

3. Expectation

3.1. Mathematical expectation

- **Definition.** Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\overline{\mathbb{R}}, \mathcal{B})$ be an extended random variable with the distribution $P_{\tilde{x}}$. The (mathematical or arithmetic) expectation (or expected value or mean) of \tilde{x} (or of its distribution) is defined by

$$E(\tilde{x}) = \int_{\Omega} \tilde{x}(\omega) dP(\omega) \quad \left(= \int_{\Omega} \tilde{x} dP \right)$$

or

$$E(\tilde{x}) = \int_{\overline{\mathbb{R}}} x dP_{\tilde{x}}(x) \quad \left(= \int_{\overline{\mathbb{R}}} x dP_{\tilde{x}} \right), \quad (\star)$$

whenever these integrals exist, i.e., whenever they are not of the form $\infty - \infty$ (indefinite).

- Obviously, if $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ is a (real-valued) random variable, we can replace $\overline{\mathbb{R}}$ by \mathbb{R} in (\star) and use the following notation:

$$E(\tilde{x}) = \int_{\mathbb{R}} x dF_{\tilde{x}}(x) \quad \left(= \int_{\mathbb{R}} x dF_{\tilde{x}} = \int_{\mathbb{R}} x dP_{\tilde{x}} \right).$$

- Recall that

$$E(\tilde{x}) = E(\tilde{x}^+) - E(\tilde{x}^-) = \int_{\{\tilde{x} \geq 0\}} \tilde{x} dP - \int_{\{\tilde{x} \leq 0\}} (-\tilde{x}) dP.$$

- From now on, when we write an expectation, it should be understood that the expectation exists.
- If $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ is a discrete random variable with probability function $f_{\tilde{x}} : \tilde{x}(\Omega) \longrightarrow [0, 1]$,

$$E(\tilde{x}) = \sum_{x \in \tilde{x}(\Omega)} x f_{\tilde{x}}(x).$$

- If $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ is an absolutely continuous random variable with density $f_{\tilde{x}} : \mathbb{R} \longrightarrow \overline{\mathbb{R}}$,

$$E(\tilde{x}) = \int_{\mathbb{R}} x f_{\tilde{x}}(x) dx.$$

- See the handout for the proof of the previous "intuitive" formula.

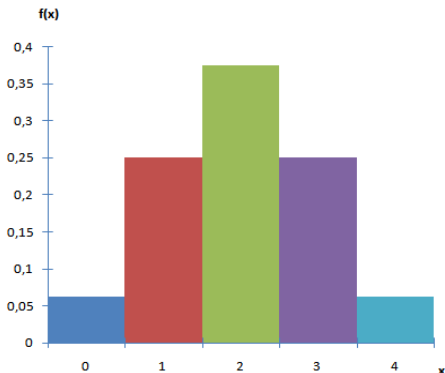
● **Example 1:**

Let \tilde{x} be the number of heads when tossing 4 coins.

$$f_{\tilde{x}}(x) = \begin{cases} 1/16 & \text{for } x = 0 \\ 4/16 & \text{for } x = 1 \\ 6/16 & \text{for } x = 2 \\ 4/16 & \text{for } x = 3 \\ 1/16 & \text{for } x = 4, \end{cases}$$

or

$$f_{\tilde{x}}(x) = \frac{1}{16} \binom{4}{x}, \quad \text{for } x = \underbrace{0, 1, 2, 3, 4}_{\tilde{x}(\Omega)}.$$



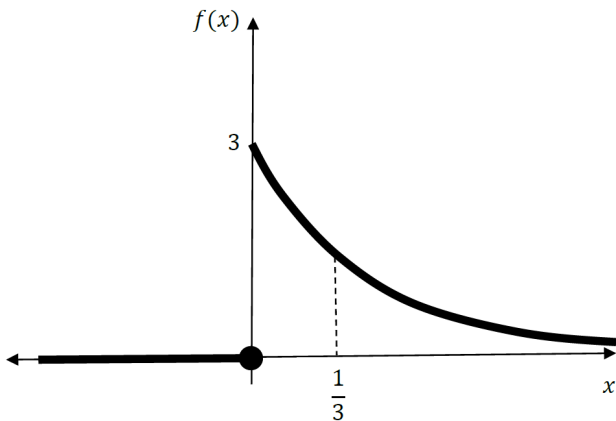
$$E(\tilde{x}) = \left(0 \cdot \frac{1}{16}\right) + \left(1 \cdot \frac{4}{16}\right) + \left(2 \cdot \frac{6}{16}\right) + \left(3 \cdot \frac{4}{16}\right) + \left(4 \cdot \frac{1}{16}\right) = 2.$$

- The median of \tilde{x} (or of its distribution) is the real number m for which $P\{\tilde{x} \leq m\} \geq 1/2$ and $P\{\tilde{x} \geq m\} \geq 1/2$. If there are two values satisfying this definition, we take the average of them. Obviously, $m = 2 = E(\tilde{x})$ in this example.

● **Example 2:**

The random variable \tilde{x} has the following density:

$$f(x) = \begin{cases} 3e^{-3x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$



$$\begin{aligned}
E(\tilde{x}) &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-\infty}^0 x \cdot 0 \cdot dx + \int_0^{\infty} \underbrace{x}_{H(x)} \underbrace{3e^{-3x}}_{g(x)} dx \\
&= 0 + [H(x)G(x)]_0^{\infty} - \int_0^{\infty} h(x)G(x) dx \\
&= \underbrace{[-xe^{-3x}]_0^{\infty}}_{-\lim_{x \rightarrow \infty} \left(\frac{x}{e^{3x}}\right) + 0 = -\lim_{x \rightarrow \infty} \left(\frac{1}{3e^{3x}}\right) = 0} - \int_0^{\infty} \underbrace{1}_{h(x)} \cdot \underbrace{(-e^{-3x})}_{G(x)} dx = \left[\frac{e^{-3x}}{-3} \right]_0^{\infty} = \frac{1}{3},
\end{aligned}$$

\uparrow
L'Hôpital's rule

where $H' = h$ and $G' = g$.

- Since the distribution of \tilde{x} is absolutely continuous in this example, its median m satisfies $P\{\tilde{x} \leq m\} = P\{\tilde{x} \geq m\}$ and, thus,

$$\begin{aligned}
P\{\tilde{x} \leq m\} &= F(m) = \int_0^m 3e^{-3x} dx = 1 - e^{-3m} = \frac{1}{2} \\
\implies m &= \frac{1}{3} \ln 2 = 0.23105 < \frac{1}{3} = E(\tilde{x}).
\end{aligned}$$

3.2. Properties of mathematical expectation

- **Proposition (the law of the unconscious statistician - LOTUS).**

Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$ be a random object with the distribution $P_{\tilde{x}}$ and let $g : (\Omega', \mathcal{F}') \longrightarrow (\mathbb{R}, \mathcal{B})$ be a Borel measurable function. If $\tilde{y} = g(\tilde{x})$, then

$$E(\tilde{y}) = \int_{\Omega} g(\tilde{x}) dP = \int_{\Omega'} g(x) dP_{\tilde{x}}(x).$$

- **Corollary 1.** Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B})$ be a random vector with the distribution $P_{\tilde{x}}$ and let $g : (\mathbb{R}^n, \mathcal{B}) \longrightarrow (\mathbb{R}, \mathcal{B})$ be a Borel measurable function. If $\tilde{y} = g(\tilde{x})$, then

$$E(\tilde{y}) = E(g(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)) = \int_{\Omega} g(\tilde{x}) dP = \int_{\mathbb{R}^n} \underbrace{g(x_1, x_2, \dots, x_n)}_{x \in \mathbb{R}^n} dP_{\tilde{x}}(x_1, x_2, \dots, x_n).$$

- **Corollary 2.** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B})$ be a random vector with the distribution $P_{\tilde{x}}$, then

$$E(\tilde{x}_i) = \int_{\Omega} \tilde{x}_i dP = \int_{\mathbb{R}^n} x_i dP_{\tilde{x}}(x_1, x_2, \dots, x_n) \quad \text{for } i = 1, 2, \dots, n.$$

- **Proof:** Just make $g = p_i$ in Corollary 1, where p_i is the projection map to the i th coordinate, that is, $p_i(x_1, x_2, \dots, x_n) = x_i$.
- **Corollary 3.** Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ be a random variable with the distribution $P_{\tilde{x}}$ and let $g : (\mathbb{R}, \mathcal{B}) \longrightarrow (\mathbb{R}, \mathcal{B})$ be a Borel measurable function. If $\tilde{y} = g(\tilde{x})$, then

$$E(\tilde{y}) = \int_{\Omega} g(\tilde{x}) dP = \int_{\mathbb{R}} g(x) dP_{\tilde{x}}(x).$$

- **Geometric** and **Harmonic** expectation (or mean) of a non-negative extended random variable, $\tilde{x} \geq 0$:

$$\text{GE}(\tilde{x}) = \exp(\text{E}[\ln \tilde{x}]) \quad \text{and} \quad \text{HE}(\tilde{x}) = (\text{E}[\tilde{x}^{-1}])^{-1} = \frac{1}{\text{E}[1/\tilde{x}]}.$$

- If $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ is a discrete random variable with probability function $f_{\tilde{x}} : \tilde{x}(\Omega) \longrightarrow [0, 1]$,

$$\begin{aligned} \text{GE}(\tilde{x}) &= \exp(\text{E}[\ln \tilde{x}]) = \exp\left(\sum_{x \in \tilde{x}(\Omega)} f_{\tilde{x}}(x) \cdot \ln x\right) \\ &= \exp\left(\sum_{x \in \tilde{x}(\Omega)} \ln \left[x^{f_{\tilde{x}}(x)}\right]\right) = \exp\left(\ln \left[\prod_{x \in \tilde{x}(\Omega)} x^{f_{\tilde{x}}(x)}\right]\right) = \prod_{x \in \tilde{x}(\Omega)} x^{f_{\tilde{x}}(x)} \end{aligned}$$

and

$$\text{HE}(\tilde{x}) = \left[\sum_{x \in \tilde{x}(\Omega)} \frac{f_{\tilde{x}}(x)}{x}\right]^{-1}.$$

- *Note:* Probability and distribution are expectations:
- Consider the measure space (Ω, \mathcal{F}, P) , then

$$P(B) = \int_B dP = \int_{\Omega} \mathbb{I}_B dP = \mathbb{E}(\mathbb{I}_B), \text{ for all } B \in \mathcal{F}.$$

- Consider the random object $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$, then

$$\begin{aligned} P_{\tilde{x}}(A) &= P\{\tilde{x} \in A\} = \int_A dP_{\tilde{x}}(x) = \int_{\Omega} \mathbb{I}_A(\tilde{x}(\omega)) dP(\omega) \\ &= \int_{\Omega'} \mathbb{I}_A(x) dP_{\tilde{x}}(x) = \mathbb{E}[\mathbb{I}_A(\tilde{x})], \text{ for all } A \in \mathcal{F}', \end{aligned}$$

- or

$$P_{\tilde{x}}(A) = P(\tilde{x}^{-1}(A)) = \mathbb{E}[\mathbb{I}_{\tilde{x}^{-1}(A)}], \text{ for all } A \in \mathcal{F}'.$$

- **Proposition.** Let $\tilde{x}(\omega) = b$ a.s. $[P]$, where b is a constant. Then, $E(\tilde{x}) \equiv \int_{\Omega} \tilde{x} dP = b$.
- **Proposition.** If c is a constant, then $E(c\tilde{x}) = cE(\tilde{x})$.
- **Proposition.** Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$ be a random object, $\{c_i\}_{i=1}^n$ a collection of constants, and $g_i : (\Omega', \mathcal{F}') \longrightarrow (\mathbb{R}, \mathcal{B})$ measurable functions, for $i = 1, \dots, n$. Then,

$$E \left[\sum_{i=1}^n c_i g_i(\tilde{x}) \right] = \sum_{i=1}^n c_i E(g_i(\tilde{x})).$$

- **Corollary.** Let $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ be a collection of random variables on (Ω, \mathcal{F}, P) and $\{c_i\}_{i=1}^n$ a collection of constants. Then,

$$E \left[\sum_{i=1}^n c_i \tilde{x}_i \right] = \sum_{i=1}^n c_i E(\tilde{x}_i).$$

3.3. L^p spaces and the Cauchy-Schwarz inequality

- **Definition.** Assume a given measure space $(\Omega, \mathcal{F}, \mu)$ and let $f : (\Omega, \mathcal{F}, \mu) \rightarrow (\mathbb{R}, \mathcal{B})$ and p be a non-negative real number, $p \geq 0$. We say that $f \in L^p$ (or $f \in L^p(\Omega, \mathcal{F}, \mu)$) if

$$\int_{\Omega} |f|^p d\mu < \infty.$$

i.e., $|f|^p$ is μ -integrable.

- **Definition.** $\|f\|_p = \left[\int_{\Omega} |f|^p d\mu \right]^{1/p}$.
- **Hölder inequality:** Let $f \in L^p$ and $g \in L^q$ with $p > 1$, $q > 1$, and $\frac{1}{p} + \frac{1}{q} = 1$, then

$$\|fg\|_1 \leq \|f\|_p \cdot \|g\|_q,$$

that is,

$$\int_{\Omega} |fg| d\mu \leq \left[\int_{\Omega} |f|^p d\mu \right]^{1/p} \cdot \left[\int_{\Omega} |g|^q d\mu \right]^{1/q}.$$

- The **Cauchy-Schwarz inequality** is the **Hölder inequality** for $p = q = 2$, i.e., if $f \in L^2$ and $g \in L^2$ then

$$\|fg\|_1 \leq \|f\|_2 \cdot \|g\|_2,$$

- or

$$\int_{\Omega} |fg| d\mu \leq \left[\int_{\Omega} |f|^2 d\mu \right]^{1/2} \cdot \left[\int_{\Omega} |g|^2 d\mu \right]^{1/2},$$

- or

$$\int_{\Omega} |fg| d\mu \leq \left[\int_{\Omega} f^2 d\mu \right]^{1/2} \cdot \left[\int_{\Omega} g^2 d\mu \right]^{1/2}.$$

- Remember that

$$\int_{\Omega} fg d\mu \leq \left| \int_{\Omega} fg d\mu \right| \leq \int_{\Omega} |fg| d\mu.$$

Proof of the Cauchy-Schwarz inequality: For every constant K and any pair of functions f and g belonging to L^2 , we have

$$\int_{\Omega} (K|f| + |g|)^2 d\mu \geq 0$$

and, thus,

$$\int_{\Omega} K^2 |f|^2 d\mu + \int_{\Omega} 2K |f| |g| d\mu + \int_{\Omega} |g|^2 d\mu \geq 0.$$

The previous inequality becomes

$$\underbrace{K^2 \int_{\Omega} f^2 d\mu}_a + K \cdot 2 \underbrace{\int_{\Omega} |fg| d\mu}_b + \underbrace{\int_{\Omega} g^2 d\mu}_c \geq 0.$$

This means that the above second order polynomial of K can have one or zero roots. That is, the discriminant $D = b^2 - 4ac$ has to be less than or equal to zero. Therefore,

$$\left[2 \int_{\Omega} |fg| d\mu \right]^2 - 4 \left[\int_{\Omega} f^2 d\mu \right] \left[\int_{\Omega} g^2 d\mu \right] \leq 0.$$

Rearranging the above inequality and dividing by 4, one obtains

$$\left[\int_{\Omega} |fg| d\mu \right]^2 \leq \left[\int_{\Omega} f^2 d\mu \right] \cdot \left[\int_{\Omega} g^2 d\mu \right]$$

and taking the positive square root in both sides, we obtain the desired inequality. *Q.E.D.*

3.4. Moments and its properties: mean and variance

- Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$.

- Definition.** Let k be a natural number, $k = 0, 1, 2, \dots$. The k th (non-central) moment of \tilde{x} is

$$\mu'_k(\tilde{x}) = \mathbb{E}(\tilde{x}^k) = \int_{\Omega} \tilde{x}^k dP = \int_{\mathbb{R}} x^k dP_{\tilde{x}}(x).$$

- Definition.** Let k be a natural number, $k = 0, 1, 2, \dots$, and assume that $\mathbb{E}(\tilde{x})$ is finite. The k th central moment of \tilde{x} is

$$\begin{aligned} \mu_k(\tilde{x}) &= \mathbb{E}[(\tilde{x} - \mathbb{E}(\tilde{x}))^k] \\ &= \int_{\Omega} (\tilde{x} - \mathbb{E}(\tilde{x}))^k dP = \int_{\mathbb{R}} (x - \mathbb{E}(\tilde{x}))^k dP_{\tilde{x}}(x). \end{aligned}$$

- Odd moments may fail to exist and, whenever they exist, they may be equal to $\pm\infty$.
- Even moments always exist but they may be equal to $+\infty$.

- Note that the k th moment of \tilde{x} is finite if and only if \tilde{x}^k is P -integrable. Therefore, the finiteness of $E(\tilde{x}^k)$ requires the previous existence of the integral $\int_{\Omega} \tilde{x}^k dP$.

- Let k be a natural number, $k = 0, 1, 2, \dots$. Since

$$E(\tilde{x}^k) = E\left[(\tilde{x}^k)^+\right] - E\left[(\tilde{x}^k)^-\right]$$

and

$$E(|\tilde{x}|^k) = E(|\tilde{x}^k|) = E\left[(\tilde{x}^k)^+\right] + E\left[(\tilde{x}^k)^-\right],$$

then $E(\tilde{x}^k)$ is finite if and only if $E(|\tilde{x}|^k)$ is finite. Note that $E(|\tilde{x}|^k)$ always exists.

- In particular, $E(\tilde{x})$ is finite if and only if $E(|\tilde{x}|)$ is finite. Moreover, if $E(\tilde{x})$ does not exist (it is of the indefinite form $\infty - \infty$) or is equal to $\pm\infty$, then $E(|\tilde{x}|) = \infty$.

- **Proposition.** If k is a natural number, $k = 0, 1, 2, \dots$ and $E(\tilde{x}^k)$ is finite, then $E(\tilde{x}^j)$ is finite for all natural numbers j smaller or equal than k , $j = 0, 1, \dots, k$.
- Recall that $E(\tilde{x}^k)$ is finite if and only if $E(|\tilde{x}|^k)$ is finite, where $k = 0, 1, 2, \dots$. Therefore, the proposition says that $\tilde{x} \in L^k(\Omega, \mathcal{F}, P) \implies \tilde{x} \in L^j(\Omega, \mathcal{F}, P)$ for all $j = 0, 1, \dots, k$, where k is a natural number.
- The proposition follows immediately from the following lemma:
- **Lemma:** If $\tilde{x} \in L^k$, where k is a non-negative real number, then $\tilde{x} \in L^j$ for all real numbers $j \in [0, k]$.
- Hence, the lemma says that $L^k(\Omega, \mathcal{F}, P) \subset L^j(\Omega, \mathcal{F}, P)$ for $0 \leq j \leq k$. This inclusion relationship does not hold in general for measures that are not finite.

Proof of the Lemma: We will prove that $E(|\tilde{x}|^j) < \infty$ for all real numbers $j \in [0, k]$ when $E(|\tilde{x}|^k) < \infty$ with $k \geq 0$. The cases where either $j = 0$ or $k = 0$ are trivial so that we only consider the case where $k > 0$ and $j \in (0, k]$ as follows:

$$\begin{aligned}
 E(|\tilde{x}|^j) &= \int_{\Omega} |\tilde{x}|^j dP = \int_{\{|\tilde{x}|^j < 1\}} |\tilde{x}|^j dP + \int_{\{|\tilde{x}|^j \geq 1\}} |\tilde{x}|^j dP \\
 &\leq \int_{\{|\tilde{x}|^j < 1\}} dP + \int_{\{|\tilde{x}|^j \geq 1\}} (|\tilde{x}|^j)^{k/j} dP + \int_{\{|\tilde{x}|^j < 1\}} |\tilde{x}|^k dP = \\
 &\int_{\{|\tilde{x}|^j < 1\}} dP + \int_{\Omega} |\tilde{x}|^k dP = P\{|\tilde{x}|^j < 1\} + E(|\tilde{x}|^k) < \infty. \quad \text{Q.E.D.}
 \end{aligned}$$

- Note 1:** If $E(\tilde{x})$ fails to exist (it is of the form $\infty - \infty$) or is equal to $\pm\infty$, then $E(\tilde{x}^k) = \infty$ when k is an even strictly positive natural number, $k = 2, 4, 6, \dots$
- To see this observe that in this case $E(|\tilde{x}|) = \infty$ and $E(\tilde{x}^k) = E(|\tilde{x}|^k) \geq 0$ when k is even. According to the previous Lemma, if $E(|\tilde{x}|) = \infty$ then $E(|\tilde{x}|^k) = E(\tilde{x}^k) = \infty$.
- Note 2:** If $E(\tilde{x})$ fails to exist (i.e., it is of the form $\infty - \infty$), then $E(\tilde{x}^k)$ fails to exist when k is an odd natural number, $k = 1, 3, 5, \dots$
- To see this observe that in this case, since $\tilde{x}^+ \geq 0$ and $\tilde{x}^- \geq 0$, $E(\tilde{x}) = E(\tilde{x}^+) - E(\tilde{x}^-) = E(|\tilde{x}^+|) - E(|\tilde{x}^-|) = \infty - \infty$. Thus, $E(|\tilde{x}^+|) = \infty$ and $E(|\tilde{x}^-|) = \infty$. According to the previous Lemma, $E(|\tilde{x}^+|) = \infty$ implies that $E\left[(|\tilde{x}^+|)^k \right] = E\left[(\tilde{x}^k)^+ \right] = \infty$ and $E(|\tilde{x}^-|) = \infty$ implies that $E\left[(|\tilde{x}^-|)^k \right] = E\left[(\tilde{x}^k)^- \right] = \infty$. Therefore, $E(\tilde{x}^k) = E\left[(\tilde{x}^k)^+ \right] - E\left[(\tilde{x}^k)^- \right]$ is of the form $\infty - \infty$.

- **Note 3:** Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\overline{\mathbb{R}}, \mathcal{B})$ be an extended real-valued random variable. If $E(\tilde{x})$ is finite, then \tilde{x} is finite a.e. with respect to P .
- From now on, when we write a moment $\mu'_k(\tilde{x})$ of a (extended) random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\overline{\mathbb{R}}, \mathcal{B})$, it should be understood that this moment exists and is finite, that is, $\tilde{x} \in L^k(\Omega, \mathcal{F}, P)$.

- *Note:*

$$\mu'_0(\tilde{x}) = \mu_0(\tilde{x}) = E(1) = 1$$

and

$$\mu_1(\tilde{x}) = E[\tilde{x} - E(\tilde{x})] = E(\tilde{x}) - E(\tilde{x}) = 0.$$

- μ'_1 is the mean of \tilde{x} and is denoted by μ (or $\mu_{\tilde{x}}$),

$$\mu = \mu'_1 = E(\tilde{x}).$$

- μ_2 is the variance of \tilde{x} and is denoted by σ^2 , $\sigma_{\tilde{x}}^2$ or $\text{Var}(\tilde{x})$,

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

- The inverse of the variance of \tilde{x} is called the precision of \tilde{x} ,

$$\tau_{\tilde{x}} = \frac{1}{\sigma_{\tilde{x}}^2}.$$

- **Proposition.** Let $\tilde{x}(\omega) = b$ a.s. $[P]$, where b is a constant. Then $\text{Var}(\tilde{x}) = \text{Var}(b) = 0$.
- **Proposition.** If c is a constant, then $\text{Var}(c\tilde{x}) = c^2\text{Var}(\tilde{x})$.
- **Definition.** The standard deviation $\sigma_{\tilde{x}}$ (or σ) of \tilde{x} is

$$\sigma_{\tilde{x}} = (\text{Var}(\tilde{x}))^{1/2}.$$

- **Proposition.** If $c > 0$ is a constant, then $\sigma_{c\tilde{x}} = c\sigma_{\tilde{x}}$.
- **Definition.** The coefficient of variation of \tilde{x} is

$$\text{CV}_{\tilde{x}} = \frac{\sigma}{|\mu|}, \quad \text{with } \mu \neq 0.$$

- **Proposition.** If c is a constant, then $\text{CV}_{c\tilde{x}} = \text{CV}_{\tilde{x}}$.

● **Proposition.**

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

or

$$\sigma^2 \equiv \mu_2 = \mu'_2 - (\mu'_1)^2 \equiv \mu'_2 - \mu^2.$$

● **Proof.**

$$\begin{aligned}\sigma^2 &= E[(\tilde{x} - \mu)^2] = E[\tilde{x}^2 - 2\tilde{x}\mu + \mu^2] = E(\tilde{x}^2) - 2\mu E(\tilde{x}) + \mu^2 \\ &= E(\tilde{x}^2) - 2\mu^2 + \mu^2 = E(\tilde{x}^2) - \mu^2. \quad \text{Q.E.D.}\end{aligned}$$

- In general: If k is a natural number, we get the following formula for the k th central moment by just making the binomial expansion:

$$\mathbb{E} \left[(\tilde{x} - \mathbb{E}(\tilde{x}))^k \right] = \sum_{n=0}^k \binom{k}{n} \mathbb{E}(\tilde{x}^n) [-\mathbb{E}(\tilde{x})]^{k-n}$$

or

$$\mu_k = \sum_{n=0}^k \binom{k}{n} \mu'_n [-\mu]^{k-n}.$$

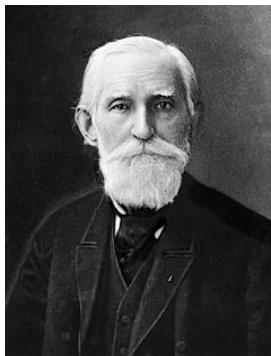
- Therefore, the central moment μ_k is finite if and only if the (non-central) moment μ'_k is finite.
- **Definition.** The coefficient of skewness or asymmetry (for $\sigma^3 \neq 0$) is

$$\text{CA} = \frac{\mu_3}{\sigma^3}.$$

- **Definition.** The coefficient of kurtosis (for $\sigma^4 > 0$) is

$$\text{CK} = \frac{\mu_4}{\sigma^4}.$$

3.5. Chebyshev's inequality



Pafnuty Chebyshev (1821 -1894)

- **Theorem (Chevyshev).** Assume that the random variable \tilde{x} has the finite mean μ and the finite variance $\sigma^2 > 0$. Then, for any $k \in (0, \infty)$,

$$P\{|\tilde{x} - \mu| < k\sigma\} \geq 1 - \frac{1}{k^2}, \quad (\text{Chevyshev's inequality})$$

or, equivalently,

$$P\{|\tilde{x} - \mu| \geq k\sigma\} \leq \frac{1}{k^2}. \quad (\text{Chevyshev's inequality})$$

- Making the change of variable $k\sigma = \varepsilon > 0$, Chebyshev's inequality becomes

$$P\{|\tilde{x} - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}, \quad (\text{Chebyshev's inequality})$$

or, equivalently,

$$P\{|\tilde{x} - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}. \quad (\text{Chebyshev's inequality})$$

- The Chebyshev's theorem says that a random variable with small variance is likely to take on values close to its mean.

Proof of the Theorem: We know from the definition of variance that

$$\sigma^2 = \mathbb{E} \left[(\tilde{x} - \mu)^2 \right] = \int_{\mathbb{R}} (x - \mu)^2 dP_{\tilde{x}}(x)$$

and, dividing the integral into three parts, we get

$$\begin{aligned} \sigma^2 &= \int_{(-\infty, \mu - k\sigma]} (x - \mu)^2 dP_{\tilde{x}} \\ &\quad + \int_{(\mu - k\sigma, \mu + k\sigma)} (x - \mu)^2 dP_{\tilde{x}} + \int_{[\mu + k\sigma, \infty)} (x - \mu)^2 dP_{\tilde{x}}. \end{aligned}$$

Since $(x - \mu)^2$ is non-negative, the three integrals are also non-negative. Therefore, we can form the inequality

$$\sigma^2 \geq \int_{(-\infty, \mu - k\sigma]} (x - \mu)^2 dP_{\tilde{x}}(x) + \int_{[\mu + k\sigma, \infty)} (x - \mu)^2 dP_{\tilde{x}}(x).$$

by deleting the second integral.

Now, since $(x - \mu)^2 \geq k^2\sigma^2$ if either $x \leq \mu - k\sigma$ or $x \geq \mu + k\sigma$, it follows that

$$\sigma^2 \geq \int_{(-\infty, \mu - k\sigma]} k^2\sigma^2 dP_{\tilde{x}}(x) + \int_{[\mu + k\sigma, \infty)} k^2\sigma^2 dP_{\tilde{x}}(x)$$

and, hence, dividing both sides by $k^2\sigma^2$, we get

$$\frac{1}{k^2} \geq \int_{(-\infty, \mu - k\sigma]} dP_{\tilde{x}}(x) + \int_{[\mu + k\sigma, \infty)} dP_{\tilde{x}}(x),$$

provided $\sigma^2 > 0$. Since the sum of the two integrals in this inequality represents the probability that \tilde{x} will take on a value less than or equal to $\mu - k\sigma$ or greater than or equal to $\mu + k\sigma$, we have thus shown that

$$P \{ |\tilde{x} - \mu| \geq k\sigma \} \leq \frac{1}{k^2}. \quad \text{Q.E.D.}$$

3.6. The moment-generating function

- **Definition.** The moment-generating function (MGF), $M_{\tilde{x}}(\cdot) : B \rightarrow \mathbb{R}$, of the random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ (or of the distribution $P_{\tilde{x}}$) is

$$M_{\tilde{x}}(t) = E(e^{t\tilde{x}}) = \int_{\Omega} e^{t\tilde{x}} dP = \int_{\mathbb{R}} e^{tx} dP_{\tilde{x}}(x) \geq 0.$$

where B is the subset of \mathbb{R} for which $E(e^{t\tilde{x}})$ is finite.

- *Notes:*
- (1) $M_{\tilde{x}}(0) = 1$.
- (2) $M_{\tilde{x}}(t)$ could fail to be finite for some set of values of $t \in \mathbb{R}$.
- (3) If there exists a real number a such that $P\{\tilde{x} < a\} = 0$ ($P\{\tilde{x} > a\} = 0$), then $M_{\tilde{x}}(t)$ is finite for all real numbers $t \leq (\geq) 0$.
- (4) If there exist two real numbers a and b such that $P\{a \leq \tilde{x} \leq b\} = 1$, then $M_{\tilde{x}}(t)$ is finite for all $t \in \mathbb{R}$.

- **Definition.** A neighborhood V of a point $x \in \mathbb{R}^n$ is a subset of \mathbb{R}^n that includes an open set containing the point x .
- Note that, if V is a neighborhood of the point x , then x lies in the interior of V .
- **Proposition.** If there exists a neighborhood V of 0 for which $M_{\tilde{x}}(t)$ is finite for all $t \in V$, then $M_{\tilde{x}}(t)$ is infinitely differentiable at $t = 0$, the moments μ'_r of \tilde{x} are finite for $r = 0, 1, 2, \dots$, and

$$\left. \frac{d^r M_{\tilde{x}}(t)}{dt^r} \right|_{t=0} = \mu'_r, \quad \text{for } r = 0, 1, 2, \dots$$

- **Proof.** Since, from the Taylor's expansion of e^y at $y = 0$, we have

$$e^y = \sum_{r=0}^{\infty} \frac{y^r}{r!},$$

it follows that

$$e^{tx} = \sum_{r=0}^{\infty} \frac{(tx)^r}{r!},$$

and, thus,

$$M_{\tilde{x}}(t) = \int_{\mathbb{R}} \overbrace{\left[\sum_{r=0}^{\infty} \frac{(tx)^r}{r!} \right]}^{e^{tx}} dP_{\tilde{x}}(x) = \sum_{r=0}^{\infty} \left[\int_{\mathbb{R}} \frac{(tx)^r}{r!} dP_{\tilde{x}}(x) \right] =$$
$$\sum_{r=0}^{\infty} \frac{t^r}{r!} \left(\underbrace{\int_{\mathbb{R}} x^r dP_{\tilde{x}}(x)}_{\mu'_r} \right) = 1 + \mu'_1 t + \mu'_2 \frac{t^2}{2!} + \mu'_3 \frac{t^3}{3!} + \dots + \mu'_r \frac{t^r}{r!} + \dots$$

Since $M_{\tilde{x}}(t) < \infty$ for all t in a neighborhood of $t = 0$, it holds that μ'_r is finite for $r = 0, 1, 2, \dots$ and, thus, $M_{\tilde{x}}(t)$ is infinitely differentiable at $t = 0$.

Finally,

$$\left. \frac{d^r M_{\tilde{x}}(t)}{dt^r} \right|_{t=0} = \left. \frac{d^r E(e^{t\tilde{x}})}{dt^r} \right|_{t=0} = E(\tilde{x}^r e^{t\tilde{x}}) \Big|_{t=0} = E(\tilde{x}^r) = \mu'_r. \quad Q.E.D.$$

- Therefore, if some moments of a random variable \tilde{x} do not exist (they are of the indefinite form $\infty - \infty$) or are not finite, then $M_{\tilde{x}}(t)$ fails to be finite in a neighborhood of $t = 0$ (and, thus, it fails to be infinitely differentiable at $t = 0$).
- However, even if a random variable has all its moments finite, the MGF may fail to be finite in a neighborhood of $t = 0$ (this is the case for the log-normal distribution).

- Since

$$M_{\tilde{x}}''(t) \equiv \frac{d^2 M_{\tilde{x}}(t)}{dt^2} = \underbrace{\mathbb{E}(\tilde{x}^2 e^{t\tilde{x}})}_{\geq 0} \geq 0,$$

the moment-generating function $M_{\tilde{x}}(t)$ is convex.

- Moreover, since

$$M_{\tilde{x}}'(t) \Big|_{t=0} \equiv \left. \frac{dM_{\tilde{x}}(t)}{dt} \right|_{t=0} = \mathbb{E}(\tilde{x}),$$

the slope of the moment-generating function $M_{\tilde{x}}(t)$ at $t = 0$ is the mean of \tilde{x} .

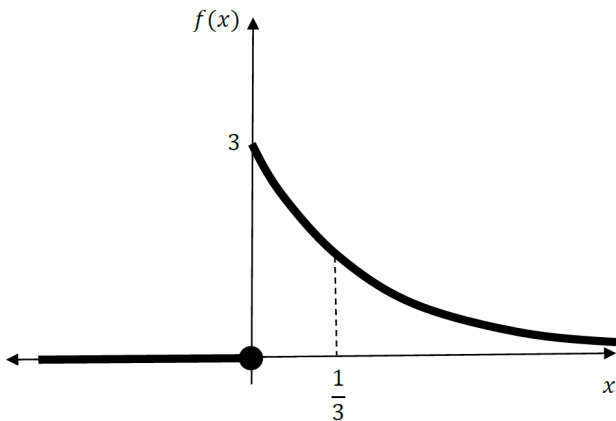
● Examples:

- The log-normal and F distributions have positive densities only for positive values. The Pareto distribution has positive density only for values larger than a strictly positive threshold. The Pareto, F , and log-normal distributions exhibit MGF's that are finite if and only if $t \leq 0$. The Student's t and Cauchy distributions have positive densities everywhere and exhibit MGF's that are not well defined (i.e., not finite) for all $t \neq 0$.
- The log-normal distribution has all its moments finite. The Pareto, F , and Student's t distributions could have some moments μ'_r , with $r \geq 1$, finite (depending on its parameters) but μ'_r for r sufficiently high becomes either indefinite (for the odd moments) or infinite (for the even moments). Finally, the Cauchy distribution has all its moments μ'_r , with $r \geq 1$, either indefinite (for the odd moments) or infinite (for the even moments).

● **Example:**

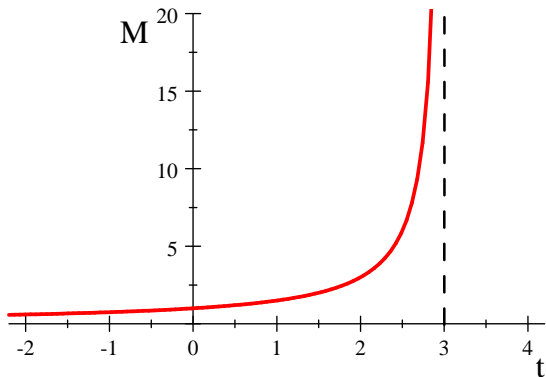
Consider the absolutely continuous random variable \tilde{x} with the density

$$f(x) = \begin{cases} 3e^{-3x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$



$$M_{\tilde{x}}(t) = \mathbb{E}(e^{t\tilde{x}}) = \int_{\mathbb{R}} e^{tx} dP_{\tilde{x}}(x) = \int_{\mathbb{R}} e^{tx} f(x) dx = \int_0^{\infty} e^{tx} 3e^{-3x} dx =$$

$$3 \int_0^{\infty} e^{(t-3)x} dx = 3 \left[\frac{e^{(t-3)x}}{t-3} \right]_0^{\infty} = 3 \cdot \left(0 - \frac{1}{t-3} \right) = \frac{3}{3-t}, \quad \text{for } t < 3.$$



$$M'_{\tilde{x}}(t) = \frac{3}{(3-t)^2} \implies M'_{\tilde{x}}(0) = \mu = \mu'_1 = E(\tilde{x}) = \frac{3}{9} = \frac{1}{3},$$

$$M''_{\tilde{x}}(t) = \frac{6}{(3-t)^3} \implies M''_{\tilde{x}}(0) = E(\tilde{x}^2) = \mu'_2 = \frac{6}{27} = \frac{2}{9},$$

$$\text{Var}(\tilde{x}) = \mu_2 = \mu'_2 - (\mu'_1)^2 = E(\tilde{x}^2) - [E(\tilde{x})]^2 = \frac{2}{9} - \left(\frac{1}{3}\right)^2 = \frac{1}{9}.$$

- **Note:** If $M_{\bar{x}}(t)$ is finite for all t in a neighborhood of 0, then $d^r M_{\bar{x}}(t) / dt^r$ is continuous at $t = 0$, for $r = 0, 1, 2, \dots$. In some cases, it could be necessary to compute the following limit to find the moment μ'_r :

$$\lim_{t \rightarrow 0} \frac{d^r M_{\bar{x}}(t)}{dt^r} = \left. \frac{d^r M_{\bar{x}}(t)}{dt^r} \right|_{t=0} = \mu'_r .$$

- **Properties of the moment-generating function:**

- 1 Let a be a constant, then $M_{\bar{x}+a}(t) = E(e^{t(\bar{x}+a)}) = e^{at} M_{\bar{x}}(t)$.
- 2 Let b be a constant, then $M_{b\bar{x}}(t) = E(e^{tb\bar{x}}) = M_{\bar{x}}(bt)$.
- 3 $M_{\frac{\bar{x}+a}{b}}(t) = e^{\frac{a}{b}t} M_{\bar{x}}\left(\frac{t}{b}\right)$.

- Recall that

$$e = \lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x = \sum_{r=0}^{\infty} \frac{1}{r!},$$

which follows from Newton's binomial theorem (please, check it).

3.7. The characteristic function and the Laplace transform

- Consider the set \mathbb{C} of complex numbers, where $\mathbb{C} \equiv \mathbb{R}^2$ is endowed with the following two *internal* operations:

(1) The (standard) sum,

$$(a, b) + (c, d) = (a + c, b + d),$$

(2) The (somewhat strange) multiplication,

$$(a, b) \cdot (c, d) = (ac - bd, ad + bc).$$

- The complex number (a, b) has a real component a and an imaginary component b . Therefore, the complex numbers of the form $(a, 0)$ are real, $(a, 0) = a \in \mathbb{R}$.

- Therefore, if $k \in \mathbb{R}$ then

$$k \cdot (a, b) = (k, 0) \cdot (a, b) = (ka, kb),$$

which agrees with the standard rule of multiplication of a vector by a scalar.

- The complex number $i \equiv (0, 1)$ is such that $i^2 = i \cdot i = (0, 1) \cdot (0, 1) = (-1, 0) = -1$. That is, $\sqrt{-1} = \pm i = (0, \pm 1)$.

- Observe that

$$(a, b) = a \cdot (1, 0) + b \cdot (0, 1) = a \cdot 1 + b \cdot i = a + bi.$$

- **Definition.** The modulus of the complex number (a, b) is its distance to $(0, 0)$, i.e., $(a^2 + b^2)^{1/2}$.

- **Definition.** The characteristic function, $\varphi_{\tilde{x}}(\cdot) : \mathbb{R} \longrightarrow \mathbb{C}$, of the random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ (or of its distribution) is

$$\begin{aligned}\varphi_{\tilde{x}}(t) &= \mathbb{E}(e^{it\tilde{x}}) = \int_{\mathbb{R}} e^{itx} dP_{\tilde{x}}(x) = \mathbb{E}[\cos(t\tilde{x}) + i\sin(t\tilde{x})] \\ &= \mathbb{E}[\cos(t\tilde{x})] + i\mathbb{E}[\sin(t\tilde{x})] = (\mathbb{E}[\cos(t\tilde{x})], \mathbb{E}[\sin(t\tilde{x})]),\end{aligned}$$

which is a well-defined (i.e., finite) complex number for all $t \in \mathbb{R}$.

- Recall that the moment-generating function, $M_{\tilde{x}}(\cdot) : B \longrightarrow \mathbb{R}$, of the random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ (or of its distribution) is

$$M_{\tilde{x}}(t) = \mathbb{E}(e^{t\tilde{x}}) = \varphi_{\tilde{x}}(-it).$$

where B is the subset of \mathbb{R} for which $\mathbb{E}(e^{t\tilde{x}})$ is finite.

- **Definition.** The Laplace transform, $\Lambda_{\tilde{x}}(\cdot) : D \rightarrow \mathbb{R}$, of the random variable $\tilde{x} : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}, \mathcal{B})$ (or of its distribution) is

$$\Lambda_{\tilde{x}}(t) = E(e^{-t\tilde{x}}) = M_{\tilde{x}}(-t) = \varphi_{\tilde{x}}(it).$$

where D is the subset of \mathbb{R} for which $E(e^{-t\tilde{x}})$ is finite.

- *Note:* Moment-generating functions and Laplace transforms may be not well-defined around $t = 0$ (i.e., they may be not finite).
- **Proposition.** Let $P_{\tilde{x}}$ and $P_{\tilde{y}}$ be the distributions of the random variables \tilde{x} and \tilde{y} , respectively. Then, $P_{\tilde{x}} = P_{\tilde{y}}$ if and only if $\varphi_{\tilde{x}} = \varphi_{\tilde{y}}$.
 - or iff $M_{\tilde{x}} = M_{\tilde{y}}$, provided they are well-defined (i.e., finite) in a neighborhood of $t = 0$,
 - or iff $\Lambda_{\tilde{x}} = \Lambda_{\tilde{y}}$, provided they are well-defined (i.e., finite) in a neighborhood of $t = 0$.

- Lévy's inversion formula:

$$P_{\tilde{x}}(a, b) + \frac{1}{2}P_{\tilde{x}}\{a, b\} = \underbrace{F_{\tilde{x}}(b) - F_{\tilde{x}}(a)}_{P_{\tilde{x}}(a, b]} - \frac{1}{2}[P_{\tilde{x}}\{b\} - P_{\tilde{x}}\{a\}]$$

$$= \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{itb}}{it} \varphi_{\tilde{x}}(t) dt.$$

- If the characteristic function $\varphi_{\tilde{x}}$ is Lebesgue integrable on \mathbb{R} , then \tilde{x} is absolutely continuous with the density

$$f_{\tilde{x}}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-itx}}{it} \varphi_{\tilde{x}}(t) dt.$$

- The Fourier transform of the distribution $P_{\tilde{x}}$ is $\mathfrak{F}_{\tilde{x}}(t) = E(e^{-it\tilde{x}}) = \varphi_{\tilde{x}}(-t)$ or $\mathfrak{F}_{\tilde{x}}(t) = E(e^{-2\pi it\tilde{x}}) = \varphi_{\tilde{x}}(-2\pi t)$, whereas the Fourier transform of the function $f : \mathbb{R} \rightarrow \mathbb{R}$ is

$$\widehat{f}(t) = \int_{\mathbb{R}} e^{-2\pi itx} f(x) dx.$$

3.8. Product moments and its properties: covariance and correlation coefficient.

- $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B})$.
- **Definition.** The (r_1, r_2, \dots, r_n) product moment of the random variables $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ is

$$\begin{aligned} \mu'_{r_1, r_2, \dots, r_n}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) &= \mathbb{E}(\tilde{x}_1^{r_1} \cdot \tilde{x}_2^{r_2} \cdot \dots \cdot \tilde{x}_n^{r_n}) \\ &= \int_{\Omega} (\tilde{x}_1^{r_1} \cdot \tilde{x}_2^{r_2} \cdot \dots \cdot \tilde{x}_n^{r_n}) dP = \int_{\mathbb{R}^n} (x_1^{r_1} \cdot x_2^{r_2} \cdot \dots \cdot x_n^{r_n}) dP_{\tilde{x}}(x_1, \dots, x_n). \end{aligned}$$

- **Definition.** The (r_1, r_2, \dots, r_n) product central moment of the random variables $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ is

$$\begin{aligned} \mu_{r_1, r_2, \dots, r_n}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) &= \\ E[(\tilde{x}_1 - E(\tilde{x}_1))^{r_1} \cdot (\tilde{x}_2 - E(\tilde{x}_2))^{r_2} \cdot \dots \cdot (\tilde{x}_n - E(\tilde{x}_n))^{r_n}] &= \\ \int_{\Omega} (\tilde{x}_1 - E(\tilde{x}_1))^{r_1} \cdot (\tilde{x}_2 - E(\tilde{x}_2))^{r_2} \cdot \dots \cdot (\tilde{x}_n - E(\tilde{x}_n))^{r_n} dP &= \\ \int_{\mathbb{R}^n} (x_1 - E(\tilde{x}_1))^{r_1} \cdot (x_2 - E(\tilde{x}_2))^{r_2} \cdot \dots \cdot (x_n - E(\tilde{x}_n))^{r_n} dP_{\tilde{x}}(x_1, \dots, x_n). \end{aligned}$$

● *Note:*

$$\mu'_r(\tilde{x}_i) = \mu'_{0,0,\dots,\underset{\uparrow}{i},\dots,0}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) = \mathbb{E}(\tilde{x}_i^r),$$

and

$$\mu_r(\tilde{x}_i) = \mu_{0,0,\dots,\underset{\uparrow}{i},\dots,0}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) = \mathbb{E}[(\tilde{x}_i - \mathbb{E}(\tilde{x}_i))^r].$$

- **Definition.** The covariance $\sigma_{\tilde{x},\tilde{y}}$ (or $\text{Cov}(\tilde{x},\tilde{y})$) between (or of) \tilde{x} and \tilde{y} is

$$\sigma_{\tilde{x},\tilde{y}} = \mu_{1,1}(\tilde{x},\tilde{y}) = \mathbb{E}[(\tilde{x} - \mathbb{E}(\tilde{x})) \cdot (\tilde{y} - \mathbb{E}(\tilde{y}))] =$$

$$\int_{\Omega} (\tilde{x} - \mathbb{E}(\tilde{x})) (\tilde{y} - \mathbb{E}(\tilde{y})) dP = \int_{\mathbb{R}^2} (x - \mathbb{E}(\tilde{x})) (y - \mathbb{E}(\tilde{y})) dP_{\tilde{x},\tilde{y}}(x, y).$$

- **Note:** $\sigma_{\tilde{x},\tilde{x}} = \sigma_{\tilde{x}}^2$ and $\sigma_{\tilde{x},\tilde{y}} = \sigma_{\tilde{y},\tilde{x}}$.

- **Proposition.**

$$\sigma_{\tilde{x}, \tilde{y}} = \mu'_{1,1}(\tilde{x}, \tilde{y}) - \mu_{\tilde{x}}\mu_{\tilde{y}}$$

or

$$\text{Cov}(\tilde{x}, \tilde{y}) = \text{E}(\tilde{x} \cdot \tilde{y}) - \text{E}(\tilde{x}) \cdot \text{E}(\tilde{y}).$$

- **Proof.**

$$\begin{aligned}\sigma_{\tilde{x}, \tilde{y}} &= \text{E} \left[(\tilde{x} - \mu_{\tilde{x}}) \cdot (\tilde{y} - \mu_{\tilde{y}}) \right] = \text{E} \left(\tilde{x}\tilde{y} - \tilde{x}\mu_{\tilde{y}} - \tilde{y}\mu_{\tilde{x}} + \mu_{\tilde{x}}\mu_{\tilde{y}} \right) \\ &= \text{E}(\tilde{x}\tilde{y}) - \mu_{\tilde{x}}\mu_{\tilde{y}} - \mu_{\tilde{y}}\mu_{\tilde{x}} + \mu_{\tilde{x}}\mu_{\tilde{y}} = \text{E}(\tilde{x} \cdot \tilde{y}) - \mu_{\tilde{x}}\mu_{\tilde{y}}. \quad \text{Q.E.D.}\end{aligned}$$

- **Definition.** The coefficient of correlation between \tilde{x} and \tilde{y} is

$$\rho_{\tilde{x},\tilde{y}} = \frac{\sigma_{\tilde{x},\tilde{y}}}{\sigma_{\tilde{x}} \cdot \sigma_{\tilde{y}}}, \quad \text{with } \sigma_{\tilde{x}} > 0 \text{ and } \sigma_{\tilde{y}} > 0.$$

- **Proposition.** $\rho_{\tilde{x},\tilde{y}} \in [-1, 1]$ or $|\rho_{\tilde{x},\tilde{y}}| \leq 1$.

- **Proof.** From the Cauchy-Schwarz inequality,

$$\begin{aligned} \left| \int_{\Omega} (\tilde{x} - E(\tilde{x})) (\tilde{y} - E(\tilde{y})) dP \right| &\leq \int_{\Omega} |(\tilde{x} - E(\tilde{x})) (\tilde{y} - E(\tilde{y}))| dP \\ &\leq \left[\int_{\Omega} (\tilde{x} - E(\tilde{x}))^2 dP \right]^{1/2} \left[\int_{\Omega} (\tilde{y} - E(\tilde{y}))^2 dP \right]^{1/2}. \end{aligned}$$

That is,

$$|\sigma_{\tilde{x},\tilde{y}}| \leq \sigma_{\tilde{x}} \cdot \sigma_{\tilde{y}} \Leftrightarrow -\sigma_{\tilde{x}} \cdot \sigma_{\tilde{y}} \leq \sigma_{\tilde{x},\tilde{y}} \leq \sigma_{\tilde{x}} \cdot \sigma_{\tilde{y}} \Leftrightarrow -1 \leq \rho_{\tilde{x},\tilde{y}} \leq 1.$$

Q.E.D.

- **Proposition.** Let a and b be constants. Then

$$\sigma_{a\tilde{x}, b\tilde{y}} = ab\sigma_{\tilde{x}, \tilde{y}}.$$

Moreover, if $a > 0$ and $b > 0$, then

$$\rho_{a\tilde{x}, b\tilde{y}} = \rho_{\tilde{x}, \tilde{y}}.$$

- **Proposition.** Let $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ be independent random variables on (Ω, \mathcal{F}, P) belonging to L^1 , then

$$E(\tilde{x}_1 \cdot \tilde{x}_2 \cdot \dots \cdot \tilde{x}_n) = E(\tilde{x}_1) \cdot E(\tilde{x}_2) \cdot \dots \cdot E(\tilde{x}_n).$$

- **Proof.** Since, from independency,

$P_{\tilde{x}} = P_{\tilde{x}_1} \times P_{\tilde{x}_2} \times \dots \times P_{\tilde{x}_n} = \prod_{i=1}^n P_{\tilde{x}_i}$, we can use Fubini's theorem so that

$$\begin{aligned} E(\tilde{x}_1 \cdot \tilde{x}_2 \cdot \dots \cdot \tilde{x}_n) &= \int_{\mathbb{R}^n} (x_1 \cdot x_2 \cdot \dots \cdot x_n) dP_{\tilde{x}}(x_1, \dots, x_n) \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} (x_1 \cdot x_2 \cdot \dots \cdot x_n) dP_{\tilde{x}_1}(x_1) dP_{\tilde{x}_2}(x_2) \dots dP_{\tilde{x}_n}(x_n) \\ &= \left[\int_{\mathbb{R}} x_1 dP_{\tilde{x}_1}(x_1) \right] \cdot \left[\int_{\mathbb{R}} x_2 dP_{\tilde{x}_2}(x_2) \right] \cdot \dots \cdot \left[\int_{\mathbb{R}} x_n dP_{\tilde{x}_n}(x_n) \right] \\ &= E(\tilde{x}_1) \cdot E(\tilde{x}_2) \cdot \dots \cdot E(\tilde{x}_n). \quad \text{Q.E.D.} \end{aligned}$$

- **Corollary.** If the random variables \tilde{x} and \tilde{y} are independent, then

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

i.e., \tilde{x} and \tilde{y} are "uncorrelated".

- **Proof.**

$$\text{Cov}(\tilde{x}, \tilde{y}) = E(\tilde{x} \cdot \tilde{y}) - E(\tilde{x}) \cdot E(\tilde{y}) = E(\tilde{x}) \cdot E(\tilde{y}) - E(\tilde{x}) \cdot E(\tilde{y}) = 0,$$

where the second equality holds because $E(\tilde{x} \cdot \tilde{y}) = E(\tilde{x}) \cdot E(\tilde{y})$ when \tilde{x} and \tilde{y} are independent. *Q.E.D.*

- Independency implies $\text{Cov}(\tilde{x}, \tilde{y}) = 0$, but the converse is not true.

- **Example:** Consider the following table summarizing the probability function $f_{\tilde{x}, \tilde{y}}(x, y)$ of the random variables \tilde{x} and \tilde{y} :

$y \backslash x$	-1	0	1	$f_{\tilde{y}}(y)$
-1	1/6	1/3	1/6	2/3
1	1/6	0	1/6	1/3
$f_{\tilde{x}}(x)$	1/3	1/3	1/3	1

- We have that $E(\tilde{x}) = 0$, $E(\tilde{y}) = -1/3$, and $E(\tilde{x} \cdot \tilde{y}) = 0$.
- Thus,

$$\sigma_{\tilde{x}, \tilde{y}} = E(\tilde{x} \cdot \tilde{y}) - E(\tilde{x}) \cdot E(\tilde{y}) = 0.$$

- However, \tilde{x} and \tilde{y} are not independent since it is not true that $f_{(\tilde{x}, \tilde{y})}(x, y) = f_{\tilde{x}}(x) \cdot f_{\tilde{y}}(y)$ for all $(x, y) \in \tilde{x}(\Omega) \times \tilde{y}(\Omega)$. For instance,

$$f(-1, -1) = \frac{1}{6} \neq f_{\tilde{x}}(-1) \cdot f_{\tilde{y}}(-1) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}.$$

3.9. Mean and variance of linear combinations of random variables: the mean vector and the covariance matrix

Proposition. Let $\{\tilde{x}_i\}_{i=1}^n$ be a collection of random variables on (Ω, \mathcal{F}, P) , $\{c_i\}_{i=1}^n$ a collection of constants, and b a constant. Then,

$$\begin{aligned}\text{Var} \left(\left[\sum_{i=1}^n c_i \tilde{x}_i \right] + b \right) &= \sum_{i=1}^n c_i^2 \text{Var} (\tilde{x}_i) + 2 \sum_{j=i+1}^n \sum_{i=1}^{n-1} c_i c_j \text{Cov} (\tilde{x}_i, \tilde{x}_j) \\ &= \sum_{i=1}^n c_i^2 \text{Var} (\tilde{x}_i) + \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{i=1}^n c_i c_j \text{Cov} (\tilde{x}_i, \tilde{x}_j) = \sum_{j=1}^n \sum_{i=1}^n c_i c_j \text{Cov} (\tilde{x}_i, \tilde{x}_j).\end{aligned}$$

Proof:

$$\begin{aligned}\text{Var} \left(\left[\sum_{i=1}^n c_i \tilde{x}_i \right] + b \right) &= \text{E} \left(\left[\left(\sum_{i=1}^n c_i \tilde{x}_i \right) + b - \text{E} \left(\left[\sum_{i=1}^n c_i \tilde{x}_i \right] + b \right) \right]^2 \right) \\ &= \text{E} \left(\left[\left(\sum_{i=1}^n c_i \tilde{x}_i \right) + b - \left(\sum_{i=1}^n c_i \text{E}(\tilde{x}_i) \right) - b \right]^2 \right) \\ &= \text{E} \left(\left[\sum_{i=1}^n c_i (\tilde{x}_i - \text{E}(\tilde{x}_i)) \right]^2 \right) \\ &= \text{E} \left[\sum_{i=1}^n c_i^2 (\tilde{x}_i - \text{E}(\tilde{x}_i))^2 + 2 \sum_{j=i+1}^n \sum_{i=1}^{n-1} c_i c_j (\tilde{x}_i - \text{E}(\tilde{x}_i)) (\tilde{x}_j - \text{E}(\tilde{x}_j)) \right] \\ &= \sum_{i=1}^n c_i^2 \text{E} \left[(\tilde{x}_i - \text{E}(\tilde{x}_i))^2 \right] + 2 \sum_{j=i+1}^n \sum_{i=1}^{n-1} c_i c_j \text{E} [(\tilde{x}_i - \text{E}(\tilde{x}_i)) (\tilde{x}_j - \text{E}(\tilde{x}_j))] \\ &= \sum_{i=1}^n c_i^2 \text{Var}(\tilde{x}_i) + 2 \sum_{j=i+1}^n \sum_{i=1}^{n-1} c_i c_j \text{Cov}(\tilde{x}_i, \tilde{x}_j). \quad \text{Q.E.D.}\end{aligned}$$

- **Corollary.** If $\{\tilde{x}_i\}_{i=1}^n$ is a collection of independent random variables, then

$$\text{Var} \left(\left[\sum_{i=1}^n c_i \tilde{x}_i \right] + b \right) = \sum_{i=1}^n c_i^2 \text{Var} (\tilde{x}_i) .$$

- **Corollary.** If the random variables \tilde{x} and \tilde{y} are independent, then

$$\text{Var} (\tilde{x} + \tilde{y}) = \text{Var} (\tilde{x}) + \text{Var} (\tilde{y}) .$$

- But, in general,

$$\text{Var} (\tilde{x} + \tilde{y}) = \text{Var} (\tilde{x}) + \text{Var} (\tilde{y}) + 2\text{Cov} (\tilde{x}, \tilde{y}) .$$

- Let $c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$ be a vector of scalars and $\tilde{x} = \begin{pmatrix} \tilde{x}_1 \\ \tilde{x}_2 \\ \vdots \\ \tilde{x}_n \end{pmatrix}$.

- $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{pmatrix}$, with $\mu_i = E(\tilde{x}_i)$, is the vector of means.

-

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}_{n \times n},$$

with $\sigma_{ij} = \text{Cov}(\tilde{x}_i, \tilde{x}_j)$, is the covariance matrix (or variance covariance matrix).

- Note:* Σ is symmetric ($\sigma_{ij} = \sigma_{ji}$) and the elements along the diagonal are the variances ($\sigma_{ii} = \sigma_i^2$).

- Then,

$$\mathbb{E}(c^T \tilde{x}) = \mathbb{E} \left(\sum_{i=1}^n c_i \tilde{x}_i \right) = c^T \mu$$

and

$$\text{Var}(c^T \tilde{x}) = \text{Var} \left(\sum_{i=1}^n c_i \tilde{x}_i \right) = c^T \Sigma c.$$

- *Note:* Since $\text{Var}(c^T \tilde{x}) = c^T \Sigma c \geq 0$ for all $c \in \mathbb{R}^n$, then the matrix Σ is symmetric positive semi-definite, which implies that $\det(\Sigma) \geq 0$.
- **Proposition.**

$$\text{Cov} \left(\tilde{y}, \sum_{i=1}^n c_i \tilde{x}_i \right) \equiv \text{Cov}(\tilde{y}, c^T \tilde{x}) = \sum_{i=1}^n c_i \text{Cov}(\tilde{y}, \tilde{x}_i).$$

- **Proof.** Exercise.

3.10. Multivariate moment-generating functions

- **Definition.** Let $\tilde{x} = (\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n) : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B})$ be a random vector. The multivariate moment-generating function, $M_{\tilde{x}}(\cdot) : B \longrightarrow \mathbb{R}$, of the random variables $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ is

$$\begin{aligned} M_{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n}(t_1, t_2, \dots, t_n) &= E(e^{t_1 \tilde{x}_1 + t_2 \tilde{x}_2 + \dots + t_n \tilde{x}_n}) \\ &= \int_{\Omega} e^{(t_1 \tilde{x}_1 + t_2 \tilde{x}_2 + \dots + t_n \tilde{x}_n)} dP = \int_{\mathbb{R}^n} e^{(t_1 x_1 + t_2 x_2 + \dots + t_n x_n)} dP_{\tilde{x}}(x_1, \dots, x_n), \end{aligned}$$

where B is the subset of \mathbb{R}^n for which $E(e^{t_1 \tilde{x}_1 + t_2 \tilde{x}_2 + \dots + t_n \tilde{x}_n})$ is finite.

- **Proposition.** If there exists a neighborhood V of $(0, 0, \dots, 0)$ for which $M_{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n}(t_1, t_2, \dots, t_n)$ is finite for all vectors $(t_1, t_2, \dots, t_n) \in V$, then $M_{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n}(t_1, t_2, \dots, t_n)$ is infinitely differentiable at $(t_1, t_2, \dots, t_n) = (0, 0, \dots, 0)$, all the product moments $\mu'_{r_1, r_2, \dots, r_n}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$ are finite and

$$\begin{aligned} &\left. \frac{\partial^{r_1 + r_2 + \dots + r_n} M_{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n}(t_1, t_2, \dots, t_n)}{\partial t_1^{r_1} \partial t_2^{r_2} \dots \partial t_n^{r_n}} \right|_{(t_1, t_2, \dots, t_n) = (0, 0, \dots, 0)} \\ &= \mu'_{r_1, r_2, \dots, r_n}(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n). \end{aligned}$$

- **Proposition.** If the random variables $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ are independent, then

$$M_{\tilde{x}_1 + \tilde{x}_2 + \dots + \tilde{x}_n}(t) = \prod_{i=1}^n M_{\tilde{x}_i}(t).$$

- **Proof.**

$$\mathbb{E} \left(e^{t(\tilde{x}_1 + \tilde{x}_2 + \dots + \tilde{x}_n)} \right) = \mathbb{E} \left(\prod_{i=1}^n e^{t\tilde{x}_i} \right) = \prod_{i=1}^n \mathbb{E} \left(e^{t\tilde{x}_i} \right). \quad Q.E.D.$$

- **Proposition.** If the random variables $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ are independent, then

$$M_{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n}(t_1, t_2, \dots, t_n) = \prod_{i=1}^n M_{\tilde{x}_i}(t_i).$$

- **Proof.**

$$\mathbb{E} \left(e^{t_1\tilde{x}_1 + t_2\tilde{x}_2 + \dots + t_n\tilde{x}_n} \right) = \mathbb{E} \left(\prod_{i=1}^n e^{t_i\tilde{x}_i} \right) = \prod_{i=1}^n \mathbb{E} \left(e^{t_i\tilde{x}_i} \right). \quad Q.E.D.$$

3.11. Conditional expectation

- **Definition.** Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$ be a random object and $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ a random variable. The conditional expectation of \tilde{y} given $\tilde{x} = x$ is a (essentially unique w.r.t. $P_{\tilde{x}}$) Borel measurable function $E(\tilde{y} | \tilde{x} = \cdot) : (\Omega', \mathcal{F}') \longrightarrow (\mathbb{R}, \mathcal{B})$ given by

$$E(\tilde{y} | \tilde{x} = x) = \int_{\Omega} \tilde{y} dP(\omega | \tilde{x} = x) = \int_{\mathbb{R}} y dP_{\tilde{y}|\tilde{x}}(y | x), \quad \text{for all } x \in \Omega'.$$

- Note that the previous conditional expectation can be seen as a random variable from (Ω', \mathcal{F}') to $(\mathbb{R}, \mathcal{B})$.
- All the properties of the (unconditional) expectation of \tilde{y} also hold a.s. $[P_{\tilde{x}}]$ for the conditional expectation of \tilde{y} given $\tilde{x} = x$.

- If the random vector (\tilde{x}, \tilde{y}) is discrete, then the conditional expectation of \tilde{y} given $\tilde{x} = x$ becomes

$$\mathbf{E}(\tilde{y} | \tilde{x} = x) = \sum_{y \in \tilde{y}(\Omega)} y f_{\tilde{y}|\tilde{x}}(y | x), \text{ for all } x \text{ with } f_{\tilde{x}}(x) > 0.$$

- If the random vector (\tilde{x}, \tilde{y}) is absolutely continuous, then the conditional expectation of \tilde{y} given $\tilde{x} = x$ becomes

$$\mathbf{E}(\tilde{y} | \tilde{x} = x) = \int_{\mathbb{R}} y f_{\tilde{y}|\tilde{x}}(y | x) dy, \text{ for all } x \text{ with } f_{\tilde{x}}(x) > 0.$$

- We can define conditional moments since moments are just expectations. For instance, the conditional variance of \tilde{y} given $\tilde{x} = x$ is

$$\begin{aligned} \text{Var}(\tilde{y} | \tilde{x} = x) &= \mathbf{E} \left([\tilde{y} - \mathbf{E}(\tilde{y} | \tilde{x} = x)]^2 | \tilde{x} = x \right) \\ &= \mathbf{E}(\tilde{y}^2 | \tilde{x} = x) - [\mathbf{E}(\tilde{y} | \tilde{x} = x)]^2, \text{ for all } x \in \Omega'. \end{aligned}$$

- If the discrete (absolutely continuous) random variables \tilde{x} and \tilde{y} are independent then

$$E(\tilde{y} | \tilde{x} = x) = E(\tilde{y}), \quad \text{for } f_{\tilde{x}}(x) > 0.$$

That is, the conditional expectation is equal to the corresponding unconditional expectation.

- In general, if the random object $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$ and the random variable $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ are independent then

$$E(\tilde{y} | \tilde{x} = x) = E(\tilde{y}), \quad \text{a.s. } [P_{\tilde{x}}].$$

- Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$ and $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega'', \mathcal{F}'')$.
- Conditional probability is a particular case of conditional expectation:

$$P(B | \tilde{x} = x) = E(\mathbb{I}_B | \tilde{x} = x), \text{ for all } B \in \mathcal{F}.$$

- The conditional distribution is also a particular case of conditional expectation:

$$P_{\tilde{y}}(A | \tilde{x} = x) = E(\mathbb{I}_A(\tilde{y}) | \tilde{x} = x), \text{ for all } A \in \mathcal{F}'',$$

or

$$P_{\tilde{y}}(A | \tilde{x} = x) = P(\tilde{y}^{-1}(A) | \tilde{x} = x) = E\left(\mathbb{I}_{\tilde{y}^{-1}(A)} | \tilde{x} = x\right),$$

for all $A \in \mathcal{F}''$.

- We can define the conditional expectation of the random variable $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ given the event $B \in \mathcal{F}$ as

$$E(\tilde{y} | B) = E(\tilde{y} | \mathbb{I}_B = 1).$$

- Theorem of total expectation (Adam's law).** Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$ be a random object and $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$. Then,

$$E(\tilde{y}) = \int_{\Omega'} E(\tilde{y} | \tilde{x} = x) dP_{\tilde{x}}(x).$$

- If the random vector (\tilde{x}, \tilde{y}) is discrete,

$$\begin{aligned} \int_{\mathbb{R}} E(\tilde{y} | \tilde{x} = x) dP_{\tilde{x}}(x) &= \sum_{x \in \tilde{x}(\Omega)} \left[\sum_{y \in \tilde{y}(\Omega)} y f_{\tilde{y} | \tilde{x}}(y | x) \right] f_{\tilde{x}}(x) \\ &= \sum_{x \in \tilde{x}(\Omega)} \sum_{y \in \tilde{y}(\Omega)} y f_{\tilde{x}, \tilde{y}}(x, y) = \sum_{y \in \tilde{y}(\Omega)} y \left[\sum_{x \in \tilde{x}(\Omega)} f_{\tilde{x}, \tilde{y}}(x, y) \right] \\ &= \sum_{y \in \tilde{y}(\Omega)} y f_{\tilde{y}}(y) = E(\tilde{y}). \end{aligned}$$

- If the random vector (\tilde{x}, \tilde{y}) is absolutely continuous,

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{E}(\tilde{y} | \tilde{x} = x) dP_{\tilde{x}}(x) &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} y f_{\tilde{y}|\tilde{x}}(y|x) dy \right] f_{\tilde{x}}(x) dx \\ &= \int_{\mathbb{R}} \left[\int_{\mathbb{R}} y f_{\tilde{x},\tilde{y}}(x,y) dy \right] dx = \int_{\mathbb{R}} \left[y \int_{\mathbb{R}} f_{\tilde{x},\tilde{y}}(x,y) dx \right] dy \\ &= \int_{\mathbb{R}} y f_{\tilde{y}}(y) dy = \mathbb{E}(\tilde{y}). \end{aligned}$$

3.12. Conditional expectation given a sigma-algebra

- **Definition.** Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega', \mathcal{F}')$ be a random object and $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ a random variable. The conditional expectation of \tilde{y} given \tilde{x} is a (essentially unique w.r.t. P) Borel measurable function $E(\tilde{y} | \tilde{x} = \tilde{x}(\cdot)) : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$ (or $E(\tilde{y} | \tilde{x})(\cdot)$) given by

$$\begin{aligned} E(\tilde{y} | \tilde{x})(\omega) &\equiv E(\tilde{y} | \tilde{x} = \tilde{x}(\omega)) \\ &= \int_{\Omega} \tilde{y}(\omega') dP(\omega' | \tilde{x} = \tilde{x}(\omega)) = \int_{\mathbb{R}} y dP_{\tilde{y} | \tilde{x}}(y | \tilde{x}(\omega)), \quad \text{for all } \omega \in \Omega. \end{aligned}$$

- Note that the previous conditional expectation can be seen as a random variable from (Ω, \mathcal{F}) to $(\mathbb{R}, \mathcal{B})$.
- Moreover, if $E(\tilde{y} | \tilde{x} = x) \equiv h(x)$ a.s. $[P_{\tilde{x}}]$, where $h : (\Omega', \mathcal{F}') \longrightarrow (\mathbb{R}, \mathcal{B})$, then $E(\tilde{y} | \tilde{x}) = h(\tilde{x})$, a.s. $[P]$ where $h(\tilde{x}) : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$, since $h(\tilde{x}(\omega)) = E(\tilde{y} | \tilde{x} = \tilde{x}(\omega))$.

- **Definition.** Let $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$ be a random object. The σ -algebra induced (or generated) on Ω by \tilde{x} is

$$\mathcal{F}(\tilde{x}) = \{ \tilde{x}^{-1}(A) \mid A \in \mathcal{F}' \},$$

that is, $\mathcal{F}(\tilde{x})$ contains the pre-images of all the measurable sets in \mathcal{F}' .

- The σ -algebra $\mathcal{F}(\tilde{x})$ is the smallest (i.e., the coarsest) σ -algebra that makes the random object \tilde{x} measurable, $\tilde{x} : (\Omega, \mathcal{F}(\tilde{x})) \longrightarrow (\Omega', \mathcal{F}')$. Hence, $\mathcal{F}(\tilde{x}) \subset \mathcal{F}$.
- Recall that, if the random variable $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B})$ is discrete, there is a countable partition $\mathcal{A} = \{A_1, A_2, \dots\}$ of Ω with $A_n = \{\omega \in \Omega \mid \tilde{x}(\omega) = x_n\}$, for all $x_n \in \tilde{x}(\Omega)$. Then, the σ -algebra $\mathcal{F}(\tilde{x})$ induced on Ω by \tilde{x} is the σ -algebra $\sigma(\mathcal{A})$ generated by the partition \mathcal{A} , i.e., the smallest σ -algebra containing \mathcal{A} .
- *Note:* If we have two random objects $\tilde{x}_1 : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega_1, \mathcal{F}_1)$ and $\tilde{x}_2 : (\Omega, \mathcal{F}, P) \longrightarrow (\Omega_2, \mathcal{F}_2)$ such that $\mathcal{F}(\tilde{x}_1) = \mathcal{F}(\tilde{x}_2)$, then

$$E(\tilde{y} \mid \tilde{x}_1) = E(\tilde{y} \mid \tilde{x}_2) \quad \text{a.s. } [P].$$

- Therefore, we can define the conditional expectation of the random variable $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ given the σ -algebra \mathcal{G} on Ω as the (essentially unique w.r.t. P) Borel measurable function $E(\tilde{y} | \mathcal{G})(\cdot) : (\Omega, \mathcal{G}) \longrightarrow (\mathbb{R}, \mathcal{B})$ given by $E(\tilde{y} | \mathcal{G}) = E(\tilde{y} | \tilde{x})$, where $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$ is any random object such that $\mathcal{G} = \mathcal{F}(\tilde{x})$.
- In particular, we can choose $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\Omega, \mathcal{G})$ to be the identity map, $\tilde{x}(\omega) = \omega$ for all $\omega \in \Omega$.
- Note that the conditional expectation of the random variable \tilde{y} given the σ -algebra \mathcal{G} can be seen as a random variable from (Ω, \mathcal{G}) to $(\mathbb{R}, \mathcal{B})$.
- *Note:* $E(\tilde{y} | \tilde{x}) = E(\tilde{y} | \mathcal{F}(\tilde{x}))$ a.s. $[P]$.

- **Definition.** The random object $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$ is \mathcal{G} -measurable if $\tilde{x}^{-1}(A) \in \mathcal{G}$ for all $A \in \mathcal{F}'$. Obviously, we must have that $\mathcal{G} \subset \mathcal{F}$.
- **Proposition.** If $\tilde{x} : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$ is a random object, then \tilde{x} is \mathcal{G} -measurable for all \mathcal{G} such that $\mathcal{F}(\tilde{x}) \subset \mathcal{G} \subset \mathcal{F}$.
- **Definition.** Let \mathcal{H} and \mathcal{G} be two σ -algebras on the set Ω with $\mathcal{H} \subset \mathcal{G}$. Then, we say that \mathcal{H} is coarser than \mathcal{G} or, equivalently, that \mathcal{G} is finer than \mathcal{H} .
- **Definition.** Let H and G be partitions of the sample space Ω . Then, we say that H is coarser than G or, equivalently, that G is finer than H , if every element of G is a subset of an element of H .

Example A:

- $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}$
- $H = \{\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}\} = \{\Omega\}$
- $G = \{\{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5\}\},$
- $F = \{\{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_5\}\}.$
- Observe that F is finer than G , while G is finer than H .

$$\{\omega_1\} \subset \{\omega_1, \omega_2, \omega_3\} \subset \Omega,$$

$$\{\omega_2\} \subset \{\omega_1, \omega_2, \omega_3\} \subset \Omega,$$

$$\{\omega_3\} \subset \{\omega_1, \omega_2, \omega_3\} \subset \Omega,$$

$$\{\omega_4\} \subset \{\omega_4, \omega_5\} \subset \Omega,$$

$$\{\omega_5\} \subset \{\omega_4, \omega_5\} \subset \Omega.$$

- Let \mathcal{G} be the σ -algebra generated by the partition G , i.e., \mathcal{G} is the smallest σ -algebra containing G .
- Similarly, in our example, \mathcal{H} is the σ -algebra generated by H , while \mathcal{F} is the σ -algebra generated by F .
- Obviously, if a countable partition H is coarser (resp. finer) than the countable partition G , then the σ -algebra \mathcal{H} generated by H is coarser (resp. finer) than the σ -algebra \mathcal{G} generated by G , i.e., $\mathcal{H} \subset \mathcal{G}$ (resp. $\mathcal{G} \subset \mathcal{H}$).
- *Note:* Not all σ -algebras are generated by partitions. For instance, there is no partition of \mathbb{R} that generates the σ -algebra \mathcal{B} of Borel sets in \mathbb{R} .

- **Example B:** Consider $\Omega = \mathbb{R}$ and the following uncountable partition of \mathbb{R} :

$$P = \{\{x\} \mid x \in \mathbb{R}\}.$$

The σ -algebra generated by P is

$$\mathcal{P} = \{A \subset \mathbb{R} \mid \text{either } A \text{ or } A^c \text{ are countable}\}.$$

Consider now the following countable partition of \mathbb{R} :

$$Q = \{(-\infty, 0), [0, \infty)\}.$$

The σ -algebra generated by Q is

$$\mathcal{Q} = \{\emptyset, \mathbb{R}, (-\infty, 0), [0, \infty)\}.$$

We see that P is finer than Q since, for all $x \in \mathbb{R}$, either $\{x\} \subset (-\infty, 0)$ or $\{x\} \subset [0, \infty)$. However, it is not true that $\mathcal{Q} \subset \mathcal{P}$ since neither $(-\infty, 0)$ nor $[0, \infty)$ are countable or have countable complements.

- In the previous Example A, we have $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$.
- $\mathcal{H} = \{\emptyset, \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}\}$,
- $\mathcal{G} = \{\emptyset, \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}, \{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5\}\}$,
and
- $\mathcal{F} = \{\emptyset, \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5\}, \{\omega_1, \omega_2, \omega_3\}, \{\omega_4, \omega_5\}, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_4\}, \{\omega_5\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_4\}, \{\omega_1, \omega_5\}, \{\omega_2, \omega_3\}, \{\omega_2, \omega_4\}, \{\omega_2, \omega_5\}, \{\omega_3, \omega_4\}, \{\omega_3, \omega_5\}, \{\omega_1, \omega_2, \omega_4\}, \{\omega_1, \omega_2, \omega_5\}, \{\omega_1, \omega_3, \omega_4\}, \{\omega_1, \omega_3, \omega_5\}, \{\omega_1, \omega_4, \omega_5\}, \{\omega_2, \omega_3, \omega_4\}, \{\omega_2, \omega_3, \omega_5\}, \{\omega_2, \omega_4, \omega_5\}, \{\omega_3, \omega_4, \omega_5\}, \{\omega_1, \omega_2, \omega_3, \omega_4\}, \{\omega_1, \omega_2, \omega_3, \omega_5\}, \{\omega_1, \omega_2, \omega_4, \omega_5\}, \{\omega_1, \omega_3, \omega_4, \omega_5\}, \{\omega_2, \omega_3, \omega_4, \omega_5\}\}$.
- In this example, \mathcal{F} is the power set 2^Ω (the finest σ -algebra on Ω) since F is the finest partition of Ω .
- Note that $\#\mathcal{F} = 2^5 = 32$.
- Moreover, in this example, \mathcal{H} is the coarsest σ -algebra on Ω , $\mathcal{H} = \{\emptyset, \Omega\}$, as H is the coarsest partition of Ω .

- **Example:** Consider the probability space $(\Omega, 2^\Omega, P)$, where the sample space is $\Omega = \{a, b, c\}$ and the probability P satisfies $P\{a\} = 1/6$, $P\{b\} = 2/3 = 4/6$, and $P\{c\} = 1/6$. Let $\tilde{x} : (\Omega, 2^\Omega, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ be a random variable defined as follows:

$$\tilde{x}(\omega) = \begin{cases} 1 & \text{if } \omega = a \\ 2 & \text{if } \omega \in \{b, c\}. \end{cases}$$

- The σ -algebra $\mathcal{F}(\tilde{x})$ induced on Ω by the random variable \tilde{x} is

$$\mathcal{F}(\tilde{x}) = \{\emptyset, \Omega, \{a\}, \{b, c\}\} \subset 2^\Omega.$$

- Then,

$$E(\tilde{x} | \mathcal{F}(\tilde{x})) = \tilde{x}(\omega) = \begin{cases} 1 & \text{if } \omega = a \\ 2 & \text{if } \omega \in \{b, c\}. \end{cases}$$

- Consider the following partition of the sample space Ω :
 $G = \{\{a, b\}, \{c\}\}$, and let \mathcal{G} be the σ -algebra generated by the partition G ,

$$\mathcal{G} = \{\emptyset, \Omega, \{a, b\}, \{c\}\} \subset 2^\Omega.$$

- Since

$$P\{\tilde{x} = 1 | \{a, b\}\} = P(\{a\} | \{a, b\}) = \frac{P(\{a\} \cap \{a, b\})}{P\{a, b\}} = \frac{P\{a\}}{P\{a, b\}} = \frac{1/6}{5/6} = \frac{1}{5},$$

$$\begin{aligned} P\{\tilde{x} = 2 | \{a, b\}\} &= P(\{b, c\} | \{a, b\}) = \frac{P(\{b, c\} \cap \{a, b\})}{P\{a, b\}} \\ &= \frac{P(b)}{P\{a, b\}} = \frac{4/6}{5/6} = \frac{4}{5}, \end{aligned}$$

$$P\{\tilde{x} = 1 | \{c\}\} = P(\{a\} | \{c\}) = \frac{P(\emptyset)}{P\{c\}} = \frac{0}{1/6} = 0,$$

$$P\{\tilde{x} = 2 | \{c\}\} = P(\{b, c\} | \{c\}) = \frac{P(\{b, c\} \cap \{c\})}{P\{c\}} = \frac{P\{c\}}{P\{c\}} = \frac{1/6}{1/6} = 1,$$

- then

$$E(\tilde{x} | \mathcal{G}) = \begin{cases} 1 \cdot \frac{1}{5} + 2 \cdot \frac{4}{5} = \frac{9}{5} & \text{if } \omega \in \{a, b\} \\ 1 \cdot 0 + 2 \cdot 1 = 2 & \text{if } \omega = c. \end{cases}$$

- Let \mathcal{H} be the coarsest σ -algebra on Ω , $\mathcal{H} = \{\emptyset, \Omega\}$. Since

$$P\{\tilde{x} = 1\} = P\{a\} = \frac{1}{6},$$

$$P\{\tilde{x} = 2\} = P\{b, c\} = \frac{5}{6},$$

then

$$E(\tilde{x} | \mathcal{H}) = E(\tilde{x}) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{5}{6} = \frac{11}{6} \quad \text{for } \omega \in \{a, b, c\} = \Omega.$$

- Finally, note that, since

$$P\{a, b\} = \frac{5}{6},$$

$$P\{c\} = \frac{1}{6},$$

then

$$E(E(\tilde{x} | \mathcal{G})) = E(E(\tilde{x} | \mathcal{G}) | \mathcal{H}) = \frac{9}{5} \cdot \frac{5}{6} + 2 \cdot \frac{1}{6} = \frac{11}{6} = E(\tilde{x} | \mathcal{H}) = E(\tilde{x}).$$

Properties of the conditional expectation given a σ -algebra:

1. (Theorem of total expectation). Let $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ and $\mathcal{G} \subset \mathcal{F}$, then $E(E(\tilde{y} | \mathcal{G})) = E(\tilde{y})$. In particular, if $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$, then

$$E(E(\tilde{y} | \tilde{x})) = E(\tilde{y}). \quad (\text{Adam's Law})$$

2. If $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ and $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ are independent random variables, then

$$E(\tilde{y} | \tilde{x}) = E(\tilde{y}) \quad \text{a.s. } [P].$$

3. Let $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ and $\mathcal{F}(\tilde{y}) \subset \mathcal{G} \subset \mathcal{F}$, then $E(\tilde{y} | \mathcal{G}) = \tilde{y}$ a.s. $[P]$. In particular, $E(\tilde{y} | \mathcal{F}) = \tilde{y}$ a.s. $[P]$, and

$$E(\tilde{y} | \tilde{y}) = E(\tilde{y} | \mathcal{F}(\tilde{y})) = \tilde{y} \quad \text{a.s. } [P].$$

4. Let $\tilde{z} : (\Omega, \mathcal{G}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ and $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ be random variables with $\mathcal{G} \subset \mathcal{F}$, then

$$E(\tilde{z} \cdot \tilde{y} | \mathcal{G}) = \tilde{z} \cdot E(\tilde{y} | \mathcal{G}) \quad \text{a.s. } [P].$$

In particular, $E(\tilde{z} \cdot \tilde{y} | \tilde{z}) = \tilde{z} \cdot E(\tilde{y} | \tilde{z})$ a.s. $[P]$.

5. If $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ and $\mathcal{H} = \{\emptyset, \Omega\}$ is the coarsest σ -algebra on Ω , then $E(\tilde{y} | \mathcal{H}) = E(\tilde{y})$ a.s. $[P]$.

3.13. The law of iterated expectations

Theorem. Let $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ and $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$, then

(a) $E(E(\tilde{y} | \mathcal{G}_1) | \mathcal{G}_2) = E(\tilde{y} | \mathcal{G}_1)$ a.s. $[P]$.

(b) The law of iterated expectations:

$$E(E(\tilde{y} | \mathcal{G}_2) | \mathcal{G}_1) = E(\tilde{y} | \mathcal{G}_1) \text{ a.s. } [P].$$

3.14. Jensen's inequality



Johan Jensen (1859 - 1925)

- **Theorem.** Let g be a Borel measurable function from the convex set $C \subset \mathbb{R}^n$ to \mathbb{R} . Then,

$$E(g(\tilde{y})) \underset{(\geq)}{\leq} g(E(\tilde{y})), \quad (\text{Jensen's inequality})$$

for all random vectors $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B})$ with $\tilde{y}(\Omega) \subset C$ (i.e., $\tilde{y}(\omega) \in C$ for all $\omega \in \Omega$) if and only if g is concave (convex) on C .

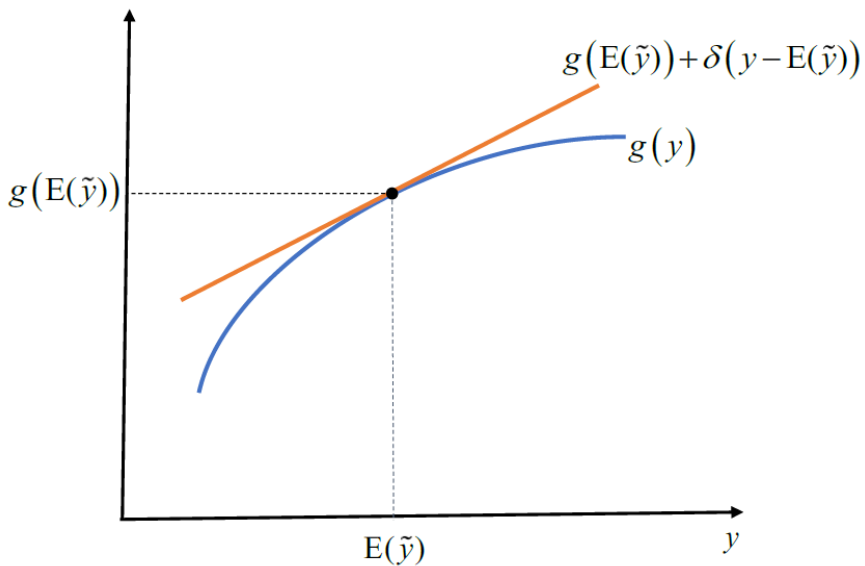
- **Proof: (IF)** If g is concave, there exists a real vector $\delta \in \mathbb{R}^n$ such that

$$g(y) \leq g(E(\tilde{y})) + \delta^\top (y - E(\tilde{y})), \text{ for all } y \in C,$$

where $E(\tilde{y}) \in C \subset \mathbb{R}^n$ is the mean vector of the random vector \tilde{y} .

Therefore,

$$\begin{aligned} \mathbf{E}(\mathbf{g}(\tilde{y})) &= \int_{\mathcal{C}} \mathbf{g}(y) dP_{\tilde{y}}(y) \leq \int_{\mathcal{C}} \left[\mathbf{g}(\mathbf{E}(\tilde{y})) + \delta^{\top} (y - \mathbf{E}(\tilde{y})) \right] dP_{\tilde{y}}(y) \\ &= \int_{\mathcal{C}} \mathbf{g}(\mathbf{E}(\tilde{y})) dP_{\tilde{y}}(y) + \int_{\mathcal{C}} \delta^{\top} (y - \mathbf{E}(\tilde{y})) dP_{\tilde{y}}(y) \\ &= \mathbf{g}(\mathbf{E}(\tilde{y})) + \delta^{\top} \underbrace{\mathbf{E}[\tilde{y} - \mathbf{E}(\tilde{y})]}_{=(0,0,\dots,0)^{\top} \in \mathbb{R}^n} = \mathbf{g}(\mathbf{E}(\tilde{y})). \end{aligned}$$



(ONLY IF) Consider a random vector $\tilde{y} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}^n, \mathcal{B})$ taking the values $x, y \in C$ with probability p and $(1 - p)$, respectively. Then, Jensen's inequality becomes

$$pg(x) + (1 - p)g(y) \leq g(px + (1 - p)y),$$

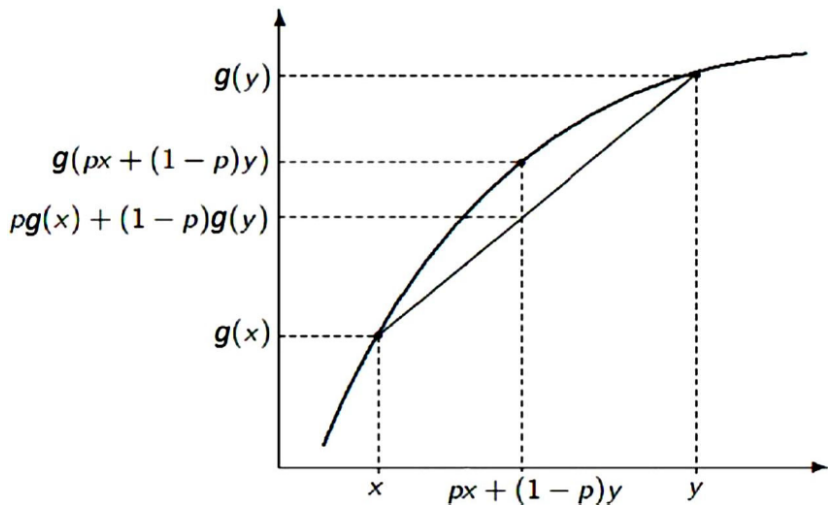
and, since this inequality holds for all values $x, y \in C$ and for all $p \in [0, 1]$, we have just obtained the definition of concavity of the function g on C . *Q.E.D.*

- Jensen's inequality also holds for the conditional expectation of \tilde{y} given $\tilde{x} = x$ (a.s. $[P_{\tilde{x}}]$) or given the σ -algebra \mathcal{G} (a.s. $[P]$):
- For g concave (convex):

$$E[g(\tilde{y}) | \tilde{x} = x] \underset{(\geq)}{\leq} g(E[\tilde{y} | \tilde{x} = x]), \quad \text{a.s. } [P_{\tilde{x}}]$$

or

$$E[g(\tilde{y}) | \mathcal{G}] \underset{(\geq)}{\leq} g(E[\tilde{y} | \mathcal{G}]), \quad \text{a.s. } [P].$$

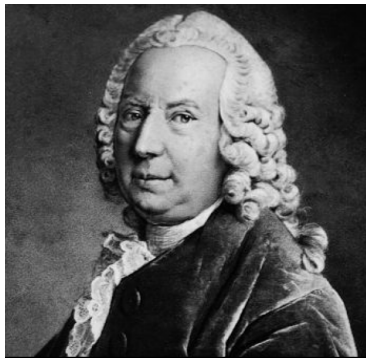


- **Application to Economics:**

- **Definition.** An expected utility maximizing individual with utility function u is risk averse if

$$u [E (\tilde{x})] \geq E [u (\tilde{x})] \quad \text{for all random variables } \tilde{x}.$$

- From Jensen's inequality we can conclude that an individual is risk averse if and only if the utility function u (which is called the Bernoulli utility function) is concave.



Daniel Bernoulli (1700 - 1782)

Proof that $E(\tilde{x}) = \int_{\mathbb{R}} x f_{\tilde{x}}(x) dx$ for absolutely continuous random variables.

Let $\tilde{x} : (\Omega, \mathcal{F}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ be an absolutely continuous random variable with density $f_{\tilde{x}}$. On the one hand, we know that

$$P_{\tilde{x}}(A) = \int_A dP_{\tilde{x}}(x) = \int_{\mathbb{R}} \mathbb{I}_A(x) dP_{\tilde{x}}(x). \quad (1)$$

On the other hand, from the definition of density, we have

$$P_{\tilde{x}}(A) = \int_A f_{\tilde{x}}(x) dx = \int_{\mathbb{R}} \mathbb{I}_A(x) f_{\tilde{x}}(x) dx, \quad (2)$$

as \tilde{x} is absolutely continuous. Therefore, combining (1) and (2) we get

$$\int_A dP_{\tilde{x}}(x) = \int_A f_{\tilde{x}}(x) dx. \quad (3)$$

Let us compute now the integral of the non-negative simple function $s : (\mathbb{R}, \mathcal{B}) \longrightarrow (\mathbb{R}, \mathcal{B})$ with respect to the probability distribution $P_{\tilde{x}}$. Recall that simple functions have the following form:

$$s(x) = \sum_{i=1}^n y_i \mathbb{I}_{A_i}(x), \quad (4)$$

where $\{A_1, A_2, \dots, A_n\}$ is a finite partition of \mathbb{R} and $s(x) = y_i$ if $x \in A_i$, $i = 1, \dots, n$. Thus,

$$\begin{aligned} \int_{\mathbb{R}} s(x) dP_{\tilde{x}}(x) &= \int_{\mathbb{R}} \sum_{i=1}^n y_i \mathbb{I}_{A_i}(x) dP_{\tilde{x}}(x) = \sum_{i=1}^n \int_{\mathbb{R}} y_i \mathbb{I}_{A_i}(x) dP_{\tilde{x}}(x) \\ &= \sum_{i=1}^n \int_{A_i} y_i dP_{\tilde{x}}(x) = \sum_{i=1}^n y_i \int_{A_i} dP_{\tilde{x}}(x) = \sum_{i=1}^n y_i \int_{A_i} f_{\tilde{x}}(x) dx \\ &= \sum_{i=1}^n \int_{A_i} s(x) f_{\tilde{x}}(x) dx = \int_{\bigcup_i A_i} s(x) f_{\tilde{x}}(x) dx = \int_{\mathbb{R}} s(x) f_{\tilde{x}}(x) dx, \end{aligned} \quad (5)$$

where the first equality comes from (4); the second from the fact that the integral of a sum is the sum of integrals; the third from the definition of indicator function; the fourth from the fact that y_i is a constant in the Borel set A_i and, thus, it can be factorized; the fifth from (3); the sixth from the fact that $s(x) = y_i$ when $x \in A_i$; the seventh from the countable additivity property of the integral; and the eighth from the fact that $\bigcup_i A_i = \mathbb{R}$.

Let us now compute the integral $\int_{\mathbb{R}} g(x) dP_{\tilde{x}}(x)$, where $g : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$. We can use the definition of integral with respect to a measure to compute the integral of the positive and negative parts of g , g^+ and g^- (which are given by $g^+(x) = \max(g, 0) \geq 0$ and $g^-(x) = \max(-g, 0) \geq 0$, respectively) as the supremum of the integrals of non-negative simple functions s and h such that $s \leq g^+$ and $h \leq g^-$, respectively. We thus have

$$\begin{aligned}
\mathbb{E}[g(\tilde{x})] &= \int_{\mathbb{R}} g(x) dP_{\tilde{x}}(x) = \int_{\mathbb{R}} g^+(x) dP_{\tilde{x}}(x) - \int_{\mathbb{R}} g^-(x) dP_{\tilde{x}}(x) \\
&= \sup_{0 \leq s \leq g^+} \int_{\mathbb{R}} s(x) dP_{\tilde{x}}(x) - \sup_{0 \leq h \leq g^-} \int_{\mathbb{R}} h(x) dP_{\tilde{x}}(x) \\
&= \sup_{0 \leq s \leq g^+} \int_{\mathbb{R}} s(x) f_{\tilde{x}}(x) dx - \sup_{0 \leq h \leq g^-} \int_{\mathbb{R}} h(x) f_{\tilde{x}}(x) dx \\
&= \int_{\mathbb{R}} g^+(x) f_{\tilde{x}}(x) dx - \int_{\mathbb{R}} g^-(x) f_{\tilde{x}}(x) dx = \int_{\mathbb{R}} g(x) f_{\tilde{x}}(x) dx, \tag{6}
\end{aligned}$$

where the first equality is the definition of expectation; the second and the third come from the definition of the integral of a Borel measurable function; the fourth from (5); and the fifth and the sixth from the fact that the density $f_{\tilde{x}}$ is non-negative a.e. with respect to Lebesgue measure and from the definition of integral.

If the previous Borel measurable function g is the identity function, i.e., $g(x) = x$ for all $x \in \mathbb{R}$, then (6) becomes

$$\mathbb{E}(\tilde{x}) = \int_{\mathbb{R}} x dP_{\tilde{x}}(x) = \int_{\mathbb{R}} x f_{\tilde{x}}(x) dx. \tag{Q.E.D.}$$

Exercises. Probability and Statistics. IDEA.

3. Expectation

1. The density function of the absolutely continuous random variable \tilde{x} is given by

$$f(x) = \begin{cases} \frac{1}{x \cdot (\ln 3)} & \text{for } 1 < x < 3 \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Find $E(\tilde{x})$, $E(\tilde{x}^2)$, and $E(\tilde{x}^3)$.

(b) Use the result of part (a) to determine the value of $E(\tilde{x}^3 + 2\tilde{x}^2 - 3\tilde{x} + 1)$.

2. A casino offers a game of chance for a single player in which a fair coin is tossed at each stage. The initial stake starts at 2 dollars and is doubled every time head appears. The first time tail appears, the game ends and the player wins whatever is in the pot. Thus, the player wins 2 dollars if tail appears on the first toss, 4 dollars if head appears on the first toss and tail on the second, 8 dollars if head appears on the first two tosses and tail on the third, and so on. Mathematically, the player wins 2^x dollars, where x is the number of tosses and the probability of having x tosses is given by the probability function $f(x) = (1/2)^x$ for $x = 1, 2, 3, \dots$. Show that $E(2^{\tilde{x}}) = \infty$ so that the fair price to pay the casino for entering the game is infinite. This is the famous **St. Petersburg paradox**, according to which a player's payoff expectation is infinite if he is to receive 2^x dollars when, in a series of flips of a balanced coin, the first tail appears on the x th flip.
3. Mr. Adams and Ms. Smith are betting on repeated flips of a coin. At the start of the game Mr. Adams has a dollars and Ms. Smith has b dollars. At each flip the loser pays the winner one dollar, and the game continues until either player is "ruined". Making use of the fact that in an equitable game each player's mathematical expectation is zero, find the probability that Mr. Adams will win Ms. Smith's b dollars before he loses his a dollars.
4. Show that if \tilde{x} is a random variable taking non-negative values with mean μ , then for any positive constant a ,

$$P\{\tilde{x} \geq a\} \leq \frac{\mu}{a}.$$

This inequality is called **Markov's inequality**.

5. Find the moment-generating function of the discrete random variable \tilde{x} , which has the probability function $f(x) = 2 \cdot (1/3)^x$ for $x = 1, 2, 3, \dots$, and use it to determine the moments μ'_1 , μ'_2 , and σ^2 .

6. Find the moment-generating function of the absolutely continuous random variable \tilde{x} whose density is given by

$$f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

and use it to find μ'_1 , μ'_2 and σ^2 .

7. If we let $R_{\tilde{x}}(t) = \ln M_{\tilde{x}}(t)$, show that $R'_{\tilde{x}}(0) = \mu$ and $R''_{\tilde{x}}(0) = \sigma^2$. Also, use these results to find the mean and the variance of a random variable \tilde{x} having the moment-generating function

$$M_{\tilde{x}}(t) = e^{4(e^t-1)}$$

8. Given the moment-generating function $M_{\tilde{x}}(t) = e^{3t+8t^2}$ of the random variable \tilde{x} , find the moment-generating function of the random variable $\tilde{z} = \frac{1}{4}(\tilde{x} - 3)$, and use it to find the mean and the variance of \tilde{z} .
9. If $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$ are random variables with the means 4, 9, 3, the variances 3, 5, 7, $\text{Cov}(\tilde{x}_1, \tilde{x}_2) = 1$, $\text{Cov}(\tilde{x}_2, \tilde{x}_3) = -2$, and $\text{Cov}(\tilde{x}_1, \tilde{x}_3) = -3$, find the mean and the variance of

(a) $\tilde{y} = 2\tilde{x}_1 - 3\tilde{x}_2 + 4\tilde{x}_3$;

(b) $\tilde{z} = \tilde{x}_1 + 2\tilde{x}_2 - \tilde{x}_3$.

10. If the joint density of \tilde{x} and \tilde{y} is given by

$$f(x, y) = \begin{cases} \frac{1}{3}(x + y) & \text{for } 0 < x < 1, 0 < y < 2 \\ 0 & \text{elsewhere,} \end{cases}$$

find the variance of $\tilde{w} = 3\tilde{x} + 4\tilde{y} - 5$.

11. A penny, which is unbalanced so that the probability of heads is 0.40, is tossed twice. What is the covariance between \tilde{z} , the number of heads obtained on the first toss, and \tilde{w} , the total number of heads obtained in the two tosses of the coin?
12. If getting a head is a success when we flip a coin, getting a six is a success when we roll a dice, and getting an ace is a success when we draw a card from an ordinary deck of 52 playing cards, find the mean and the standard deviation of the total number of successes when we
- (a) flip a balanced coin, roll a balanced dice, and then draw a card from a well-shuffled deck;
- (b) flip a balanced coin three times, roll a balanced dice twice, and then draw a card from a well-shuffled deck.

13. If the joint density function of two random variables \tilde{x} and \tilde{y} is given by

$$f(x, y) = \begin{cases} \frac{2}{3}(x + 2y) & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{elsewhere,} \end{cases}$$

- (a) find the conditional mean and the conditional variance of \tilde{x} given $\tilde{y} = 1/2$;
 (b) find the covariance between \tilde{x} and \tilde{y} .

14. The random variables \tilde{x} and \tilde{y} have the joint density

$$f(x, y) = \begin{cases} -\frac{1}{2} \ln(x \cdot y) & \text{for } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Taking into account that, if $F(x) = x \cdot (\ln x) - x$, then $F'(x) = \ln x$,

- (a) find the marginal density of \tilde{x} , $f_{\tilde{x}}(x)$.
 (b) find the conditional density of \tilde{x} given $\tilde{y} = y$, $f_{\tilde{x}|\tilde{y}}(x|y)$.
 (c) are \tilde{x} and \tilde{y} independent?
 (d) compute the expectation (or mean) of \tilde{x} , $E(\tilde{x})$.
 (e) compute the expectation of \tilde{x} given \tilde{y} , $E(\tilde{x}|\tilde{y})$.
 (f) find $P\{\tilde{x} > 2\tilde{y}\}$.
15. (a) Let \tilde{y} be an absolutely continuous random variable with density function (pdf) $g(y)$, $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$. If $\tilde{y} = y$, n independent observations $\tilde{x}_1, \dots, \tilde{x}_n$ are taken, where each \tilde{x}_i is absolutely continuous and has the same density $h(x_i; y)$, $h(\cdot; y) : \mathbb{R} \rightarrow \overline{\mathbb{R}}$. Indicate how to compute the conditional expectation of \tilde{y} given $\tilde{x}_1 = x_1, \dots, \tilde{x}_n = x_n$.
 (b) Let \tilde{y} be a discrete random variable with probability function (pmf) $g(y)$, $g : \tilde{y}(\Omega) \rightarrow [0, 1]$. If $\tilde{y} = y$, n independent observations $\tilde{x}_1, \dots, \tilde{x}_n$ are taken, where each \tilde{x}_i is discrete and has the same probability function $h(x; y)$, $h(\cdot; y) : \tilde{x}(\Omega) \rightarrow [0, 1]$. Indicate how to compute the conditional expectation of \tilde{y} given $\tilde{x}_1 = x_1, \dots, \tilde{x}_n = x_n$.
- Note:* Your answers to (a) and (b) should use only the functions $g(\cdot)$ and $h(\cdot; \cdot)$ and the values x_1, \dots, x_n .
16. Let $\{B_1, B_2, \dots\}$ be a countable collection of events with strictly positive probability that constitutes a partition of the sample space Ω . Establish the following version of the theorem of the total expectation:

$$\text{If } E(\tilde{x}) \text{ exists, then } E(\tilde{x}) = \sum_n P(B_n)E(\tilde{x}|B_n).$$

Recall that $E(\tilde{x}|B) \equiv E(\tilde{x}|\mathbb{I}_B = 1)$, where B is an event and \mathbb{I}_B is the indicator function of the set B .

17. Find the covariance of the two random variables whose joint density is given by

$$f(x, y) = \begin{cases} 2 & \text{for } x > 0, y > 0, x + y < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

18. We have a urn with 2 white balls and 3 black balls. We first extract one ball and we put it back in the urn. If the ball is white, we next extract two balls with no replacement. However, if the ball is black, we next extract only one ball. Find the moment-generating function of the number of black balls extracted in the second round.

19. Find the moment-generating function of the discrete random variable taking the values 0, 1, 2, ... with the probability function

$$f(k) = P\{\tilde{x} = k\} = q^k \cdot p, \quad k = 0, 1, 2, \dots, \quad q + p = 1, \quad q > 0, \quad p > 0$$

20. The moments of a random variable are $\mu'_1 = 1$, $\mu'_2 = 2$, and in general $\mu'_k = k$, for every strictly positive natural number k . Find its moment-generating function.

21. Is the function $g(t) = (2 + t)^2$ the moment-generating function of some random variable?

22. Consider a discrete random variable \tilde{x} with the following probability function:

$$f(x) = \begin{cases} p_0 & \text{for } x = 0 \\ p_1 & \text{for } x = 1 \\ 1 - p_0 - p_1 & \text{for } x = 2. \end{cases}$$

- (a) Find the moment-generating function of \tilde{x} .
 (b) Using the moment-generating function, compute $E(\tilde{x})$.
 (c) Using the moment-generating function, compute $E(\tilde{x}^2)$.
 (d) What is the variance of \tilde{x} ?

23. The probability function of the discrete random variable \tilde{x} is

$$f_{\tilde{x}}(x) = \frac{k}{x} \quad \text{for } x = 1, 2, 3.$$

- (a) The value of the constant k is $6/11$. Why?
 (b) Find the moment-generating function of the random variable \tilde{x} .
 (c) Compute $E(\tilde{x})$, $E(\tilde{x}^2)$ and $E(\tilde{x}^4)$ by using the moment-generating function.
 (d) Compute $\text{Var}(\tilde{x})$.
 (e) Compute $\text{Var}(\tilde{y})$, where $\tilde{y} = 3\tilde{x}^2 + 4$.

24. Consider a Borel measurable function f on the measure space $(\Omega, \mathcal{F}, \mu)$,

$$f : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B}).$$

(a) Assume that μ is a finite measure. Prove that, if $f \in L^s$, then $f \in L^r$ for $0 \leq r \leq s < \infty$.

(b) Assume that μ is the Lebesgue measure on the σ -algebra \mathcal{B} of the Borel sets of \mathbb{R} , i.e., $\Omega = \mathbb{R}$ and $\mathcal{F} = \mathcal{B}$, and

$$f(x) = \begin{cases} \frac{1}{x} & \text{for } x \geq 1 \\ 0 & \text{for } x < 1. \end{cases}$$

Does f belong to the L^1 space? Does f belong to the L^2 space? Compare the answers to the previous two questions with the result of part (a) and make a comment.

25. (a) Find the distribution function $F(x, y)$ of the random vector (\tilde{x}, \tilde{y}) whose density function is

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & \text{for } 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise,} \end{cases}$$

and check that $f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$ a.e. with respect to Lebesgue measure.

(b) Find $\text{Var}(\tilde{x} | \tilde{y} = 1/3)$.

26. Let (Ω, \mathcal{F}, P) be a probability space. Assume that the sample space Ω is the closed interval of real numbers $[0, 1]$, \mathcal{F} is the σ -algebra of Borel sets in the interval $[0, 1]$, and the probability P on (Ω, \mathcal{F}) is the Lebesgue measure on $[0, 1]$. Consider the random variables \tilde{x} , \tilde{y} , and \tilde{z} on (Ω, \mathcal{F}, P) defined as follows:

$$\tilde{x}(\omega) = \begin{cases} 1 & \text{if } \omega \in \left[0, \frac{1}{2}\right] \\ 0 & \text{otherwise,} \end{cases}$$

$$\tilde{y}(\omega) = \begin{cases} 1 & \text{if } \omega \in \left[\frac{1}{2}, 1\right] \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\tilde{z}(\omega) = \begin{cases} 1 & \text{if } \omega \in \left[0, \frac{1}{4}\right] \cup \left[\frac{1}{2}, \frac{3}{4}\right] \\ 0 & \text{otherwise.} \end{cases}$$

- (a) Find the joint probability function of the random variables \tilde{x} , \tilde{y} , and \tilde{z} .
 (b) Are \tilde{x} and \tilde{y} independent? Are \tilde{x} and \tilde{z} independent? Are \tilde{y} and \tilde{z} independent?
 (c) Find $\text{Cov}(\tilde{x}, \tilde{y})$, $\text{Cov}(\tilde{x}, \tilde{z})$, and $\text{Cov}(\tilde{y}, \tilde{z})$.

27. The distribution of the random vector (\tilde{x}, \tilde{y}) has the following density function:

$$f_{(\tilde{x}, \tilde{y})}(x, y) = \begin{cases} ky & \text{if } x < 5, y > 0, x - 2y > 3 \\ 0 & \text{otherwise.} \end{cases}$$

- (a) Prove that $k = 3$.
 (b) Compute the expectations $E(\tilde{x})$ and $E(\tilde{y})$.
 (c) Compute the variance $\text{Var}(\tilde{x})$.
 (d) Compute the covariance $\text{Cov}(\tilde{x}, \tilde{y})$ and the correlation coefficient ρ between \tilde{x} and \tilde{y} . Are \tilde{x} and \tilde{y} independent random variables?
 (e) Find the marginal density function of \tilde{y} .
 (f) Find the conditional density function of \tilde{x} given $\tilde{y} = 3/4$, $f_{\tilde{x}|\tilde{y}}(x|3/4)$.
 (g) Compute the conditional expectation of \tilde{x} given $\tilde{y} = 3/4$, $E(\tilde{x}|\tilde{y} = 3/4)$.
 (h) Compute the conditional variance of \tilde{x} given $\tilde{y} = 3/4$, $\text{Var}(\tilde{x}|\tilde{y} = 3/4)$.

28. Consider the random experiment consisting of rolling a balanced dice. Let $\tilde{x} : (\Omega, 2^\Omega) \rightarrow (\mathbb{R}, \mathcal{B})$ be the random variable representing the prize in euros you get as a function of the number of dots $\omega \in \Omega$ obtained when rolling the dice. The random variable \tilde{x} is defined as follows:

$$\tilde{x}(\omega) = \begin{cases} 0 & \text{if } \omega \in \{1, 4\} \\ 1 & \text{if } \omega = 3 \\ 2 & \text{if } \omega \in \{2, 5\} \\ 3 & \text{if } \omega = 6 \end{cases}$$

- (a) Find the σ -algebra $\mathcal{F}(\tilde{x})$ induced on Ω by the random variable \tilde{x} . Recall that $\mathcal{F}(\tilde{x})$ is the smallest σ -algebra \mathcal{F} on Ω for which the random variable $\tilde{x} : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$ is measurable.

Consider the following partition of the sample space $\Omega : G = \{\{1\}, \{2, 3, 4\}, \{5, 6\}\}$ and let \mathcal{G} be the σ -algebra generated by the partition G .

- (b) Compute the following three conditional expectations of \tilde{x} given a σ -algebra: $E(\tilde{x}|\mathcal{F}(\tilde{x}))$, $E(\tilde{x}|\mathcal{G})$, and $E(\tilde{x}|\mathcal{H})$, where $\mathcal{H} = \{\Omega, \emptyset\}$ is the coarsest σ -algebra on Ω . Recall that a conditional expectation given the σ -algebra \mathcal{F} is a function defined on the sample space Ω that is \mathcal{F} -measurable, $E(\tilde{x}|\mathcal{F}) : (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B})$.

29. Let $\tilde{y}, \tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_m, \tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ be random variables on the probability space (Ω, \mathcal{F}, P) and $b_j, j = 1, \dots, m, c_i, i = 1, \dots, n$, are scalars.

(a) Prove that

$$\text{Cov} \left(\tilde{y}, \sum_{i=1}^n c_i \tilde{x}_i \right) = \sum_{i=1}^n c_i \text{Cov} (\tilde{y}, \tilde{x}_i).$$

(b) Prove that

$$\text{Cov} \left(\sum_{j=1}^m b_j \tilde{y}_j, \sum_{i=1}^n c_i \tilde{x}_i \right) = \sum_{j=1}^m \sum_{i=1}^n b_j c_i \text{Cov} (\tilde{y}_j, \tilde{x}_i).$$

30. Let \tilde{x} and \tilde{y} be real-valued random variables on the probability space (Ω, \mathcal{F}, P) . Both random variables have a strictly positive, finite variance.

(a) One of the following equalities is always true. Which one? Justify your answer.

(i) $\text{Var} (\tilde{x}) = \text{Var} (\text{E} (\tilde{x} | \tilde{y}))$

(ii) $\text{Var} (\tilde{x}) = \text{E} (\text{Var} (\tilde{x} | \tilde{y}))$

(iii) $\text{Var} (\tilde{x}) = \text{E} (\text{Var} (\tilde{x} | \tilde{y})) + \text{Var} (\text{E} (\tilde{x} | \tilde{y}))$

(iv) $\text{Var} (\tilde{x}) = \text{E} (\text{Var} (\tilde{x})) + \text{Var} (\text{E} (\tilde{x} | \tilde{y}))$

(b) Assume now that \tilde{x} and \tilde{y} are independent. In this case, in addition to the one found in part (a), are there other equalities that are true? Which ones? Justify your answer.

(c) Let $h : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ be a Borel measurable function. Find the value of $\text{E} [(\tilde{x} - \text{E} (\tilde{x} | \tilde{y})) \cdot h (\tilde{y})]$ and the value of $\text{Cov} [(\tilde{x} - \text{E} (\tilde{x} | \tilde{y})), h (\tilde{y})]$.

Hint: In this exercise you should use the theorem of total expectation, according to which, if \tilde{x} and \tilde{y} are real-valued random variables on the probability space (Ω, \mathcal{F}, P) , then

$$\text{E} (\tilde{x}) = \text{E} (\text{E} (\tilde{x} | \tilde{y})).$$

31. Consider the random experiment consisting of rolling a balanced dice. Let $\tilde{x} : (\Omega, 2^\Omega) \rightarrow (\mathbb{R}, \mathcal{B})$ be the random variable representing the prize in euros you get as a function of the number of dots $\omega \in \Omega$ obtained when rolling the dice. The random variable \tilde{x} is defined as follows:

$$\tilde{x}(\omega) = \begin{cases} 1 & \text{if } \omega \in \{1, 2, 3\} \\ 2 & \text{if } \omega \in \{4, 5\} \\ 3 & \text{if } \omega \in \{6\} \end{cases}$$

(a) Find the σ -algebra $\mathcal{F} (\tilde{x})$ induced on Ω by the random variable \tilde{x} .

Consider the following partition of the sample space Ω :

$$G = \{\{1, 2\}, \{3, 4\}, \{5, 6\}\}$$

and let \mathcal{G} be the σ -algebra generated by the partition G .

(b) Compute the following three conditional expectations of \tilde{x} given a σ -algebra: $E(\tilde{x} | \mathcal{F}(\tilde{x}))$, $E(\tilde{x} | \mathcal{G})$, and $E(\tilde{x} | \mathcal{H})$, where $\mathcal{H} = \{\Omega, \emptyset\}$ is the coarsest σ -algebra on Ω .

32. Prove that the characteristic function $\varphi_{\tilde{x}}(t) = E(e^{it\tilde{x}})$ of the random variable \tilde{x} is equal to the following complex valued function:

$$E[\cos(t\tilde{x}) + i \sin(t\tilde{x})], \text{ which is equal to } (E[\cos(t\tilde{x})], E[\sin(t\tilde{x})]).$$

Hint: You should make the corresponding Taylor's expansions around zero.

33. In this exercise, we will use both Markov's and Jensen's inequalities to derive some lower bounds on the moment-generating function of a random variable.

(a) Use Markov's inequality (which is given in Exercise 4 of this list) to prove that the moment-generating function $M_{\tilde{y}}(t)$ of a random variable \tilde{y} satisfies

$$M_{\tilde{y}}(t) \geq e^{ct} \cdot P\{\tilde{y} \geq c\}, \text{ for all } c \in \mathbb{R} \text{ and all } t \geq 0 \text{ where } M_{\tilde{y}}(t) \text{ is well defined.}$$

Similarly, prove that

$$M_{\tilde{y}}(t) \geq e^{ct} \cdot P\{\tilde{y} \leq c\}, \text{ for all } c \in \mathbb{R} \text{ and all } t \leq 0 \text{ where } M_{\tilde{y}}(t) \text{ is well defined.}$$

(b) Use Jensen's inequality to prove that the moment-generating function $M_{\tilde{x}}(t)$ of any random variable \tilde{x} with finite mean μ is bounded below by $e^{\mu t}$ for all $t \in \mathbb{R}$ where $M_{\tilde{x}}(t)$ is well defined.

34. Consider the random experiment consisting of rolling a balanced dice. Let $\tilde{x} : (\Omega, 2^\Omega) \rightarrow (\mathbb{R}, \mathcal{B})$ be the random variable representing the prize in euros you get as a function of the number of dots $\omega \in \Omega$ obtained when rolling the dice. The random variable \tilde{x} is defined as follows:

$$\tilde{x}(\omega) = \begin{cases} 2 & \text{if } \omega \in \{1, 2\} \\ 3 & \text{if } \omega \in \{3, 4, 5, 6\} \end{cases},$$

and $\mathcal{F}(\tilde{x})$ is the σ -algebra induced on Ω by the random variable \tilde{x} .

Consider the following partition of the sample space Ω :

$$G = \{\{1, 2, 3\}, \{4, 5, 6\}\}$$

and \mathcal{G} is the σ -algebra generated by the partition G .

Compute the following three conditional expectations of \tilde{x} given a σ -algebra: $E(\tilde{x} | \mathcal{F}(\tilde{x}))$, $E(\tilde{x} | \mathcal{G})$, and $E(\tilde{x} | \mathcal{H})$, where $\mathcal{H} = \{\Omega, \emptyset\}$ is the coarsest σ -algebra on Ω .

35. Prove that, for every non-negative random variable \tilde{x} , the following inequalities hold:

$$\mathbf{E}(\tilde{x}) \geq \mathbf{GE}(\tilde{x}) \geq \mathbf{HE}(\tilde{x}).$$

36. (a) Prove that, for every random variable \tilde{x} ,

$$(\mathbf{E}[\tilde{x}^2])^{1/2} \geq \mathbf{E}(\tilde{x}).$$

Note: the expression in the left-hand side is called the quadratic expectation (or mean) of \tilde{x} .

(b) Let us define the generalized (or power) expectation (or mean) of a non-negative random variable \tilde{x} as

$$\mathbf{E}_p(\tilde{x}) = (\mathbf{E}[\tilde{x}^p])^{1/p}, \quad \text{with } p \in \mathbb{R}.$$

Note: $\mathbf{E}_1(\tilde{x})$ is the arithmetic mean of \tilde{x} and $\mathbf{E}_2(\tilde{x})$ is the quadratic mean of \tilde{x} .

Prove the following:

(i) $\mathbf{E}_0(\tilde{x}) \equiv \lim_{p \rightarrow 0} \mathbf{E}_p(\tilde{x}) = \mathbf{GE}(\tilde{x}).$

(ii) $\mathbf{E}_{-1}(\tilde{x}) = \mathbf{HE}(\tilde{x}).$

(iii) $\mathbf{E}_\infty(\tilde{x}) \equiv \lim_{p \rightarrow \infty} \mathbf{E}_p(\tilde{x}) = x_{\max}$, where $x_{\max} = \max\{\text{supp}(P_{\tilde{x}})\}.$

(iv) $\mathbf{E}_{-\infty}(\tilde{x}) \equiv \lim_{p \rightarrow -\infty} \mathbf{E}_p(\tilde{x}) = x_{\min}$, where $x_{\min} = \min\{\text{supp}(P_{\tilde{x}})\}.$

(v) If $q > p$, where $q \in \mathbb{R}$ and $p \in \mathbb{R}$, then $\mathbf{E}_q(\tilde{x}) \geq \mathbf{E}_p(\tilde{x}).$

37. Prove that the irrational number e satisfies the following:

(a)

$$e = \sum_{r=0}^{\infty} \frac{1}{r!} = \lim_{x \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x.$$

Hint: Use Taylor's expansion for the first equality and Newton's binomial theorem for the second.

(b) $e \in (2, 3)$. *Hint:* Compute the Lebesgue integral $\int_{(1,e]} (\ln x)^3 dx.$