4 Mathematical expectation and integral

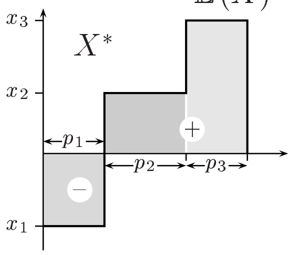
4a Bounded random variables, and Riemann integrals

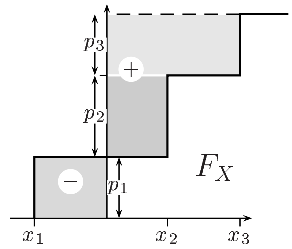
If a random variable X takes on only a finite number of values $x_1, \dots, x_n$, its expectation is defined by

(4a1) $$\mathbb{E}\,(X) = x_1 p_1 + \dots + x_n p_n$$

where $p_k = \mathbb{P}\,\big(\, X = x_k \,\big)$; that is well-known from Introduction to Probability. In terms of the quantile function,

(4a2) $$\mathbb{E}\,(X) = \int_0^1 X^*(p)\, dp\,.$$





In terms of the cumulative distribution function,

(4a3) $$\mathbb{E}\,(X) = -\int_{-\infty}^0 F_X(x)\, dx + \int_0^\infty \big(1 - F_X(x)\big)\, dx\,.$$

These formulas will be generalized to all bounded random variables (and then to some unbounded r.v.)

Recall the notion of Riemann integral (on $(0,1)$ or another bounded interval). Given a bounded function f, we consider all step functions f_-, f_+ satisfying $f_- \le f \le f_+$, and corresponding intervals $[\int f_-, \int f_+]$ (these elementary integrals being defined naturally). If the intersection of all the intervals contains only a single number, then the number is called the Riemann integral of f, denoted $\int f$, and f is called Riemann integrable. Otherwise f is not Riemann integrable. So,

(4a4) $$\int f = \sup \int f_- = \inf \int f_+\,.$$

It is well-known from Analysis that a continuous function on a closed bounded interval is always bounded and Riemann integrable. If a function has a finite number of discontinuities, but is bounded, then it is Riemann integrable. However, we are mostly interested in increasing functions. Fortunately, no continuity conditions are needed here.

4a5 Lemma. Any bounded increasing function $f : (0,1) \to [a,b]$ is Riemann integrable, and[1]

$$\int_0^1 f(p-)\, dp = \int_0^1 f(p)\, dp = \int_0^1 f(p+)\, dp\,.$$

[1] $\int f(p-)\, dp$ is just a Riemann integral, but the function is $p \mapsto f(p-)$, the left continuous version of f.

Proof. Given any $n \in \{1, 2, \dots\}$, we define step functions $f_-, f_+ : (0,1) \to [a,b]$ by

$$\begin{aligned} f_-(p) &= f\big(\tfrac{k-1}{n}\big) \\ f_+(p) &= f\big(\tfrac{k+2}{n}\big) \end{aligned} \qquad \text{for } p \in \big[\tfrac{k}{n}, \tfrac{k+1}{n}\big)\,;$$

here $k \in \{0, \dots, n-1\}$, and $f\big(\frac{-1}{n}\big) = a$, $f\big(\frac{n+1}{n}\big) = b$. We have

$$f_-(p) \le f(p-) \le f(p) \le f(p+) \le f_+(p)\,,$$

$$\int_0^1 f_-(p)\,dp = \frac{1}{n}\sum_{k=0}^{n-1} f\big(\tfrac{k-1}{n}\big)\,, \qquad \int_0^1 f_+(p)\,dp = \frac{1}{n}\sum_{k=0}^{n-1} f\big(\tfrac{k+2}{n}\big)\,,$$

$$\int f_+ - \int f_- = \frac{1}{n}\Big(f\big(\tfrac{n+1}{n}\big) + f(1) + f\big(\tfrac{n-1}{n}\big) - f\big(\tfrac{1}{n}\big) - f(0) - f\big(\tfrac{-1}{n}\big)\Big) \le \frac{3(b-a)}{n} \xrightarrow[n\to\infty]{} 0\,,$$

therefore all the three functions $f(p-), f(p), f(p+)$ are Riemann integrable, and all the three integrals are equal to $\sup \int f_- = \inf \int f_+$, where supremum and infimum are taken over functions f_-, f_+ considered above. $\square$

4a6 Lemma. If two random variables $X, Y : \Omega \to \mathbb{R}$ satisfy[2] $X \le Y$, then $X^* \le Y^*$ in the sense that

$$\forall p \in (0,1) \quad X^*(p-) \le Y^*(p-) \text{ and } X^*(p+) \le Y^*(p+)\,.$$

Proof. We have $Y \le u \implies X \le u$ for any u; hence $\{Y \le u\} \subset \{X \le u\}$ and $\mathbb{P}\big(Y \le u\big) \le \mathbb{P}\big(X \le u\big)$, that is,

$$\forall u \in \mathbb{R} \quad F_Y(u) \le F_X(u)\,.$$

However, $p \le F_X(u) \iff X^*(p-) \le u$ and $p \le F_Y(u) \iff Y^*(p-) \le u$ (see (2b13)). So, $Y^*(p-) \le u \implies p \le F_Y(u) \implies p \le F_X(u) \implies X^*(p-) \le u$ for all u; taking $u = Y^*(p-)$ we get

$$X^*(p-) \le Y^*(p-)\,.$$

The second inequality, $X^*(p+) \le Y^*(p+)$, follows from the first one, since $X^*(p+) = \lim_{\varepsilon\to 0+} X^*\big((p+\varepsilon)-\big)$, and the same for Y^*. $\square$

4a7 Definition. For any bounded[3] random variable X, its expectation $\mathbb{E}\,X$ is defined by

$$\mathbb{E}\,X = \int_0^1 X^*(p)\,dp\,,$$

where the integral is treated as Riemann integral.

According to Lemma 4a5, the definition is correct; the integral exists and does not depend on values of X^* at jumping points.

[2]As usual, "$X \le Y$" means $X(\omega) \le Y(\omega)$ for all $\omega \in \Omega$, or for almost all $\omega \in \Omega$, that is, $\mathbb{P}\big(\{\omega : X(\omega) \le Y(\omega)\}\big) = 1$. A set of zero probability may be neglected.

[3]That is, $\mathbb{P}\big(a \le X \le b\big) = 1$ for some $a, b \in \mathbb{R}$. A set of zero probability may be neglected, as usual.

Note some properties of expectation.

(4a8) $\mathbb{E}\, X = x_1 p_1 + \dots + x_n p_n$ if X takes on only a finite number of values;

(4a9) if X and Y are identically distributed, then $\mathbb{E}\, X = \mathbb{E}\, Y$;

(4a10) if $X \le Y$ then $\mathbb{E}\, X \le \mathbb{E}\, Y$.

Any other definition of expectation (for bounded random variables), satisfying (4a8), (4a9), (4a10) must be equivalent to Def. 4a7. Indeed, (4a9) combined with (2b17) implies

$$\mathbb{E}\, X = \mathbb{E}\, X^* .$$

Further, (4a8) implies $\mathbb{E}\, f = \int f$ for all step functions on $(0,1)$. Finally, (4a10) combined with (4a4) implies

$$\mathbb{E}\, X^* = \int X^* ;$$

so, the equality $\mathbb{E}\, X = \int X^*$ is deduced from (4a8), (4a9), (4a10).

4a11 Lemma. Let $X, Y : \Omega \to \mathbb{R}$ be two random variables, each having a finite number of values. Then

$$\mathbb{E}\,(X + Y) = \mathbb{E}\, X + \mathbb{E}\, Y .$$

This fact was proven in Introduction to Probability; I recall the idea of the proof:

$$\mathbb{P}\big(X = x_k,\, Y = y_l \big) = p_{kl} ;$$
$$\mathbb{E}\,(X + Y) = \sum_{k,l} (x_k + y_l) p_{kl} = \sum_{k,l} x_k p_{kl} + \sum_{k,l} y_l p_{kl} = \mathbb{E}\, X + \mathbb{E}\, Y .$$

4a12 Lemma. For any bounded random variable $X : \Omega \to \mathbb{R}$ and any $\varepsilon > 0$ there exist random variables $X_-, X_+ : \Omega \to \mathbb{R}$, each having a finite number of values, such that

$$X_- \le X \le X_+ ,$$
$$\mathbb{E}\,(X_+) - \mathbb{E}\,(X_-) \le \varepsilon .$$

Proof. Define[4]

$$X_-(\omega) = k\varepsilon, \quad X_+(\omega) = (k+1)\varepsilon \quad \text{when } X(\omega) \in [k\varepsilon, (k+1)\varepsilon) ,$$

then $X_+ \le X_- + \varepsilon$, therefore $\mathbb{E}\,(X_+) \le \mathbb{E}\,(X_-) + \varepsilon$. Boundedness of X ensures that k takes on only a finite number of values. □

Though $(X + Y)^*$ is not determined by X^*, Y^*, its integral is determined.

4a13 Theorem. For any bounded random variables $X, Y : \Omega \to \mathbb{R}$,[5]

$$\mathbb{E}\,(X + Y) = \mathbb{E}\, X + \mathbb{E}\, Y .$$

[4] Note that sets $\{\omega : X(\omega) \in [k\varepsilon, (k+1)\varepsilon)\}$ belong to the σ-field. More properties are neither needed nor ensured.

[5] See also 5e11.

Proof. By Lemma 4a12, for any $\varepsilon > 0$ there exist X_-, X_+, Y_-, Y_+, each having a finite number of values and such that $X_- \le X \le X_+$, $\mathbb{E}(X_+) - \mathbb{E}(X_-) \le \varepsilon$, $Y_- \le Y \le Y_+$, $\mathbb{E}(Y_+) - \mathbb{E}(Y_-) \le \varepsilon$. We have $X_- + Y_- \le X + Y \le X_+ + Y_+$, therefore (by Def. 4a7), $\mathbb{E}(X_- + Y_-) \le \mathbb{E}(X+Y) \le \mathbb{E}(X_+ + Y_+)$. By Lemma 4a11, $\mathbb{E}(X_- + Y_-) = \mathbb{E}(X_-) + \mathbb{E}(Y_-)$ and $\mathbb{E}(X_+ + Y_+) = \mathbb{E}(X_+) + \mathbb{E}(Y_+)$. So,

$$\mathbb{E}(X_-) + \mathbb{E}(Y_-) \le \mathbb{E}(X+Y) \le \mathbb{E}(X_+) + \mathbb{E}(Y_+) .$$

On the other hand, $\mathbb{E}(X_-) \le \mathbb{E}X \le \mathbb{E}(X_+)$ and $\mathbb{E}(Y_-) \le \mathbb{E}Y \le \mathbb{E}(Y_+)$, therefore

$$\mathbb{E}(X_-) + \mathbb{E}(Y_-) \le \mathbb{E}X + \mathbb{E}Y \le \mathbb{E}(X_+) + \mathbb{E}(Y_+) .$$

The interval $[\,\mathbb{E}(X_-) + \mathbb{E}(Y_-), \mathbb{E}(X_+) + \mathbb{E}(Y_+)\,]$ of length $\le 2\varepsilon$ contains both $\mathbb{E}(X+Y)$ and $\mathbb{E}X + \mathbb{E}Y$, therefore

$$\big| \mathbb{E}(X+Y) - (\mathbb{E}X + \mathbb{E}Y) \big| \le 2\varepsilon .$$

The inequality holds for all $\varepsilon > 0$, therefore $\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y$. $\square$

Also, $\mathbb{E}(cX) = c\mathbb{E}X$, which follows from (3a5) and Def. 4a7. Collect some properties of expectation:

(4a14) $\quad \mathbb{E}(cX + dY) = c\mathbb{E}X + d\mathbb{E}Y\,;$ (linearity)

(4a15) $\quad X \ge 0 \quad\Longrightarrow\quad \mathbb{E}X \ge 0\,;$ (positivity)

(4a16) $\quad X \ge 0 \text{ and } \mathbb{E}X = 0 \quad\Longrightarrow\quad X = 0\,.$ (strict positivity)

(Of course, "$X \ge 0$" means $\mathbb{P}\big(X \ge 0\big) = 1$, and "$X = 0$" means $\mathbb{P}\big(X = 0\big) = 1$.) Positivity is proven as follows: $X \ge 0 \implies X^* \ge 0 \implies \int X^* \ge 0$. Strict positivity follows from Markov inequality (already known for the discrete case from Introoduction to Probability):

(4a17) $$X \ge 0 \quad\Longrightarrow\quad \mathbb{P}\big(X \ge c\big) \le \frac{1}{c}\mathbb{E}X \quad \text{for } c \in (0,\infty)\,,$$

since $X \ge c\mathbf{1}_{X \ge c}$.

The following result generalizes (4a3). Again, integrals are taken in fact on finite intervals. Riemann integrability is ensured by monotonicity.

4a18 Lemma. For any bounded random variable X,

$$\mathbb{E}X = -\int_{-\infty}^0 F_X(x)\,dx + \int_0^\infty (1 - F_X(x))\,dx\,.$$

Proof. By Lemma 4a12, for any $\varepsilon > 0$ there exist X_-, X_+, each having a finite number of values and such that $X_- \le X \le X_+$, $\mathbb{E}(X_+) - \mathbb{E}(X_-) \le \varepsilon$. We have $F_{X_-} \ge F_X \ge F_{X_+}$, therefore, using (4a3),

$$\underbrace{-\textstyle\int_{-\infty}^0 F_{X_-} + \int_0^\infty (1 - F_{X_-})}_{\mathbb{E}X_-} \le -\textstyle\int_{-\infty}^0 F_X + \int_0^\infty (1 - F_X) \le \underbrace{-\textstyle\int_{-\infty}^0 F_{X_+} + \int_0^\infty (1 - F_{X_+})}_{\mathbb{E}X_+}\,.$$

The interval $[\mathbb{E}(X_-), \mathbb{E}(X_+)]$ of length $\le \varepsilon$ contains both $\mathbb{E}X$ and $-\int_{-\infty}^0 F_X + \int_0^\infty (1-F_X)$, therefore

$$\Big|\mathbb{E}X - \big(-\textstyle\int_{-\infty}^0 F_X + \int_0^\infty (1-F_X) \big)\Big| \le \varepsilon\,.$$

The inequality holds for all $\varepsilon > 0$, therefore $\mathbb{E}X = -\int_{-\infty}^0 F_X + \int_0^\infty (1-F_X)$. □

Smooth case. Assume that F_X is continuous on $[a,b]$, $F_X(a) = 0$, $F_X(b) = 1$, and the density $f_X = F_X'$ exists and is continuous on $[a,b]$.[6] Applying integration by parts to 4a18 we express the expectation in terms of the density:[7]

$$\begin{aligned}
\mathbb{E}X &= -\int_a^0 F_X(x)\,dx + \int_0^b \big(1-F_X(x)\big)\,dx = \\
&= -\bigg(xF_X(x)\Big|_a^0 - \int_a^0 x\,dF_X(x) \bigg) + \\
&\quad + \bigg(x\big(1-F_X(x)\big)\Big|_0^b - \int_0^b x\,d\big(1-F_X(x)\big) \bigg) = \\
&= \int_a^0 x f_X(x)\,dx + \int_0^b x f_X(x)\,dx
\end{aligned}$$

since $xF_X(x)$ vanishes at a and 0, and $x\big(1-F_X(x)\big)$ vanishes at 0 and b. Finally,

$$\mathbb{E}X = \int_{-\infty}^{+\infty} x f_X(x)\,dx\,. \tag{4a19}$$

Comparing it with the elementary formula (4a4), $\mathbb{E}X = \sum x_k p_k$, we see that $\int$ corresponds to $\sum$, x corresponds to x_k, $f_X(x)\,dx$ corresponds to p_k.

Sometimes the density f_X is continuous on (a,b) but not on $[a,b]$; limits $f_X(a+)$, $f_X(b-)$ may be infinite, or maybe they do not exist at all. In such cases we may use integration by parts on $[a+\varepsilon, b-\varepsilon]$, $\varepsilon > 0$, and take the limit $\varepsilon \to 0$, assuming (as before) that F_X is continuous on $[a,b]$ (no atoms):

$$\begin{aligned}
\mathbb{E}X &= \lim_{\varepsilon\to 0} \bigg(-\int_{a-\varepsilon}^0 F_X(x)\,dx + \int_0^{b-\varepsilon} \big(1-F_X(x)\big)\,dx \bigg) = \dots \\
\dots &= \lim_{\varepsilon\to 0} \int_{a+\varepsilon}^{b-\varepsilon} x f_X(x)\,dx\,.
\end{aligned}$$

Intermediate calculations are as before; say, $\lim_{\varepsilon\to 0}\big(xF_X(x)\big|_{a+\varepsilon}^0 = 0 - \lim_{\varepsilon\to 0}\big((a+\varepsilon)F_X(a+\varepsilon)\big) = 0$. Existence of $\lim_{\varepsilon\to 0}\int_{a+\varepsilon}^{b-\varepsilon} x f_X(x)\,dx$ is not self-evident, it follows from existence of $\lim_{\varepsilon\to 0}\big(-\int_{a-\varepsilon}^0 F_X(x)\,dx + \int_0^{b-\varepsilon}\big(1-F_X(x)\big)\,dx\big)$; here existence is self-evident, since integrands are bounded.

Can we conclude that (4a19) still holds? It depends on the meaning of integral. If it is Riemann integral then the answer is "yes" when $x f_X(x)$ is bounded, otherwise "no". However, there is a notion *"improper integral"* convenient in presence of singularities. Improper

[6] It means that the derivative $F_X'(x)$ exists for all $x \in (a,b)$ and has finite limits at a and b.

[7] I assume that $a < 0 < b$. Other cases are similar but simpler.

integral $\int_a^b \dots$ is defined as the limit $\lim_{\varepsilon\to 0}\int_{a+\varepsilon}^{b-\varepsilon}\dots$; existence of the limit means (by definition) existence of the improper integral. In that sense (4a19) holds when f_X is discontinuous at a, b. The same argument works if the density has a finite number of singular points. Say, if there are singularities at a, b, and some $c \in (a,b)$, then the improper integral is $\lim_{\varepsilon\to 0}\bigl(\int_{a+\varepsilon}^{c-\varepsilon} x f_X(x)\,dx + \int_{c+\varepsilon}^{b-\varepsilon} x f_X(x)\,dx\bigr)$.

4b Unbounded random variables, and integrability

If the function X^* is not bounded on $(0,1)$ then it cannot be Riemann integrable on $(0,1)$, but still it is Riemann integrable on any $(\varepsilon, 1-\varepsilon)$, and we may use the notion of improper integral. However, the integral may diverge, and we must be careful about possible cancellation of $(+\infty)$ and $(-\infty)$.

If X is bounded from below but unbounded from above,[8] then X^* is bounded near 0 but unbounded near 1, and we define

(4b1) $$\mathbb{E}X = \lim_{\varepsilon\to 0}\int_0^{1-\varepsilon} X^*(p)\,dp \quad \in (-\infty, +\infty]\,.$$

The limit exists due to monotonicity, but may be infinite. If the limit is finite then X is called *integrable*, and $\mathbb{E}X$ is a number. If the limit is $+\infty$ then X is called non-integrable, and $\mathbb{E}X = +\infty$.

If X is bounded from above but unbounded from below, the situation is similar:

(4b2) $$\mathbb{E}X = \lim_{\varepsilon\to 0}\int_\varepsilon^{1} X^*(p)\,dp \quad \in [-\infty, +\infty)\,.$$

If X is unbounded both from above and from below, the situation is more complicated. Expectation is *not* defined as $\lim_{\varepsilon\to 0}\int_\varepsilon^{1-\varepsilon} X^*(p)\,dp$; rather, it is defined by

(4b3) $$\mathbb{E}X = \lim_{\varepsilon\to 0}\int_\varepsilon^{p_0} X^*(p)\,dp + \lim_{\varepsilon\to 0}\int_{p_0}^{1-\varepsilon} X^*(p)\,dp\,,$$

where $p_0 \in (0,1)$ may be chosen arbitrarily (it does not affect the result). If both limits are finite, then X is called integrable, and $\mathbb{E}X$ is a number. Otherwise X is called non-integrable. If the first limit is finite while the second is infinite, then $\mathbb{E}X = +\infty$. If the first limit is $(-\infty)$ while the second limit is finite then $\mathbb{E}X = -\infty$. If the first limit is $(-\infty)$ while the second is $(+\infty)$ then $\mathbb{E}X$ is not defined (never write $\mathbb{E}X = \infty$ in this case).

The formula $\mathbb{E}X = \lim_{\varepsilon\to 0}\int_\varepsilon^{1-\varepsilon} X^*(p)\,dp$ gives correct answers in three cases (number + number = number; number + $(+\infty) = +\infty$; $(-\infty)$ + number $= -\infty$) but can give a wrong answer in the fourth case ($(-\infty) + (+\infty)$ is undefined).

[8]Sets of zero probability may be neglected, as usual.

4b4 Example. Consider a random variable X distributed exponentially with parameter 1 (as in 2c8):

$$\begin{aligned} F_X(x) &= 1 - e^{-x} && \text{for } x \in [0,\infty)\,, \\ X^*(p) &= -\ln(1-p) = \ln\frac{1}{1-p} && \text{for } p \in (0,1)\,, \\ f_X(x) &= e^{-x} && \text{for } x \in (0,\infty)\,. \end{aligned}$$

We have

$$\mathbb{E}\,X = \lim_{\varepsilon\to 0}\int_0^{1-\varepsilon} X^*(p)\,dp = -\lim_{\varepsilon\to 0}\int_0^{1-\varepsilon}\ln(1-p)\,dp\,;$$

the integral may be calculated by parts: $\int \ln(1-p)\,dp = p\ln(1-p) - \int p\,d\ln(1-p) = p\ln(1-p) + \int p\cdot\frac{1}{1-p}\,dp = p\ln(1-p) + \int\bigl(-1+\frac{1}{1-p}\bigr)\,dp = p\ln(1-p) - p - \ln(1-p)$; thus,

$$\begin{aligned} \mathbb{E}\,X &= -\lim_{\varepsilon\to 0}\bigl(p\ln(1-p) - p - \ln(1-p)\bigr)\Big|_0^{1-\varepsilon} = -\lim_{\varepsilon\to 0}\bigl((1-\varepsilon)\ln\varepsilon - 1 + \varepsilon - \ln\varepsilon\bigr) = \\ &= \lim_{\varepsilon\to 0}(\varepsilon\ln\varepsilon) + 1 = 1\,, \end{aligned}$$

since $\varepsilon\ln\varepsilon \to 0$ for $\varepsilon\to 0$, which is well-known from Analysis:[9] $\ln\varepsilon$ tends to $-\infty$ much slower than ε tends to 0.

It would be easier to calculate $\mathbb{E}\,X$ via 4a18 or (4a19); however, do they work for unbounded random variables? Let X be unbounded from above, but bounded from below, say, $X \ge 0$.

4b5 Lemma. For any $X \ge 0$,[10]

$$\lim_{\varepsilon\to 0}\int_0^{1-\varepsilon} X^*(p)\,dp = \lim_{M\to\infty}\int_0^M \bigl(1 - F_X(x)\bigr)\,dx \quad \in [0,+\infty]\,.$$

Proof. Denote for short $I_1(\varepsilon) = \int_0^{1-\varepsilon} X^*(p)\,dp$ and $I_2(M) = \int_0^M \bigl(1-F_X(x)\bigr)\,dx$; we have to prove that $I_1(0+) = I_2(\infty)$.[11] For any $M \in (0,\infty)$ introduce the *truncated* random variable

$$X_M = \min(M,X) = \begin{cases} X & \text{if } X \le M, \\ M & \text{if } X \ge M. \end{cases}$$

[9] And easy to guess from numeric examples; say, $0.000\,001\cdot\log_{10} 0.000\,001 = -0.000\,006$, a small number.
[10] Either both limits are finite and equal, or they are both $+\infty$.
[11] Of course, $I_1(0+)$ means $\lim_{\varepsilon\to 0+} I_1(\varepsilon)$ and $I_2(\infty)$ means $\lim_{M\to+\infty} I_2(M)$.

We have

$$F_{X_M}(x) = \begin{cases} F_X(x) & \text{for } x < M, \\ 1 & \text{for } X \ge M; \end{cases}$$

$$X_M^*(p) = \min\bigl(M, X^*(p)\bigr) = \begin{cases} X^*(p) & \text{if } X^*(p) \le M, \\ M & \text{if } X^*(p) \ge M, \end{cases}$$

$$\mathbb{E}\, X_M = \int_0^1 \min\bigl(M, X^*(p)\bigr)\, dp\,;$$

$$\mathbb{E}\, X_M = \underbrace{\int_0^M \bigl(1 - F_X(x)\bigr)\, dx}_{I_2(M)} \qquad \text{(by (4a18))}\,.$$

Comparing the two formulas for $\mathbb{E}\, X_M$ we see that $I_2(M)$ may be written as an integral in p, not x,[12]

$$I_2(M) = \int_0^1 \min\bigl(M, X^*(p)\bigr)\, dp\,.$$

In order to prove that $I_1(0+) = I_2(\infty)$ we'll prove both $I_1(0+) \le I_2(\infty)$ and $I_1(0+) \ge I_2(\infty)$.

Proof of $I_1(0+) \le I_2(\infty)$. If $M \ge X^*(1-\varepsilon)$ then

$$\underbrace{\int_0^{1-\varepsilon} X^*(p)\, dp}_{I_1(\varepsilon)} = \int_0^{1-\varepsilon} \min\bigl(M, X^*(p)\bigr)\, dp \le \underbrace{\int_0^1 \min\bigl(M, X^*(p)\bigr)\, dp}_{I_2(M)}\,,$$

that is, $I_1(\varepsilon) \le I_2(M)$ for M large enough. It follows that $I_1(\varepsilon) \le I_2(\infty)$ for all ε, therefore $I_1(0+) \le I_2(\infty)$.

Proof of $I_1(0+) \ge I_2(\infty)$. We have

$$\underbrace{\int_0^{1-\varepsilon} X^*(p)\, dp}_{I_1(\varepsilon)} \ge \int_0^{1-\varepsilon} \min\bigl(M, X^*(p)\bigr)\, dp = \underbrace{\int_0^1 \min\bigl(M, X^*(p)\bigr)\, dp}_{I_2(M)} - \underbrace{\int_{1-\varepsilon}^1 \min\bigl(M, X^*(p)\bigr)\, dp}_{\le M\varepsilon}\,,$$

that is, $I_1(\varepsilon) \ge I_2(M) - M\varepsilon$. It follows that $I_1(0+) \ge I_2(M)$ for all M, therefore $I_1(0+) \ge I_2(\infty)$. □

4b6 Example. Let X be as in 4b4, then

$$\mathbb{E}\, X = \int_0^\infty \bigl(1 - F_X(x)\bigr)\, dx = \int_0^\infty e^{-x}\, dx = 1\,.$$

Turn to (4a19) for $X \ge 0$ unbounded from above. Integration by parts gives

$$I_2(M) = \int_0^M \bigl(1 - F_X(x)\bigr)\, dx = x\bigl(1 - F_X(x)\bigr)\Big|_0^M - \int_0^M x\, d\bigl(1 - F_X(x)\bigr) = \\ = M\bigl(1 - F_X(M)\bigr) + \int_0^M x\, f_X(x)\, dx\,.$$

[12] That is a good luck, since $I_1(\varepsilon)$ is defined as an integral in p.

Denote $\varepsilon = 1 - F_X(M)$, then M is an $(1-\varepsilon)$-quantile of X; in terms of X^* it means that $X^*\big((1-\varepsilon)-\big) \le M \le X^*\big((1-\varepsilon)+\big)$. Therefore

$$\begin{gathered} I_2(M) = \int_0^1 \min\big(M, X^*(p)\big)\,dp = \int_0^{1-\varepsilon} \dots + \int_{1-\varepsilon}^1 \dots = \int_0^{1-\varepsilon} X^*(p)\,dp + M\varepsilon\,; \\ \int_0^M x\,f_X(x)\,dx = I_2(M) - M\varepsilon = \int_0^{1-\varepsilon} X^*(p)\,dp = I_1(\varepsilon)\,; \end{gathered}$$

so,

$$\int_0^M x\,f_X(x)\,dx = I_1\big(1 - F_X(M)\big) \to \mathbb{E}\,X \quad \text{for } M \to \infty\,.$$

We have now three improper integrals for the expectation of $X : \Omega \to [0,\infty)$:

$$\mathbb{E}\,X = \int_0^1 X^*(p)\,dp = \int_0^\infty \big(1 - F_X(x)\big)\,dx = \int_0^\infty x\,f_X(x)\,dx \tag{4b7}$$

(of course, the third integral is for the smooth case only). Similarly, for $X : \Omega \to (-\infty, +\infty)$,

$$\mathbb{E}\,X = \int_0^1 X^*(p)\,dp = -\int_{-\infty}^0 F_X(x)\,dx + \int_0^\infty \big(1 - F_X(x)\big)\,dx = \int_{-\infty}^{+\infty} x\,f_X(x)\,dx\,; \tag{4b8}$$

if one of these formulas gives $(-\infty) + (+\infty)$, then the other two formulas do the same.

4b9 Example. Let X be as in 4b6, then

$$\mathbb{E}\,X = \int_0^\infty x\,f_X(x)\,dx = \int_0^\infty x\,e^{-x}\,dx = 1\,.$$

4c Moments

If X is a random variable then X^2 is also a random variable, and we may consider its expectation $\mathbb{E}\,(X^2)$, called the second moment of X. (Do not confuse $\mathbb{E}\,(X^2)$ with $(\mathbb{E}\,X)^2$; their difference is the variance of X, as you know from Introduction to Probability.) More generally, we may ask about $\mathbb{E}\,\varphi(X)$ for some function φ. Of course, we may deal only with the new random variable $Y = \varphi(X)$, calculate Y^*, or F_Y, or f_Y, and then calculate $\mathbb{E}\,Y$. However, the following formula is more convenient:

$$\mathbb{E}\,\varphi(X) = \int_0^1 \varphi\big(X^*(p)\big)\,dp \tag{4c1}$$

for any $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\varphi\big(X^*(\cdot)\big)$ is Riemann integrable and $\varphi(X)$ is a random variable.[13] Sketch of the proof: first, (4c1) holds for every *step* function φ (since X and X^* are identically distributed, as well as $\varphi(X)$ and $\varphi(X^*)$). Second, consider step functions φ_-, φ_+ such that $\varphi_- \le \varphi \le \varphi_+$, and proceed as before.

[13] See also 5e2.

If φ is piecewise monotone and bounded on the support of X, then $\varphi\big(X^*(\cdot)\big)$ is Riemann integrable, and (4c1) is applicable. If φ is unbounded (on the support of X) then we treat the integral in (4c1) as improper.

Moments:

$$\text{(4c2)}\qquad \begin{aligned} \mathbb{E}\,X^{2n} &= \int_0^1 \big(X^*(p)\big)^{2n}\,dp \in [0,+\infty]\,; \\ \mathbb{E}\,X^{2n+1} &= \int_0^1 \big(X^*(p)\big)^{2n+1}\,dp \in \mathbb{R} \qquad \text{if } \int_0^1 |X^*(p)|^{2n+1}\,dp < \infty\,. \end{aligned}$$

Moment generating function:

$$\text{(4c3)}\qquad \mathrm{MGF}_X(t) = \mathbb{E}\,e^{tX} = \int_0^1 e^{tX^*(p)}\,dp \in (0,+\infty]\,.$$

Smooth case.

If a random variable $X : \Omega \to [a,b]$ has a density f_X continuous and strictly positive on $[a,b]$, then we may convert an integral in p into integral in x:

$$x = X^*(p), \quad p = F_X(x), \quad dp = f_X(x)\,dx,$$
$$\int_0^1 \varphi\big(X^*(p)\big)\,dp = \int_a^b \varphi(x)\,f_X(x)\,dx,$$

and (4c1) turns into

$$\text{(4c4)}\qquad \mathbb{E}\,\varphi(X) = \int_{-\infty}^{+\infty} \varphi(x) f_X(x)\,dx\,.$$

For unbounded X and/or φ,

$$\text{(4c5)}\qquad \mathbb{E}\,\varphi(X) = \int_{-\infty}^{+\infty} \varphi(x) f_X(x)\,dx \qquad \text{if } \int_{-\infty}^{+\infty} |\varphi(x)| f_X(x)\,dx < \infty\,,$$

integrals being improper. In particular, moments and their generating function:

$$\text{(4c6)}\qquad \begin{aligned} \mathbb{E}\,X^{2n} &= \int_{-\infty}^{+\infty} x^{2n} f_X(x)\,dx \in [0,+\infty]\,; \\ \mathbb{E}\,X^{2n+1} &= \int_{-\infty}^{+\infty} x^{2n+1} f_X(x)\,dx \in \mathbb{R} \qquad \text{if } \int_{-\infty}^{+\infty} |x|^{2n+1} f_X(x)\,dx < \infty\,; \\ \mathrm{MGF}_X(t) &= \int_{-\infty}^{+\infty} e^{tx} f_X(x)\,dx \in (0,+\infty]\,. \end{aligned}$$

Note that the density $f_X(x)$ is always outside of the function φ. Compare it with elementary formulas of Introduction to Probability: $\mathbb{E}\,X^n = \sum_k x_k^n p_k$, $\mathbb{E}\,e^{tX} = \sum_k e^{tx_k} p_k$.

4c7 Example. Cauchy distribution (see 1c1): $X^*(p) = \tan \pi(\omega - \frac{1}{2})$, $F_X(x) = \frac{1}{2} + \frac{1}{\pi} \arctan x$, $f_X(x) = \frac{1}{\pi} \frac{1}{1+x^2}$. Does $\mathbb{E} X$ exist? The function X^* is unbounded near 0 and 1. Let $p = 1 - \varepsilon$, $\varepsilon \to 0$, then $X^*(p) = \tan \pi(\frac{1}{2} - \varepsilon) = \tan(\frac{\pi}{2} - \pi\varepsilon) = \frac{1}{\tan \pi\varepsilon} \sim \frac{1}{\pi} \cdot \frac{1}{\varepsilon}$, thus $\int_{p_0}^1 X^*(p)\,dp = +\infty$. Similarly, $\int_0^{p_0} X^*(p)\,dp = -\infty$. It means that $\mathbb{E} X$ is undefined. However, $\sqrt[3]{X}$ is integrable, and $\mathbb{E} \sqrt[3]{X} = 0$. Indeed, $\sqrt[3]{X^*(1-\varepsilon)} \sim \text{const} \cdot \frac{1}{\varepsilon^{1/3}}$, thus $\int_{p_0}^1 \sqrt[3]{X^*(p)}\,dp < \infty$; similarly, $\int_0^{p_0} \sqrt[3]{X^*(p)}\,dp > -\infty$; and $\int_0^1 \sqrt[3]{X^*(p)}\,dp = 0$ by symmetry: $X^*(1-p) = -X^*(p)$. Alternatively, we may use density; $\int x f_X(x)\,dx = \frac{1}{\pi} \int \frac{x}{1+x^2}\,dx$ diverges near $\pm\infty$, since $\frac{x}{1+x^2} \sim \frac{1}{x}$ for $x \to \pm\infty$. However, $\frac{\sqrt[3]{x}}{1+x^2} \sim \frac{1}{x^{5/3}}$, thus $\int \frac{\sqrt[3]{x}}{1+x^2}\,dx$ converges near $\pm\infty$, and is equal to 0 due to symmetry: $f_X(-x) = f_X(x)$.

You may ask, what is really wrong in saying that $\mathbb{E} X = 0$ by symmetry? Well, let us perform a numeric experiment. Take a sample of, say, 1000 independent values of X. To this end, we take 1000 independent random numbers $\omega_1, \dots, \omega_{1000}$ according to the uniform distribution on $(0, 1)$, and calculate $x_k = \tan \pi(\omega_k - \frac{1}{2})$; here is my sample (sorted):

k	1	2	3	4	...	997	998	999	1000
ω_k	.000703	.001391	.002178	.002693	...	.991265	.992999	.995051	.998198
x_k	-452.5	-228.7	-146.1	-118.1	...	36.4	45.4	64.3	176.6
$\sqrt[3]{x_k}$	-7.677	-6.115	-5.267	-4.907	...	3.315	3.568	4.006	5.610

Calculate mean values:

$$\frac{x_1 + \dots + x_{1000}}{1000} = -1.23; \qquad \frac{\sqrt[3]{x_1} + \dots + \sqrt[3]{x_{1000}}}{1000} = 0.0021$$

You see, the mean of $\sqrt[3]{x_k}$ is close to 0, which cannot be said about the mean of x_k. The experiment confirms the theory; $\mathbb{E} \sqrt[3]{X} = 0$, but $\mathbb{E} X$ is undefined. In fact, it is well-known that $\frac{1}{n}(X_1 + \dots + X_n)$ has the Cauchy distribution, again (with no scaling).

Cauchy	Normal $N(0,1)$
$\dfrac{X_1 + \dots + X_n}{n} \sim X_1$	$\dfrac{X_1 + \dots + X_n}{n} \sim \dfrac{1}{\sqrt{n}} X_1$

In general, $\mathbb{E} |X|^{n+1} < \infty$ implies $\mathbb{E} |X|^n < \infty$. If the moment generating function is finite on a neighborhood $(-\varepsilon, \varepsilon)$ of 0, then $\mathbb{E} |X|^n < \infty$ for all n.

For every random variable X,

$$e^{tX} = 1 + tX + \frac{1}{2!} t^2 X^2 + \dots ;$$

if X is bounded then the series converges uniformly, therefore

$$\text{MGF}_X(t) = 1 + t\mathbb{E}(X) + \frac{1}{2!} t^2 \mathbb{E}(X^2) + \dots \tag{4c8}$$

For unbounded X it is still true, provided that $\text{MGF}_X(\pm\varepsilon) < \infty$ for some $\varepsilon > 0$.

4c9 Example. The exponential distribution, again (see (4b4), (4b6), (4b9)): $f_X(x) = e^{-x}$ for $x > 0$;

$$\mathbb{E}\, e^{tX} = \int_0^\infty e^{tx} e^{-x}\,dx = \int_0^\infty e^{-(1-t)x}\,dx ;$$

for $t < 1$ the integral converges to $\frac{1}{1-t}$,

$$\mathrm{MGF}_X(t) = \frac{1}{1-t} = 1 + t + t^2 + \dots$$

Comparing it with (4c8) we get

$$\mathbb{E}\, X^n = n!$$

4c10 Example. Let X be uniform on $(0,1)$, then $f_X(x) = 1$ for $x \in (0,1)$; $\mathbb{E}\, e^{tX} = \int_0^1 e^{tx}\,dx$;

$$\mathrm{MGF}_X(t) = \frac{e^t - 1}{t} = \frac{1}{t}\biggl(-1 + 1 + t + \frac{t^2}{2!} + \frac{t^3}{3!} + \dots \biggr) = 1 + \frac{t}{2!} + \frac{t^2}{3!} + \dots\,.$$

Comparing it with (4c8) we get

$$\mathbb{E}\, X^n = \frac{1}{n+1}$$

which is easy to find without MGF.