Continuous random variable

Chung Chan cchan@inc.cuhk.edu.hk

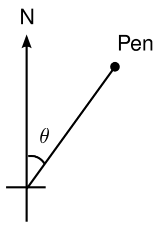

Fig. 1. Pen-drop example

I. Probability measure

Given a sample space Ω, the probability measure $\Pr$ satisfies the following Komogorov axioms:

1) $\Pr(B) \geq 0$ for event $B \in \Omega$.
2) $\Pr(\Omega) = 1$.
3) $\sum_i \Pr(\mathcal{E}_i) = \Pr(\bigcup_i \mathcal{E}_i)$ for any sequence $E_1, E_2, \dots \subseteq \Omega$ of disjoint events, i.e. $E_i \cap E_j = \emptyset$.

Example 1 (Pen-drop) A pen is dropped on the ground. The angle θ it points to with respect to the North direction is from the sample space $\Omega = (0, 2\pi]$ as shown in Fig. 1. You receive a reward of

$$\mathsf{X}(\theta) := \frac{\theta}{2\pi} \qquad \theta \in \Omega$$

Every angle is assumed to be equally likely. □

Some questions of interest are:

1) How much do you earn on average?
   i.e. the expectation $\mathrm{E}[\mathsf{X}]$.
2) What is the chance that you earn more than average?
   i.e. $\Pr\{\mathsf{X} > \mathrm{E}[\mathsf{X}]\}$.
3) What is the variance?
   i.e. $\mathrm{Var}[\mathsf{X}] = \mathrm{E}[\mathsf{X}^2] - \mathrm{E}[\mathsf{X}]^2$.

X is a random variable that maps Ω to the unit interval $(0, 1]$. The probability mass function (PMF) is defined as

$$P_{\mathsf{X}}(\alpha) := \Pr\{\mathsf{X} = \alpha\}$$

By the uniformity assumption that all angles are equally likely, we have $P_{\mathsf{X}}(\alpha) = p \geq 0$ independent of $\alpha \in (0, 1]$.

Claim 1 $\Pr\{\mathsf{X} = \alpha\} = 0$ *for all* $\alpha \in (0, 1]$ □

Proof Suppose to the contrary that $\Pr\{\mathsf{X} = \alpha\} = p > 0$. Choose a positive integer k sufficiently large such that $kp > 1$. An possible choice is $k = \lceil \frac{1}{p} + 1 \rceil$. Choose a finite set $A := \{\alpha_1, \alpha_2, \dots, \alpha_k\}$ of distinct values from $(0, 1]$. Then,

$$\Pr\{\mathsf{X} \in A\} = \sum_{i=1}^{k} \Pr\{\mathsf{X} = \alpha_i\} = kp > 1.$$

We have the desired contradiction since

$$\Pr\{\mathsf{X} \in A\} = 1 - \Pr\{\mathsf{X} \notin A\} \leq 1.$$

The first equality follows from the first and third axioms, while the second inequality follows from the second axiom. ■

Can we compute the expectation using the following formula?

$$\mathrm{E}[\mathsf{X}] = \sum_{\alpha \in (0,1]} \alpha \underbrace{P_{\mathsf{X}}(\alpha)}_{=0}. \tag{1}$$

Is the sum equal to 0? The answers are no, because the sum is over an uncountably infinite set, and is therefore not well-defined. The following false claim has the same issue.

Claim 2 (False) *The pen never drops to any angle. i.e. it stands up vertically, perpendicular to the ground.* □

Proof (False) Since $\Pr(\theta) = 0$ for any $\theta \in \Omega$, it is impossible that the pen drops to any angle. i.e.

$$\Pr\{\emptyset\} + \sum_{\theta \in \Omega} \underbrace{\Pr(\theta)}_{=0} \overset{(*)}{=} \Pr\{\Omega\} = 1$$

where the first and second equalities are by the third and the second axioms respectively. It follows that $\Pr\{\emptyset\} = 1$, and so the pen does not drop to any angle. ■

What is wrong in the above proof is the equality (*). The third axiom is valid only to a sequence of disjoint events. The set Ω of all angles is uncountable, and therefore cannot be enumerated as a sequence of singletons. Indeed, we can write $\Pr\{\emptyset\} + \Pr\{\Omega\} = \Pr\{\Omega\}$ by the third axiom and deduce that $\Pr\{\emptyset\} = 0$ instead.

II. Cumulative distribution function (CDF)

The main reason why the expectation cannot be computed from PMF in Example 1 is because $P_{\mathsf{X}}(\alpha) = 0$ does not say anything about the distribution of X. e.g. $\Pr\{\mathsf{X} \in (a, b]\}$ for $0 < a < b \leq 1$ cannot be computed from $\sum_{\alpha \in (a,b]} P_{\mathsf{X}}(\alpha)$ since the sum is not well-defined. What we need is a better characterization of the distribution.

The CDF F_{X} of X is defined as

$$F_{\mathsf{X}}(\alpha) := \Pr\{\mathsf{X} \leq \alpha\} \qquad \alpha \in \mathbb{R} \tag{2}$$

It follows that the probability of other events can be obtained from the CDF. e.g.

$$\begin{aligned} \Pr\{\mathsf{X} \in (a, b]\} &= \Pr\{\mathsf{X} \leq b\} - \Pr\{\mathsf{X} \leq a\} \\ &= F_{\mathsf{X}}(b) - F_{\mathsf{X}}(a) \end{aligned}$$

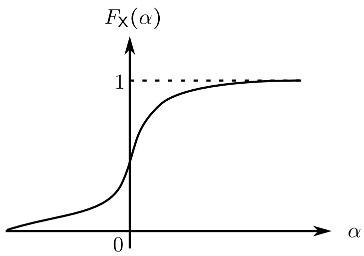

Fig. 2. A typical CDF (2)

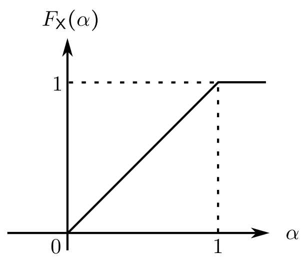

Fig. 3. CDF in (3) for Example 1

The typical shape of a CDF is shown in Fig. 2, with

$$\begin{aligned} \lim_{\alpha\to-\infty} F_{\mathsf{X}}(\alpha) &= \Pr(\emptyset) = 0 \\ \lim_{\alpha\to\infty} F_{\mathsf{X}}(\alpha) &= \Pr(\Omega) = 1 \\ F_{\mathsf{X}}(\alpha) &\leq F_{\mathsf{X}}(\beta) \qquad\qquad \forall \alpha \leq \beta \end{aligned}$$

The CDF for Example 1 is,

$$F_{\mathsf{X}}(\alpha) = \begin{cases} 0 & \alpha \leq 0 \\ \alpha & \alpha \in (0,1] \\ 1 & \alpha > 1 \end{cases} \tag{3}$$

which is shown in Fig. 3. In particular, $F_{\mathsf{X}}(\alpha) = \alpha$ for $(0,1]$ is by the uniformity assumption that every angle is equally likely.

How do we compute $\mathrm{E}[\mathsf{X}]$ from the cdf? The idea is quantization, i.e. rounding X up or down to certain decimal places. Consider Example 1, and divide the unit interval $(0,1]$ into n parts as shown in Fig. 4, i.e.

$$(0,1] = \bigcup_{i=1}^{n} (\tfrac{i-1}{n}, \tfrac{i}{n}].$$

Set $\underline{\mathsf{X}} = \frac{i-1}{n}$ and $\overline{\mathsf{X}} = \frac{i}{n}$ if and only if $\mathsf{X} \in (0,1]$. Since $\underline{\mathsf{X}} \leq \mathsf{X} \leq \overline{\mathsf{X}}$ by definition, we have

$$\mathrm{E}[\underline{\mathsf{X}}] \leq \mathrm{E}[\mathsf{X}] \leq \mathrm{E}[\overline{\mathsf{X}}]. \tag{4}$$

Note that $\underline{\mathsf{X}}$ is a discrete random variable, and so its expectation can be computed easily from its PMF by (1).

$$\begin{aligned} \mathrm{E}[\underline{\mathsf{X}}] &= \sum_{i=1}^{n} \frac{i-1}{n} \underbrace{\Pr\{\mathsf{X} \in (\tfrac{i-1}{n}, \tfrac{i}{n}]\}}_{=F_{\mathsf{X}}(\frac{i}{n}) - F_{\mathsf{X}}(\frac{i-1}{n}) = \frac{1}{n}} \\ &= \frac{1}{n^2} \sum_{i=1}^{n} (i-1) \\ &= \frac{1}{2}\left[1 - \frac{1}{n}\right] \xrightarrow{n\to\infty} \frac{1}{2} \end{aligned}$$

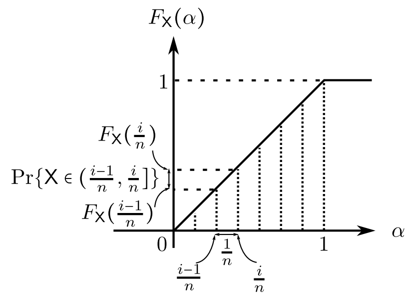

Fig. 4. Quantizing $\alpha \in (0,1]$ for (3)

Similarly, we can show that

$$\mathrm{E}[\overline{\mathsf{X}}] = \frac{1}{2}\left[1 + \frac{1}{n}\right] \xrightarrow{n\to\infty} \frac{1}{2},$$

and so $\mathrm{E}[\mathsf{X}] = \frac{1}{2}$ by (4).

For general CDF, we can partition the real line into intervals of size Δ as shown in Fig. 5, and compute the expectation as the limit

$$\mathrm{E}[\mathsf{X}] = \lim_{\Delta\to 0} \sum_{i=-\infty}^{\infty} (i-1)\Delta[F_{\mathsf{X}}(i\Delta) - F_{\mathsf{X}}((i-1)\Delta)].$$

In the language of Calculus, the above limit gives the following definite integral,

$$\mathrm{E}[\mathsf{X}] = \int_{-\infty}^{\infty} \beta dF_{\mathsf{X}}(\beta). \tag{5}$$

The term in the integral corresponds to the area of the yellow bar illustrated in Fig. 6. Integrating the term with respect to β over the entire real line, the quantity is equal to the positive area of the yellow region (i.e. for $\beta \geq 0$) in Fig. 7 minus the positive area of the red region (i.e. for $\beta \leq 0$). We can also exchange the axes as shown in Fig. 8 and see that the integral is the area of the inverse F_{X}^{-1} of the CDF. i.e.

$$\begin{aligned} \mathrm{E}[\mathsf{X}] &= \int_0^{\infty} (1 - F_{\mathsf{X}}(\beta))d\beta - \int_{-\infty}^{0} F_{\mathsf{X}}(\beta)d\beta \qquad (6) \\ &= \int_0^1 F_{\mathsf{X}}^{-1}(y)dy \qquad (7) \end{aligned}$$

For the CDF in (3) for Example 1, we have for $F_{\mathsf{X}}^{-1}(y) = y$ for $y \in (0,1)$ and so

$$\mathrm{E}[\mathsf{X}] = \int_0^1 y dy = \frac{1}{2}.$$

As expected, this is computing the area of the upper triangle in Fig. 3 bounded by the y-axis and the CDF.

Consider an alternative award of

$$\mathsf{X}_1 := -\ln \mathsf{X}.$$

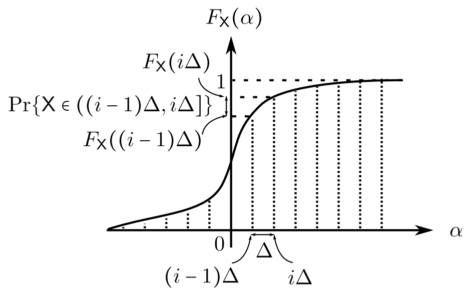

Fig. 5. Quantizing $\alpha \in \mathbb{R}$ for general CDF.

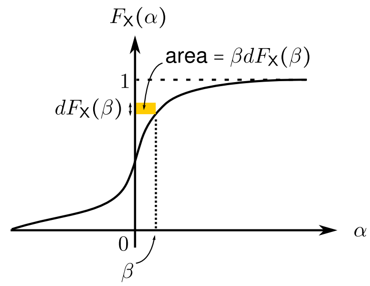

Fig. 6. A term in the integral for the expectation in (5).

Will you choose X_1 over X? To compute the expectation $\mathrm{E}[\mathsf{X}_1]$, we first compute the inverse CDF as follows.

$$\begin{aligned}
F_{\mathsf{X}_1}(\alpha) &= \Pr\{\mathsf{X}_1 \leq \alpha\} \\
&= \Pr\{-\ln \mathsf{X} \leq \alpha\} \\
&= \Pr\{\ln \mathsf{X} \geq -\alpha\} \\
&= \Pr\{\mathsf{X} \geq e^{-\alpha}\} \\
&= 1 - F_{\mathsf{X}}(e^{-\alpha}) \\
&= 1 - \begin{cases} 0 & e^{-\alpha} \leq 0 \\ e^{-\alpha} & e^{-\alpha} \in (0,1] \\ 1 & e^{-\alpha} \geq 1 \end{cases} && \text{by (3)} \\
&= 1 - e^{-\alpha} && \alpha \in [0, \infty) \\
F_{\mathsf{X}_1}^{-1}(y) &= -\ln(1-y) && y \in (0,1)
\end{aligned}$$

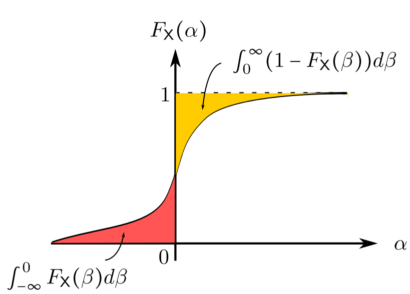

Fig. 7. Expectation (5) from CDF.

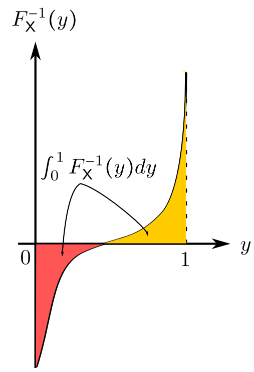

Fig. 8. Expectation (5) from inverse CDF.

The expectation can be computed from (7) as follow.

$$\begin{aligned}
\mathrm{E}[\mathsf{X}_1] &= \int_0^1 -\ln(1-y)dy \\
&= \int_0^1 \ln(1-y)d(1-y) \\
&= \int_1^0 \ln z \, dz && \text{with z:=1-y} \\
&= z \ln z|_1^0 - \int_1^0 z \, d \ln z && \text{integration by parts} \\
&= 1.
\end{aligned}$$

Thus, X_1 has a larger expectation than X.

Consider another reward defined as

$$\mathsf{X}_2 := \mathsf{X} + \ln \mathsf{X}.$$

Does it have a larger expectation? Note that the CDF for X_2 is difficult to compute because the inverse of the function $x \mapsto x + \ln x$ is not simple. However, by the linearity of expectation,

$$\mathrm{E}[\mathsf{X}_2] = \mathrm{E}[\mathsf{X}] + \mathrm{E}[\ln \mathsf{X}] = \mathrm{E}[\mathsf{X}] - \mathrm{E}[\mathsf{X}_1] = \frac{1}{2} - 1 = -\frac{1}{2}.$$

This is clearly not a good reward.

III. Probability density function

Consider yet another reward defined as

$$\mathsf{X}_3 := -\mathsf{X} \ln \mathsf{X}.$$

Once again, finding the CDF is difficult because the inverse of $x \mapsto -x \ln x$ is not simple. Even though $\mathsf{X}_3 = \mathsf{X}\mathsf{X}_1$, we cannot equate $\mathrm{E}[\mathsf{X}_3]$ to $\mathrm{E}[\mathsf{X}]\,\mathrm{E}[\mathsf{X}_1] = \frac{1}{2}$ because X and X_1 are not independent.

To compute this expectation, we define another characterization of the distribution, called the probability density function (PDF).

The PDF f_{X} of X is a function that satisfies

$$F_{\mathsf{X}}(\alpha) = \int_{-\infty}^{\alpha} f_{\mathsf{X}}(\beta) d\beta \qquad \forall \beta \in \mathbb{R}. \tag{8}$$

Probability of $\mathsf{X} \in (a,b]$ corresponds to the area under the PDF f_{X} over $(a,b]$. The expectation of any function $g(\mathsf{X})$ can be computed as

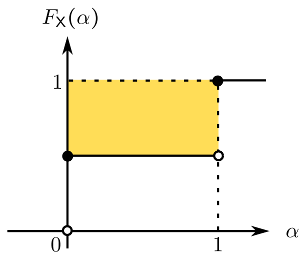

Fig. 9. The CDF of a uniformly random bit

$$\mathrm{E}[g(\mathsf{X})] = \int_{-\infty}^{\infty} g(\beta) f_{\mathsf{X}}(\beta) d\beta. \tag{9}$$

The question is whether f_{X} exists, and if so, how to find it. Assuming f_{X} exists, then

$$\frac{d}{d\alpha} F_{\mathsf{X}}(\alpha) = f_{\mathsf{X}}(\alpha)$$

for all α at which F_{X} is differentiable. For the CDF in (3) for Example 1, we have

$$f_{\mathsf{X}}(\alpha) = \begin{cases} 0 & \alpha < 0 \\ 1 & \alpha \in (0,1) \\ 0 & \alpha > 1 \end{cases}$$

The expectation of X_3 can be computed from (9) as follows.

$$\mathrm{E}[\mathsf{X}_3] = \mathrm{E}[-\mathsf{X} \ln \mathsf{X}] = \int_0^1 -\beta \ln \beta d\beta = \frac{1}{4}.$$

Note that this is smaller than $\mathrm{E}[\mathsf{X}] = \frac{1}{2}$, $\mathrm{E}[\mathsf{X}_1] = 1$ and also $\mathrm{E}[\mathsf{X}] \, \mathrm{E}[\mathsf{X}_1] = \frac{1}{2}$.

Although PDF gives us a simple way to compute expectation of complicated functions of random variables, it does not always exist. For example, consider a uniformly random bit X with the following CDF as shown in Fig. 9.

$$F_{\mathsf{X}}(\alpha) = \begin{cases} 0 & \alpha < 0 \\ \frac{1}{2} & \alpha \in [0,1) \\ 1 & \alpha \geq 1 \end{cases}$$

Taking the derivative, we have $f_{\mathsf{X}}(\alpha) = 0$ for $\alpha \in \mathbb{R} \setminus \{0,1\}$. However, the area under the curve is zero, and therefore does not satisfy (8). Note also that the expectation of $\frac{1}{2}$ can be computed from the CDF as (7) but not the PDF (9). The problem is due to the discontinuity of the CDF at 0 and 1. It can be shown that if the CDF is absolutely continuous (which is a stronger notion than continuity), the desired PDF satisfying (8) exists. This motivates the following definition.

X is a continuous random variable if its CDF is absolutely continuous.