \longrightarrow [a,b]$, is bijective (or a one-to-one correspondence). That is, g must be strictly increasing ($g' > 0$ a.e.) or strictly decreasing ($g' < 0$ a.e.) on $g^{-1}([a,b])$. Therefore, if $x = a$ then $y = g^{-1}(a)$, whereas if $x = b$ then $y = g^{-1}(b)$.

From the theory of Riemann integration, recall that

$$\int_a^b f(x)dx = F(b) - F(a), \text{ where } F' = f \text{ on } [a,b].$$

Therefore,

$$\int_b^a f(x)dx = F(a) - F(b) = -\int_a^b f(x)dx. \tag{2}$$

Note that the primitive (or antiderivative) of $f(g(y)) \cdot g'(y)$ is $F(g(y))$ as follows from the chain rule,

$$\frac{dF(g(y))}{dy} = F'(g(y)) \cdot g'(y) = f(g(y)) \cdot g'(y).$$

Therefore,

$$\begin{aligned} \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y))g'(y)dy &= F(g(y))\Big|_{g^{-1}(a)}^{g^{-1}(b)} = F(g(g^{-1}(b))) - F(g(g^{-1}(a))) \\ &= F(b) - F(a) = \int_a^b f(x)dx. \end{aligned} \tag{3}$$

On the one hand, if $g' > 0$ a.e. (g is strictly increasing) then $g^{-1}(a) < g^{-1}(b)$ and, thus, $g^{-1}([a,b])$ is the interval $\left[g^{-1}(a), g^{-1}(b)\right]$. On the other hand, if $g' < 0$

a.e. (g is strictly decreasing) then $g^{-1}(a) > g^{-1}(b)$ and, thus, $g^{-1}([a,b])$ is the interval $\left[g^{-1}(b), g^{-1}(a)\right]$.

If $g' > 0$ a.e. then

$$\int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y))g'(y)dy = \int_{[g^{-1}(a),g^{-1}(b)]} f(g(y))g'(y)dy = \int_{g^{-1}([a,b])} f(g(y))\underbrace{g'(y)}_{>0}dy. \quad (4)$$

However, if $g' < 0$ a.e. then

$$\int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y))g'(y)dy = -\int_{g^{-1}(b)}^{g^{-1}(a)} f(g(y))g'(y)dy = \int_{g^{-1}(b)}^{g^{-1}(a)} f(g(y))\left[-g'(y)\right]dy$$

$$= \int_{[g^{-1}(b),g^{-1}(a)]} f(g(y))\left[-g'(y)\right]dy = \int_{g^{-1}([a,b])} f(g(y))\underbrace{\left[-g'(y)\right]}_{>0}dy, \quad (5)$$

where the first equality comes from (2).

Combining (1), (3), (4), and (5), we get

$$\int_{[a,b]} f(x)dx = \int_a^b f(x)dx = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(y))g'(y)dy = \int_{g^{-1}([a,b])} f(g(y))\left|g'(y)\right|dy.$$

The previous formula holds both for g strictly increasing and for g strictly decreasing.

1. The Gamma Distribution

1.A. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Proof.

1^{st} step: Let $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}\,dy$. We make the following change of variable:

$$y = g(z) = \frac{1}{2}z^2 \Rightarrow \frac{dy}{dz} = g'(z) = z, \quad \text{for } y > 0,\ z > 0.$$

$$\Rightarrow \Gamma(\alpha) = \int_0^\infty \frac{1}{2^{\alpha-1}} z^{2\alpha-2} e^{-\frac{1}{2}z^2} z dz = 2^{1-\alpha}\int_0^\infty z^{2\alpha-1}e^{-\frac{1}{2}z^2}dz.$$

$$\Longrightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{2}\int_0^\infty e^{-\frac{1}{2}z^2}dz.$$

2^{nd} step:

$$\left[\Gamma\left(\frac{1}{2}\right)\right]^2 = 2\left[\int_0^\infty e^{-\frac{1}{2}z^2}dz\right]\left[\int_0^\infty e^{-\frac{1}{2}x^2}dx\right] = 2\int_0^\infty\int_0^\infty e^{-\frac{1}{2}(z^2+x^2)}dzdx.$$

Let us make the change to polar coordinates:

$$\Rightarrow \int_0^\infty\int_0^\infty e^{-\frac{1}{2}(z^2+x^2)}dzdx = \int_{\mathbb{R}_+^2} e^{-\frac{1}{2}(z^2+x^2)}d(z,x) = \int_0^\infty\int_0^{\pi/2} e^{-\frac{1}{2}r^2}rd\theta dr$$

$$= \left(\int_0^\infty e^{-\frac{1}{2}r^2}rdr\right)\left(\int_0^{\pi/2}d\theta\right) = \left[-e^{-\frac{1}{2}r^2}\right]_0^\infty \cdot [\theta]_0^{\pi/2} = \frac{\pi}{2}.$$

$$\Longrightarrow \left[\Gamma\left(\frac{1}{2}\right)\right]^2 = 2\left(\frac{\pi}{2}\right) = \pi \Longrightarrow \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

$- - - - - - - - - - - - - - - - -$

1.B. $\displaystyle\int_0^\infty e^{-\frac{1}{2}z^2}dz = \sqrt{\frac{\pi}{2}}.$

Proof. Observe that,

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{2}\int_0^\infty e^{-\frac{1}{2}z^2}dz = \sqrt{\pi},$$

where the first equality comes from step 1 in 1.A. and the second one comes from step 2 in 1.A. Then,

$$\int_0^\infty e^{-\frac{1}{2}z^2}dz = \sqrt{\frac{\pi}{2}}.$$

– – – – – – – – – – – – – – – –

1.C. $\displaystyle\mu_r' = \frac{\beta^r\Gamma(\alpha+r)}{\Gamma(\alpha)}.$

Proof.

$$\mu_r' = \int_0^\infty x^r \frac{1}{\beta^\alpha\Gamma(\alpha)}\, x^{\alpha-1}e^{-x/\beta}\, dx.$$

Making the change of variable

$$x = g(y) = \beta y \ \Leftrightarrow y = g^{-1}(x) = \frac{x}{\beta}, \text{ so that } \frac{dx}{dy} = g'(y) = \beta > 0,$$

$$\begin{aligned}\mu_r' &= \int_0^\infty \beta^r y^r \frac{1}{\beta^\alpha\Gamma(\alpha)}\beta^{\alpha-1}y^{\alpha-1}e^{-y}\beta\, dy \\ &= \frac{\beta^r}{\Gamma(\alpha)}\underbrace{\int_0^\infty y^{\alpha+r-1}e^{-y}dy}_{\Gamma(\alpha+r)} = \frac{\beta^r\Gamma(\alpha+r)}{\Gamma(\alpha)}.\end{aligned}$$

– – – – – – – – – – – – – – – –

1.D. $M_{\tilde{x}}(t) = (1-\beta t)^{-\alpha}$ if $t < 1/\beta$.

Proof.

$$M_{\tilde{x}}(t) = \int_0^\infty e^{tx} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}\, dx = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-x\left(\frac{1}{\beta}-t\right)} dx.$$

Change of variable:

$$x = g(y) = \frac{y}{\frac{1}{\beta}-t} \Leftrightarrow y = g^{-1}(x) = x\left(\frac{1}{\beta}-t\right),$$

so that

$$\frac{dx}{dy} = g'(y) = \frac{1}{\frac{1}{\beta}-t} > 0, \quad \text{if } t < 1/\beta.$$

$$M_{\tilde{x}}(t) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\frac{1}{\beta}-t}\right)^{\alpha-1} e^{-y} \left(\frac{1}{\frac{1}{\beta}-t}\right) dy$$

$$= \frac{1}{\beta^\alpha \Gamma(\alpha)\left(\frac{1}{\beta}-t\right)^\alpha} \underbrace{\int_0^\infty y^{\alpha-1} e^{-y} dy}_{\Gamma(\alpha)} = \frac{1}{\beta^\alpha \left(\frac{1}{\beta}-t\right)^\alpha}$$

$$= \frac{1}{(1-\beta t)^\alpha} = (1-\beta t)^{-\alpha} \quad \text{if } t < 1/\beta.$$

– – – – – – – – – – – – – – – –

2. The Beta Distribution

$$\int_0^1 f(x;\alpha,\beta)dx = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}dx = 1, \quad \text{with } \alpha>0,\, \beta>0.$$

Proof. We just have to prove that

$$\Gamma(\alpha)\Gamma(\beta) = \Gamma(\alpha+\beta)\cdot \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx. \tag{1}$$

Note that

$$\Gamma(\alpha)\Gamma(\beta) = \left[\int_0^\infty y^{\alpha-1}e^{-y}dy\right]\cdot\left[\int_0^\infty z^{\beta-1}e^{-z}dz\right]$$

$$= \int_0^\infty\int_0^\infty y^{\alpha-1}z^{\beta-1}e^{-(y+z)}dydz = \int_{\mathbb{R}^2_{++}} y^{\alpha-1}z^{\beta-1}e^{-(y+z)}d(y,z).$$

Let us make the following change of variable $(y,z)=g(s,x)$, where $y=sx\in(0,\infty)$ and $z=s(1-x)\in(0,\infty)$. Note that the inverse function $(s,x)=g^{-1}(y,z)$ is given by $s=y+z\in(0,\infty)$ and $x=\dfrac{y}{y+z}\in(0,1)$. Thus,

$$\left|J_{g(s,x)}\right| = \left|\det\begin{bmatrix} x & s \\ 1-x & -s\end{bmatrix}\right| = |-s| = s,$$

so that

$$\Gamma(\alpha)\Gamma(\beta) = \int_{\mathbb{R}^2_{++}} y^{\alpha-1}z^{\beta-1}e^{-(y+z)}d(y,z) = \int_{(0,\infty)\times(0,1)} (sx)^{\alpha-1}\left(s(1-x)\right)^{\beta-1}e^{-s}sd(s,x) =$$

$$\int_0^\infty\int_0^1 (sx)^{\alpha-1}\left(s(1-x)\right)^{\beta-1}e^{-s}sdxds = \underbrace{\left[\int_0^\infty s^{\alpha-1}s^{\beta-1}e^{-s}sds\right]}_{\underbrace{\int_0^\infty s^{\alpha+\beta-1}e^{-s}ds}_{\Gamma(\alpha+\beta)}}\cdot\left[\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx\right].$$

Therefore, we have proved (1).

Note: The function

$$\Xi(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

is called the beta function. Thus, the density of the beta distribution can be written as

$$f(x;\alpha,\beta) = \frac{1}{\Xi(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}.$$

– – – – – – – – – – – – – – – – –

3. The Normal distribution

3.A. $\int_{-\infty}^{\infty} n(x;\mu,\sigma) = 1.$

Proof. Let

$$x = g(z) = \mu + \sigma z \Longleftrightarrow z = g^{-1}(x) = \frac{x-\mu}{\sigma} \Rightarrow \frac{dx}{dz} = g'(z) = \sigma > 0.$$

Then,

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sigma} e^{-\frac{1}{2}z^2} \sigma\, dz$$
$$= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}z^2}\, dz = \sqrt{\frac{2}{\pi}} \cdot \sqrt{\frac{\pi}{2}} = 1,$$

where the second equality comes from the symmetry of the function $e^{-\frac{1}{2}z^2}$ and the third equality comes from 1.B.

— — — — — — — — — — — — — — — —

3.B. $M_{\tilde{x}}(t) = e^{\mu t + \frac{1}{2}t^2\sigma^2}$.

Proof. We will use in this proof a technique called "completing the square".

$$M_{\tilde{x}}(t) = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[-2xt\sigma^2 + (x-\mu)^2]} dx$$

Observe that

$$-2xt\sigma^2 + (x-\mu)^2 = [x - (\mu + t\sigma^2)]^2 - 2\mu t\sigma^2 - t^2\sigma^4.$$

Then, the last integral equals

$$e^{\frac{1}{2\sigma^2}[2\mu t\sigma^2 + t^2\sigma^4]} \underbrace{\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[x-(\mu+t\sigma^2)]^2} dx}_{1} = e^{\mu t + \frac{1}{2}t^2\sigma^2},$$

since $\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[x-(\mu+t\sigma^2)]^2} dx$ is the integral of the normal density function with parameters $\mu + t\sigma^2$ and σ, $n(x;\mu + t\sigma^2, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[x-(\mu+t\sigma^2)]^2}$.

The Standard Normal Distribution Function N(0,1)

z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Also, for z = 4.0, 5.0, and 6.0 the probabilities are 0.49997, 0.4999997, 0.499999999.

Exercises. Probability and Statistics. IDEA.
4. Special Distributions

1. If $\tilde{x}$ has the discrete uniform distribution with $f(x) = 1/k$ for $x = 1, 2, ..., k$ show that

   (a) its mean is $\mu = \dfrac{k+1}{2}$;

   (b) its variance is $\sigma^2 = \dfrac{k^2-1}{12}$.

2. If $\tilde{x}$ has the discrete uniform distribution $f(x) = 1/k$ for $x = 1, 2, ..., k$, show that its moment-generating function is given by

$$M_{\tilde{x}}(t) = \frac{e^t\left(1-e^{kt}\right)}{k\left(1-e^t\right)}$$

   Also find the mean of this distribution by evaluating $\lim\limits_{t\to 0} M'_{\tilde{x}}(t)$, and compare the result obtained in part (a) of Exercise 1.

3. A multiple-choice test consists of eight questions and three answers to each question (of which only one is correct). If a student answers each question by rolling a balanced dice and checking the first answer if he gets a 1 or 2, the second answer if he gets a 3 or 4, and the third answer if he gets a 5 or 6, what is the probability that he will get exactly four correct answers?

4. A quality control engineer wants to check whether (in accordance with specifications) 95 percent of the electronic components shipped by her company are without flaws. To this end, she randomly selects 20 components from each large lot ready to be shipped and passes the lot if they are all without flaws; otherwise each component in the lot is checked. Assuming that the lots are so large that we can use the binomial distribution as an approximation, find the probabilities that she will commit the error of

   (a) holding a lot for a complete check even though 95 percent of the components are without flaws;

   (b) passing a lot even though only 90 percent of the components are without flaws;

   (c) passing a lot even though only 80 percent of the components are without flaws;

   (d) passing a lot even though only 70 percent of the components are without flaws.

5. Show that the moment-generating function of the geometric distribution is given by

$$M_{\tilde{x}}(t) = \frac{\theta e^t}{1-e^t(1-\theta)}$$

   in a neighborhood of $t = 0$ and use it to verify that $\mu = \dfrac{1}{\theta}$ and $\sigma^2 = \dfrac{1-\theta}{\theta^2}$.

6. According to the Mendelian theory of heredity, if plants with round yellow seeds are crossbred with plants with wrinkled green seeds, the probabilities of getting a plant that produces round yellow seeds, wrinkled yellow seeds, round green seeds, or wrinkled green seeds are, respectively, 9/16, 3/16, 3/16, and 1/16. What is the probability that among nine plants thus obtained there will be four that produce round yellow seeds, two that produce wrinkled yellow seeds, three that produce round green seeds, and none that produce wrinkled green seeds?

7. Show that if a random variable has a uniform density on the interval (α, β), the probability that it will take on a value less than $\alpha + p(\beta - \alpha)$ is equal to p.

8. A random variable is said to have a **Cauchy distribution** if its density is given by

$$f(x) = \frac{\beta/\pi}{(x-\alpha)^2 + \beta^2} \qquad \text{for } -\infty < x < \infty$$

Show that for this distribution the moment μ_1' does not exist and the moment μ_2' is not finite. Note that this means that a random variable having the Cauchy distribution does not belong to the L^1 space. For the purpose of this example set $\alpha = 0$ and $\beta = 1$.

Hint: If $f(x) = \arctan(x)$, where the function $\arctan(\cdot) : \mathbb{R} \longrightarrow (-\pi/2, \pi/2)$ is the inverse function of $\tan(\cdot) : (-\pi/2, \pi/2) \longrightarrow \mathbb{R}$, then the derivative of f is $f'(x) = \dfrac{1}{1+x^2}$. Recall also that $\tan\left(\dfrac{\pi}{4}\right) = 1$, $\tan\left(-\dfrac{\pi}{4}\right) = -1$, $\lim_{\theta \to \pi/2} \tan(\theta) = \infty$, $\lim_{\theta \to -\pi/2} \tan(\theta) = -\infty$ and $\tan(0) = 0$.

9. Show that if a random variable has an exponential density with parameter θ, the probability that it will take on a value less than $-\theta \cdot \ln(1-p)$ is equal to p for $0 \leq p < 1$.

10. A random variable $\tilde{x}$ has a **Pareto distribution** if its density is given by

$$f(x) = \begin{cases} \dfrac{\alpha k^\alpha}{x^{\alpha+1}} & \text{for } x > k \\ \\ 0 & \text{elsewhere,} \end{cases}$$

with $\alpha > 0$ and $k > 0$. Show that the moment μ_r' is finite only if $r < \alpha$. Note that this means that $\tilde{x} \in \mathrm{L}^r$ only for $r < \alpha$.

11. Karl Pearson, one of the founders of modern statistics, showed that the differential equation

$$\frac{f'(x)}{f(x)} = \frac{d - x}{a + bx + cx^2}$$

yields (for appropriate values of the constants a, b, c, and d) the densities of the most important distributions of statistics. Verify that the differential equation gives the density f of

(a) the gamma distribution when $a = c = 0$, $b > 0$, and $d > -b$;

(b) the exponential distribution when $a = c = d = 0$ and $b > 0$;

(c) the beta distribution when $a = 0$, $b = -c$, $\frac{d-1}{b} < 1$, and $\frac{d}{b} > -1$;

(d) the normal distribution when $b = c = 0$ and $a > 0$.

12. A point X is chosen on the segment AB, whose midpoint is C and whose length is a. If $\tilde{x}$, the distance from X to A, is a random variable having a uniform density on the interval $(0, a)$, what is the probability that AX, BX, and AC will form a triangle?

13. The mileage (in thousands of miles) which car owners get with a certain kind of tire is a random variable having an exponential distribution with $\theta = 40$. Find the probabilities that one of this tires will last

    (a) at least 20 000 miles;

    (b) at most 30 000 miles.

14. Twice more differentiating the moment-generating function of the normal distribution, verify that $\mu_3 = 0$ and $\mu_4 = 3\sigma^4$.

15. If $\tilde{x}$ is a random variable having a normal distribution, what are the probabilities of getting a value

    (a) within one standard deviation of the mean;

    (b) within two standard deviations of the mean;

    (c) within three standard deviations of the mean;

    (d) within four standard deviations of the mean?

16. Suppose that during periods of transcendental meditation the reduction of a person's oxygen consumption is a random variable having a normal distribution with $\mu = 37.6$ cc per minute and $\sigma = 4.6$ cc per minute. Find the probabilities that during a period of transcendental meditation a person's oxygen consumption will be reduced by

    (a) at least 44.5 cc per minute;

    (b) at most 35.0 cc per minute;

    (c) anywhere from 30.0 to 40.0 cc per minute.

17. If $\tilde{x}$ and $\tilde{y}$ have the circular normal distribution with $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 12$, find

    (a) the probability of getting a point (x, y) inside the circle $x^2 + y^2 \leq 36$;

    (b) the value of c for which the probability of getting a point (x, y) inside the circle $x^2 + y^2 \leq c^2$ is 0.80.

18. The joint distribution of the random variables $\tilde{x}_1,\ \tilde{x}_2,\ \tilde{x}_3$ is $\mathrm{MN}(\mu, \Sigma)$ with

$$\mu = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 9 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 2 \end{pmatrix}.$$

(a) Compute $P\left\{\tilde{x}_1 + \tilde{x}_2 + \tilde{x}_3 \le 3\right\}$.

(b) Find the conditional density of $(\tilde{x}_1,\ \tilde{x}_2)^\intercal$ given $\tilde{x}_3 = 0$.

(c) Find a so that $\tilde{z} = a\tilde{x}_1 + \tilde{x}_2$ is independent of $\tilde{x}_3$.

19. Find the moment-generating function of the Pascal (or negative binomial) distribution having the probability function $b^*(x; k, \theta)$ for $x = k, k+1, \ldots$ and use this moment-generating function to prove that the mean of the Pascal distribution is k/θ. *Hint:* You should use the results of Exercise 6 of List 1.

20. The probability of giving birth to a male is 0.51. Abstracting from multiple pregnancies, find

(a) the probability that a couple had to wait until the fourth child in order to get exactly one male (i.e., the first male is obtained in the fourth trial);

(b) the probability that a couple had to wait until the fourth child in order to get exactly two males (i.e., the second male is obtained in the fourth trial);

(c) the expected number of children that a couple should have in order to get exactly two males.

Hint: For (c) you should use the result of the previous Exercise 19 of this list.

21. The random vector $(\tilde{x}, \tilde{y})$ is distributed according to the following density function:

$$f(x,y) = \begin{cases} \dfrac{1}{\pi} e^{-\frac{1}{2}\left(x^2+y^2\right)} & \text{if } x > 0 \text{ and } y \le 0, \text{ or if } x \le 0 \text{ and } y > 0 \\[2ex] 0 & \text{otherwise.} \end{cases}$$

(a) Show that the marginal distributions of the random variables $\tilde{x}$ and $\tilde{y}$ are standard normal.

(b) Are the random variables $\tilde{x}$ and $\tilde{y}$ independent?

(c) Is the random vector $(\tilde{x}, \tilde{y})$ multivariate normally distributed? Any comment about these results?

22. The distribution of a random vector $(\tilde{x}, \tilde{y})$ is characterized by the following density function:

$$f(x,y) = \begin{cases} k\left(x^2+y^2\right)^{1/2} & \text{when } (x,y) \in A \\[2ex] 0 & \text{when } (x,y) \notin A, \end{cases}$$

Find the value of the constant k

(a) when $A = \{(x, y) \in \mathbb{R}^2 \mid 4 < x^2 + y^2 < 9,\ 0 < y < x\}$;

(b) when $A = \{(x, y) \in \mathbb{R}^2 \mid 9 < x^2 + y^2 < 16,\ y > 0\}$.

23. Find the area of a circle with radius r by means of the following two methods:

(a) Compute the integral with respect to the Lebesgue measure on $\mathbb{R}^2$,

$$\int_C d(x, y), \quad \text{where } C = \left\{(x, y) \in \mathbb{R}^2 \,\middle|\, x^2 + y^2 \le r^2\right\},$$

through a change to polar coordinates. Why this integral is equal to the area of a circle?

(b) Compute an integral with respect to the Lebesgue measure on $\mathbb{R}$ using the change of variable $x = r \sin \alpha$.

Hint: Remember that

$$\cos^2 \alpha = \frac{1 + \cos(2\alpha)}{2} \quad \text{since} \quad \cos(\alpha + \beta) = \cos \alpha \cdot \cos \beta - \sin \alpha \cdot \sin \beta.$$

24. Find the following integral using the change to polar coordinates ($x = r \cos \alpha$, $y = r \sin \alpha$):

$$\int_C \frac{e^{-\left(x^2+y^2\right)}}{\pi} d(x, y), \quad \text{where } C \text{ is the region of } \mathbb{R}^2 \text{ such that } x^2 + y^2 \le 36,$$

that is, the region C is a circle with its radius length equal to 6 and its center located at the point $(0, 0)$.

25. The four blood groups are distributed in a given (very large) population according to the following percentages:

$$0 = 45\% \qquad A = 43\% \qquad B = 8\% \qquad AB = 4\%$$

If we randomly and independently choose people, compute the probability:

(a) That when we pick 10 individuals, 4 of them will be of the A group, 2 of the B group, and 1 of the AB group;

(b) That when we pick three individuals, one of them will be able to donate blood to the other two (the 0 group is the universal donor and the AB group is the universal receiver);

(c) That we will have to wait until the fifth volunteer to find someone of the AB group, if volunteers are called for an emergency.

26. We have 7 white balls, 3 red balls, and 2 blue balls in an urn.

    (a) What is the probability that, if we pick 5 balls <u>without replacement</u>, we will get 3 white balls, 2 red balls and none blue?

    (b) What is the probability that, if we pick 5 balls <u>with replacement</u> (that is, if we put back in the urn the ball we have picked before the next extraction), we will also get 3 white balls, 2 red balls and none blue?

    (c) Which distribution is used to solve for part (a)? The multinomial or the multivariate hypergeometric? And for part (b)? Why?

27. (a) Prove that

$$\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}.$$

    (b) Find $\int_0^{\infty} e^{-x^2} dx$.

28. Suppose that the random variable $\tilde{z}$ has a standard normal distribution, $\tilde{z} \sim \mathrm{N}(0,1)$. Let $\tilde{x}$ be a discrete random variable with the following probability function:

$$f(x) = \begin{cases} \dfrac{1}{2} & \text{for } x = 1 \\ \\ \dfrac{1}{2} & \text{for } x = -1. \end{cases}$$

    Assume that $\tilde{z}$ and $\tilde{x}$ are independent. Define the random variable $\tilde{y} = \tilde{x} \cdot \tilde{z}$.

    Prove that

    (a) the random variables $\tilde{z}$ and $\tilde{y}$ are not multivariate normally distributed. *Hint:* In order to prove it you could compute the probability $P\{\tilde{z} + \tilde{y} = 0\}$ and extract the appropriate conclusion;

    (b) the random variable $\tilde{y}$ has a standard normal distribution. *Hint:* you should use the theorem of total probability to compute the probability $P\{\tilde{y} \leq y\}$ and check that this probability is equal to the distribution function $\mathrm{N}(y)$ of a standard normal random variable;

    (c) $\mathrm{Cov}(\tilde{z}, \tilde{y}) = 0$, i.e., the random variables $\tilde{z}$ and $\tilde{y}$ are uncorrelated. *Hint:* you should use the fact that $\mathrm{E}(\tilde{z} \cdot \tilde{y}) = \mathrm{E}(\mathrm{E}(\tilde{z} \cdot \tilde{y} \,|\tilde{x}))$;

    (d) the random variables $\tilde{z}$ and $\tilde{y}$ are not independent. *Hint:* In order to prove it you could compute $P(\{\tilde{y} > 1\} \cap \{\tilde{z} \in [0, 1/2)\})$ and extract the appropriate conclusion.

    Note: this exercise provides an example of uncorrelated (but not independent) normal random variables ($\tilde{z}$ and $\tilde{y}$) that are not multivariate normal.

29. Prove that the mean of the beta distribution with parameters α and β is $\dfrac{\alpha}{\alpha+\beta}$.

30. The distribution of a random vector $(\tilde{x}, \tilde{y})$ is characterized by the following density function:

$$f(x,y) = \begin{cases} \dfrac{k}{(x^2+y^2)^3} & \text{when } (x,y) \in D \\ 0 & \text{when } (x,y) \notin D, \end{cases}$$

where $D = \{(x,y) \in \mathbb{R}^2 \,|\, 1 < x^2 + y^2 < 3,\ x < 0,\ y < 0\}$.

Find the value of the constant k.

31. We randomly and independently pick two points A and B on a circumference with radius r. Find the expected value of the distance between A and B.

32. Let $\tilde{z}$ be a standard normal random variable. Prove that

$$P\{\tilde{z} \geq b\} = P\{\tilde{z} \leq -b\} \leq e^{-b^2/2}, \quad \text{for all } b \geq 0.$$

Hint: for this exercise use the result in part (a) of Exercise 33 of List 3.

33. Assume that there are 6 balls in a box and 3 of them are white, 2 are black, and 1 is green.

(a) We extract 4 balls with no replacement from the box. What is the probability of extracting 2 white balls and 2 black balls.

(b) We extract 4 balls with replacement from the box. What is the probability of extracting 2 white balls and 2 black balls.

34. Prove that the logistic density function is symmetric around the location parameter m. That is, $l(m-x,m,s) = l(m+x,m,s)$ for all $x \in \mathbb{R}$.

– – – – – – – – – – – – – –