ECS315 2012/1 Part V Dr.Prapun

12 Three Types of Random Variables

12.1. Review: You may recall[47] the following properties for cdf of discrete random variables. These properties hold for any kind of random variables.

(a) The cdf is defined as $F_X(x) = P[X \le x]$. This is valid for any type of random variables.

(b) Moreover, the cdf for any kind of random variable must satisfies three properties which we have discussed earlier:

CDF1 F_X is non-decreasing

CDF2 F_X is right continuous

CDF3 $\lim\limits_{x\to-\infty} F_X(x) = 0$ and $\lim\limits_{x\to\infty} F_X(x) = 1$.

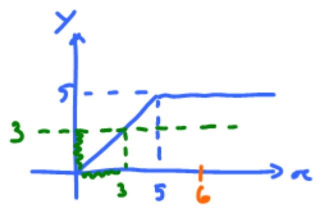
(c) $P[X = x] = F_X(x) - F_X(x^-)$ = the jump or saltus in F at x.

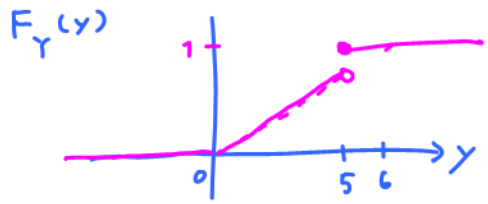
Theorem 12.2. If you find a function F that satisfies CDF1, CDF2, and CDF3 above, then F is a cdf of some random variable.

[47] If you don't know these properties by now, you should review them as soon as possible.

Example 12.3. Consider an input X to a device whose output Y will be the same as the input if the input level does not exceed 5. For input level that exceeds 5, the output will be saturated at 5. Suppose $X \sim \mathcal{U}(0,6)$. Find $F_Y(y)$.

$X \to \boxed{g(\cdot)} \to Y$





For $0 < y < 5$

$F_Y(3) = P[Y \leq 3] = P[X \leq 3] = \frac{1}{2}$

$F_Y(y) = \int_0^y f_X(x)\,dx = \frac{y}{6}.$

12.4. We can categorize random variables into three types according to its cdf:

(a) If $F_X(x)$ is piecewise flat with discontinuous jumps, then X is **discrete**.

(b) If $F_X(x)$ is a continuous function, then X is **continuous**.

(c) If $F_X(x)$ is a piecewise continuous function with discontinuities, then X is **mixed**.

cdf.

discrete

cont

mixed

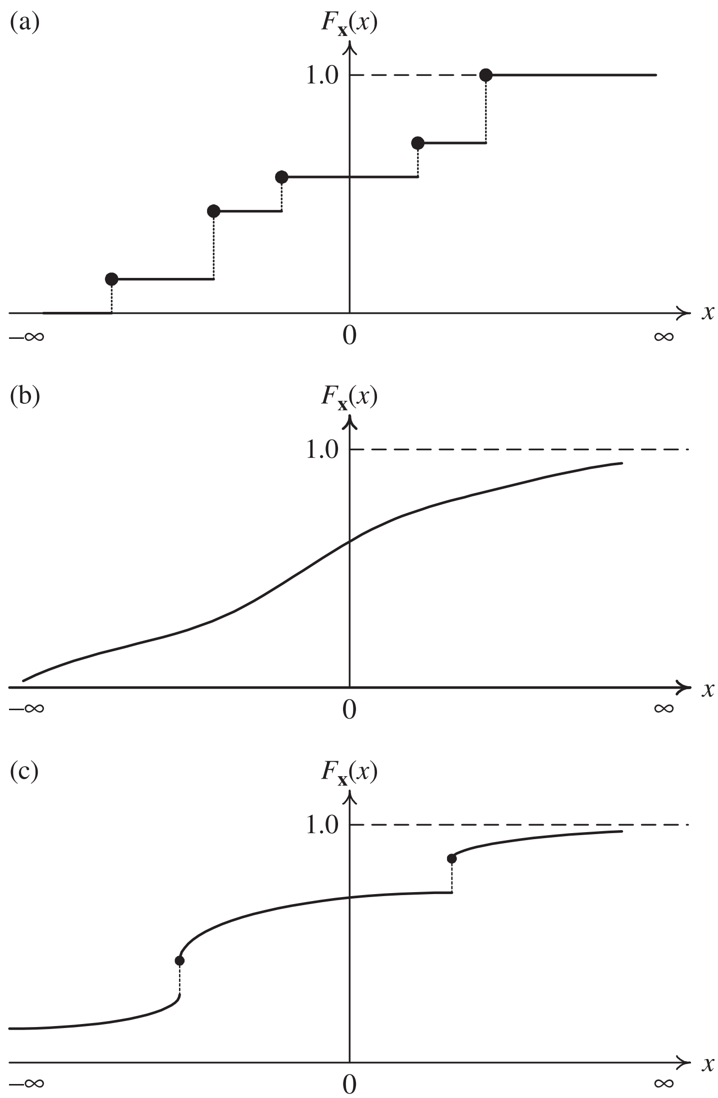


Figure 21: Typical cdfs: (a) a discrete random variable, (b) a continuous random variable, and (c) a mixed random variable [17, Fig. 3.2].

We have seen in Example 12.3 that some function can turn a continuous random variable into a mixed random variable. Next, we will work on an example where a continuous random variable is turned into a discrete random variable.

Example 12.5. Let $X \sim \mathcal{U}(0,1)$ and $Y = g(X)$ where

$$g(x) = \begin{cases} 1, & x < 0.6 \\ 0, & x \geq 0.6. \end{cases}$$

Before going deeply into the math, it is helpful to think about the nature of the derived random variable Y. The definition of $g(x)$ tells us that Y has only two possible values, $Y = 0$ and $Y = 1$. Thus, Y is a discrete random variable.

$P[Y=0] = P[X \geq 0.6] = 0.4$

$P[Y=1] = 1 - 0.4 = 0.6$ $\quad$ $Y \sim \text{Bernoulli}(0.6)$

Example 12.6. In MATLAB, we have the `rand` command to generate $\mathcal{U}(0,1)$. If we want to generate a Bernoulli random variable with success probability p, what can we do?

$Y = [\text{rand} < p]$

13 Transform methods: Characteristic Functions and Moment Generating functions

Definition 13.1. The characteristic function of a random variable X is defined by

$$\varphi_X(v) = \mathbb{E}\left[e^{jvX}\right].$$

Remarks:

(a) If X is a continuous random variable with density f_X, then

$$\varphi_X(v) = \int_{-\infty}^{+\infty} e^{jvx} f_X(x)dx,$$

which is the *Fourier transform* of f_X evaluated at $-v$. More precisely,

$$\varphi_X(v) = \mathcal{F}\{f_X\}(\omega)\big|_{\omega=-v}. \tag{35}$$

(b) Many references use u or t instead of v.

Example 13.2. You may have learned that the Fourier transform of a Gaussian waveform is a Gaussian waveform. In fact, when $X \sim \mathcal{N}(m, \sigma^2)$,

$$\mathcal{F}\{f_X\}(\omega) = \int_{-\infty}^{\infty} f_X(x)\, e^{-j\omega x} dx = e^{-j\omega m - \frac{1}{2}\omega^2\sigma^2}.$$

Using (35), we have

$$\varphi_X(v) = e^{jvm - \frac{1}{2}v^2\sigma^2}.$$

Example 13.3. For $X \sim \mathcal{E}(\lambda)$, we have $\varphi_X(v) = \frac{\lambda}{\lambda - jv}$.

$$\varphi_X(v) = \mathbb{E}[e^{jvX}] = \int_0^\infty e^{jvx} \lambda e^{-\lambda x} dx = \lambda \int_0^\infty e^{-(\lambda - jv)x} dx$$

As with the Fourier transform, we can build a large list of commonly used characteristic functions. (You probably remember that rectangular function in time domain gives a sinc function in frequency domain.) When you see a random variable that has the same form of characteristic function as the one that you know, you can quickly make a conclusion about the type and property of that random variable.

Example 13.4. Suppose a random variable X has the characteristic function $\varphi_X(v) = \frac{2}{2-jv}$. You can quickly conclude that it is an exponential random variable with parameter 2.

For many random variables, it is easy to find its expected value or any moments via the characteristic function. This can be done via the following result.

13.5. $\varphi^{(k)}(v) = j^k \mathbb{E}\left[X^k e^{jvX}\right]$ and $\varphi^{(k)}(0) = j^k \mathbb{E}\left[X^k\right]$.

$$\varphi_X(v) = \mathbb{E}\left[e^{jvX}\right]$$

$$\frac{d}{dv}\varphi_X(v) = \mathbb{E}\left[\frac{d}{dv}e^{jvX}\right] = \mathbb{E}\left[jXe^{jvX}\right]$$

$$\varphi_X'(0) = j\mathbb{E}X$$

Example 13.6. When $X \sim \mathcal{E}(\lambda)$,

(a) $\mathbb{E}X = \frac{1}{\lambda}$.

(b) $\operatorname{Var} X = \frac{1}{\lambda^2}$. $\mathbb{E}[X^2]$

Exercise 13.7 (F2011)**.** Continue from Example 13.2.

(a) Show that for $X \sim \mathcal{N}(m, \sigma^2)$, we have

  (i) $\mathbb{E}X = m$

  (ii) $\mathbb{E}\left[X^2\right] = \sigma^2 + m^2$.

(b) for $X \sim \mathcal{N}(3, 4)$, find $\mathbb{E}\left[X^3\right]$.

One very important properties of characteristic function is that it is very easy to find the characteristic function of a sum of independent random variables.

13.8. Suppose $X \perp\!\!\!\perp Y$. Let $Z = X + Y$. Then, the characteristic function of Z is the product of the characteristic functions of X and Y:

$$\varphi_Z(v) = \varphi_X(v)\varphi_Y(v)$$

$$\varphi_Z(\omega) = \mathbb{E}\left[e^{j\omega Z}\right] = \mathbb{E}\left[e^{j\omega(X+Y)}\right] = \mathbb{E}\left[e^{j\omega X} e^{j\omega Y}\right] \overset{\text{indp}}{=} \mathbb{E}\left[e^{j\omega X}\right]\mathbb{E}\left[e^{j\omega Y}\right]$$

Remark: Can you relate this property to the property of the Fourier transform?

Example 13.9. Use 13.8 to show that the sum of two independent Gaussian random variables is still a Gaussian random variable:

Exercise 13.10. Continue from Example 10.41. Suppose $\Lambda_1 \sim \mathcal{P}(\lambda_1)$ and $\Lambda_2 \sim \mathcal{P}(\lambda_2)$ are independent. Let $\Lambda = \Lambda_1 + \Lambda_2$. Use 13.8 to show that $\Lambda \sim \mathcal{P}(\lambda_1 + \lambda_2)$.

Exercise 13.11. Continue from Example 10.42 Suppose $B_1 \sim \mathcal{B}(n_1, p)$ and $B_2 \sim \mathcal{B}(n_2, p)$ are independent. Let $B = B_1 + B_2$. Use 13.8 to show that $B \sim \mathcal{B}(n_1 + n_2, p)$.

14 Limiting Theorems

14.1 Law of Large Numbers (LLN)

Definition 14.1. Let $X_1, X_2, \ldots, X_n$ be a collection of random variables with a common mean $\mathbb{E}[X_i] = m$ for all i. In practice, since we do not know m, we use the numerical average, or sample mean,

$$M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

in place of the true, but unknown value, m.

Q: Can this procedure of using M_n as an estimate of m be justified in some sense?

A: This can be done via the law of large number.

14.2. The law of large number basically says that if you have a sequence of i.i.d random variables $X_1, X_2, \ldots$. Then the sample means $M_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ will converge to the actual mean as $n \to \infty$.

(expected value)

14.3. LLN is easy to see via the property of variance. Note that

$$\mathbb{E}[M_n] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}X_i = m$$

and

$$\operatorname{Var}[M_n] = \operatorname{Var}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{Var} X_i = \frac{1}{n}\sigma^2, \tag{36}$$

Remarks:

(a) For (36) to hold, it is sufficient to have uncorrelated X_i's.

(b) From (36), we also have

$$\sigma_{M_n} = \frac{1}{\sqrt{n}}\sigma. \tag{37}$$

In words, "when uncorrelated (or independent) random variables each having the same distribution are averaged together, the standard deviation is reduced according to the square root law." [22, p 142].

Exercise 14.4 (F2011)**.** Consider i.i.d. random variables $X_1, X_2, \ldots, X_{10}$. Define the sample mean M by

$$M = \frac{1}{10}\sum_{k=1}^{10} X_k.$$

Let

$$V_1 = \frac{1}{10}\sum_{k=1}^{10} \left(X_k - \mathbb{E}\left[X_k\right]\right)^2.$$

and

$$V_2 = \frac{1}{10}\sum_{j=1}^{10} \left(X_j - M\right)^2.$$

Suppose $\mathbb{E}\left[X_k\right] = 1$ and $\mathrm{Var}[X_k] = 2$.

(a) Find $\mathbb{E}\left[M\right]$.

(b) Find $\mathrm{Var}[M]$.

(c) Find $\mathbb{E}\left[V_1\right]$.

(d) Find $\mathbb{E}\left[V_2\right]$.

14.2 Central Limit Theorem (CLT)

In practice, there are many random variables that arise as a sum of many other random variables. In this section, we consider the sum

$$S_n = \sum_{i=1}^{n} X_i \tag{38}$$

where the X_i are i.i.d. with common mean m and common variance σ^2.

- Note that when we talk about X_i being i.i.d., the definition is that they are independent and identically distributed. It is then convenient to talk about a random variable X which shares the same distribution (pdf/pmf) with these X_i. This allow us to write

$$X_i \overset{\text{i.i.d.}}{\sim} X, \tag{39}$$

which is much more compact than saying that the X_i are i.i.d. with the same distribution (pdf/pmf) as X. Moreover, we can also use $\mathbb{E}X$ and σ_X^2 for the common expected value and variance of the X_i.

Q: How does S_n behave?

For the S_n defined above, there are many cases for which we know the pmf/pdf of S_n.

Example 14.5. When the X_i are i.i.d. Bernoulli(p),

Example 14.6. When the X_i are i.i.d. $\mathcal{N}(m, \sigma^2)$,

Note that it is not difficult to find the characteristic function of S_n if we know the common characteristic function $\varphi_X(v)$ of the X_i:

$$\varphi_{S_n}(v) = (\varphi_X(v))^n .$$

If we are lucky, as in the case for the sum of Gaussian random variables in Example 14.6 above, we get $\varphi_{S_n}(v)$ that is of the form that we know. However, $\varphi_{S_n}(v)$ will usually be something we haven't seen before or difficult to find the inverse transform. This is one of the reason why having a way to approximate the sum S_n would be very useful.

There are also some situations where the distribution of the X_i is unknown or difficult to find. In which case, it would be amazing if we can say something about the distribution of S_n.

In the previous section, we consider the sample mean of identically distributed random variables. More specifically, we consider the random variable $M_n = \frac{1}{n} S_n$. We found that M_n will converge to m as n increases to ∞. Here, we don't want to rescale the sum S_n by the factor $\frac{1}{n}$.

14.7 (Approximation of densities and pmfs using the CLT). The actual statement of the CLT is a bit difficult to state. So, we first give you the interpretation/insight from CLT which is very easy to remember and use:

For n large enough, we can approximate S_n by a Gaussian random variable with the same mean and variance as S_n.

Note that the mean and variance of S_n is nm and $n\sigma^2$, respectively. Hence, for n large enough we can approximate S_n by $\mathcal{N}\left(nm, n\sigma^2\right)$. In particular,

(a) $F_{S_n}(s) \approx \Phi\left(\frac{s-nm}{\sigma\sqrt{n}}\right)$.

(b) If the X_i are continuous random variable, then

$$f_{S_n}(s) \approx \frac{1}{\sqrt{2\pi}\sigma\sqrt{n}} e^{-\frac{1}{2}\left(\frac{x-nm}{\sigma\sqrt{n}}\right)^2}.$$

(c) If the X_i are integer-valued, then

$$P[S_n = k] = P\left[k - \frac{1}{2} < S_n \le k + \frac{1}{2}\right] \approx \frac{1}{\sqrt{2\pi}\sigma\sqrt{n}} e^{-\frac{1}{2}\left(\frac{k-nm}{\sigma\sqrt{n}}\right)^2}.$$

[9, eq (5.14), p. 213].

The approximation is best for k near nm [9, p. 211].

Example 14.8. Approximation for Binomial Distribution: For $X \sim \mathcal{B}(n,p)$, when n is large, binomial distribution becomes difficult to compute directly because of the need to calculate factorial terms.

(a) When p is not close to either 0 or 1 so that the variance is also large, we can use CLT to approxmiate

$$P[X = k] \approx \frac{1}{\sqrt{2\pi \operatorname{Var} X}} e^{-\frac{(k-\mathbb{E}X)^2}{2\operatorname{Var} X}} \tag{40}$$

$$= \frac{1}{\sqrt{2\pi np(1-p)}} e^{-\frac{(k-np)^2}{2np(1-p)}}. \tag{41}$$

This is called Laplace approximation to the Binomial distribution [26, p. 282].

(b) When p is small, the binomial distribution can be approximated by $\mathcal{P}(np)$ as discussed in Section **??**.

(c) If p is very close to 1, then $n - X$ will behave approximately Poisson.

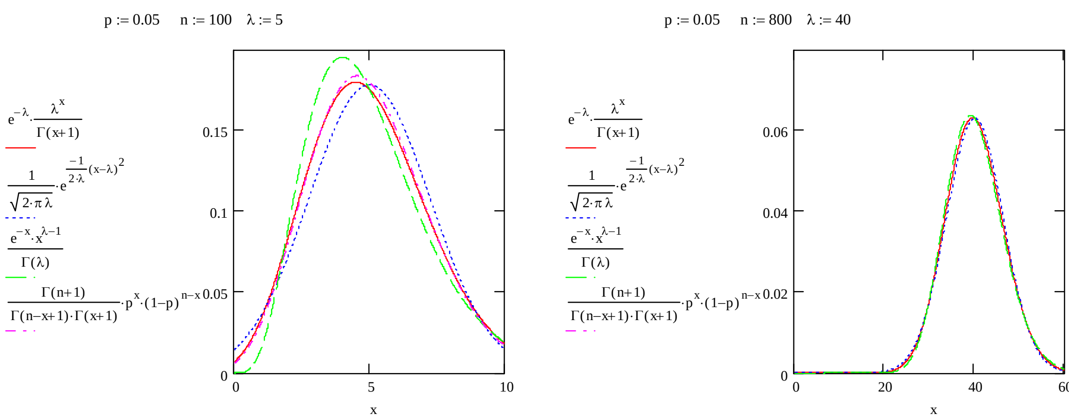


Figure 22: Gaussian approximation to Binomial, Poisson distribution, and Gamma distribution.

Exercise 14.9 (F2011)**.** Continue from Exercise 6.50. The stronger person (Kakashi) should win the competition if n is very large. (By the law of large numbers, the proportion of fights that Kakashi wins should be close to 55%.) However, because the results are random and n can not be very large, we can not guarantee that Kakashi will win. However, it may be good enough if the probability that Kakashi wins the competition is greater than 0.85.

We want to find the minimal value of n such that the probability that Kakashi wins the competition is greater than 0.85.

Let N be the number of fights that Kakashi wins among the n fights. Then, we need

$$P\left[N > \frac{n}{2}\right] \geq 0.85. \tag{42}$$

Use the central limit theorem and Table 3.1 or Table 3.2 from [Yates and Goodman] to approximate the minimal value of n such that (42) is satisfied.

14.10. A more precise statement for CLT can be expressed via the convergence of the characteristic function. In particular, suppose that $(X_k)_{k\geq 1}$ is a sequence of i.i.d. random variables with mean m and variance $0 < \sigma^2 < \infty$. Let $S_n = \sum_{k=1}^{n} X_k$. It can be shown that

(a) the characteristic function of $\frac{S_n - mn}{\sigma\sqrt{n}}$ converges pointwise to the characteristic function of $\mathcal{N}(0,1)$ and that

(b) the characteristic function of $\frac{S_n - mn}{\sqrt{n}}$ converges pointwise to the characteristic function of $\mathcal{N}(0,\sigma)$.

To see this, let $Z_k = \frac{X_k - m}{\sigma} \overset{\text{iid}}{\sim} Z$ and $Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Z_k$. Then, $\mathbb{E}Z = 0$, $\operatorname{Var} Z = 1$, and $\varphi_{Y_n}(t) = \left(\varphi_Z\left(\frac{t}{\sqrt{n}}\right)\right)^n$. By approximating $e^x \approx 1 + x + \frac{1}{2}x^2$. We have $\varphi_X(t) \approx 1 + jt\mathbb{E}X - \frac{1}{2}t^2\mathbb{E}\left[X^2\right]$ and

$$\varphi_{Y_n}(t) = \left(1 - \frac{1}{2}\frac{t^2}{n}\right)^n \to e^{-\frac{t^2}{2}},$$

which is the characteristic function of $\mathcal{N}(0,1)$.

- The case of Bernoulli(1/2) was derived by Abraham de Moivre around 1733. The case of Bernoulli(p) for $0 < p < 1$ was considered by Pierre-Simon Laplace [9, p. 208].