Continuous Random Variables

Henry D. Pfister

ECE 581 Random Signals and Noise

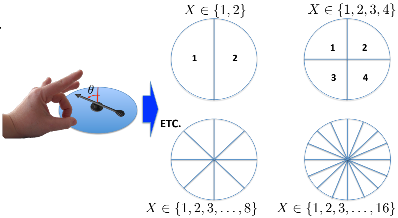
Lecture 4

Slides courtesy of Christ Richmond with slight modifications.

Table of contents

A Large Number of Outcomes

- Consider a game spinner



- Divide disc into n sectors ("pie slices") of equal size, and define discrete r.v. X as the number for sector that spinner arrow falls in.
  - Let X PMF be $P_X(x) = \frac{1}{n}$ for $X \in \{1, 2, \ldots, n\}$, and zero otherwise.
  - Clearly, as $n \to \infty$, i.e. more and more pie slices allowed:
    - $\#$ sectors becomes infinite
    - $\Pr(X = x) = P_X(x) = \frac{1}{n} \to 0$, i.e. probability of particular sector is zero.

      −This does not mean that obtaining an outcome of $X = x$ is impossible, only that there are infinite possibilities.
    - So how might we handle such scenarios, or similar ones?

Continuous Sample Spaces

- What about r.v. given by angle $\theta \in [0, 2\pi]$?...or more generally what about a r.v. X defined on the real interval $[a, b]$, i.e. $S_X = \{x : a \leq x \leq b\}$?
  - The real interval $[a, b]$ contains an infinite # (uncountable) possible values for X.
  - By axioms of probability, the total probability on $[a, b]$ must be unity.
  - So how might we handle such scenarios?...or in general a r.v. whose value may fall anywhere within continuous ranges?

Probability Density Function (PDF)

- The PDF of a r.v. X is denoted $f_X(x)$ and defined such that, for any reasonable set $B \subseteq \mathbb{R}$,

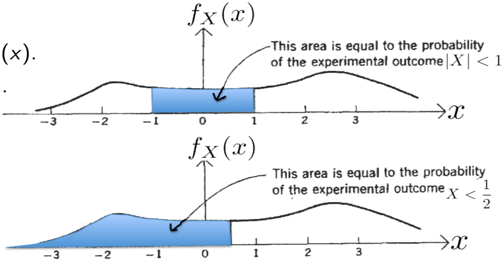
$$\Pr(X \in B) = \int_B f_X(x)dx,$$

i.e. the probability of X falling in set B is given by integrating the PDF over the set B.

- For example, if $B = \{x : x_1 \le x \le x_2\}$ then

$$\Pr(X \in B) = \Pr(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f_X(x)dx.$$

  - Notation $X \sim f_X(x)$ indicates that r.v. X has PDF $f_X(x)$.
  - Probability is given by area under the density function.



Probability Density Function (PDF)

- The area under $f_X(x)$ for interval of width zero, i.e. a single point, is equal to zero:

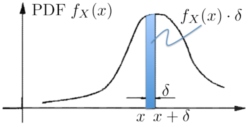
$$\Pr(X = x_1) = \int_{x_1}^{x_1} f_X(x)dx = 0. \tag{1}$$

  - Not because outcome $X = x_1$ is impossible but due to infinitely many possibilities nearby x_1.
  - Thus, including or excluding endpoints in an interval has no effect on the probability, i.e.

$$\Pr(x_1 \le X \le x_2) = \Pr(x_1 < X \le x_2) = \Pr(x_1 \le X < x_2) = \Pr(x_1 < X < x_2).$$

- The axioms of probability require the PDF to possess the following properties:
  a) $f_X(x) \ge 0$ for all $x \in S_X$ [otherwise negative probability would occur].
  b) $\int_{-\infty}^{\infty} f_X(x)\,dx = \Pr(-\infty < X < \infty) = 1$ [probability of universal set is unity].

- Small $\delta > 0$,



$$\Pr(x < X \le x + \delta) = \int_{x}^{x+\delta} f_X(u)du \approx f_X(x) \cdot \delta \tag{2}$$

  - PDF $f_X(x)$ provides a measure of "probability per unit length." not probability
  - Thus, used to compute but not equal to a probability and not restricted to be ≤ 1.

Example PDF: Uniform R.V.

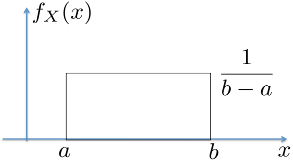
- Consider r.v. X with range $S_X = \{x : a \le x \le b\}$ with PDF

$$f_X(x) = \begin{cases} \gamma, & a \le x \le b \\ 0, & \text{Otherwise.} \end{cases}$$

Note that $$\int_a^b f_X(x)\,dx = \gamma \int_a^b dx = \gamma \cdot (b-a) = 1 \implies \gamma = \frac{1}{b-a}.$$

- Thus, r.v. uniformly distributed on interval $[a, b]$ has PDF

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{Otherwise.} \end{cases} \quad (3)$$



  - Indicator function $\mathbb{1}_S(x) \triangleq \{1,$ if $x \in S$, and 0 otherwise $\implies f_X(x) = \frac{1}{b-a}\mathbb{1}_{[a,b]}(x)$.
  - For $b = a + \epsilon$ and small $\epsilon > 0$, we have $[a, b] = [a, a+\epsilon]$ and $b - a = \epsilon$ and $f_X(x) = \frac{1}{\epsilon}$ for $x \in [a, a+\epsilon]$
  - Thus, PDF can assume arbitrarily large values near a for small enough $\epsilon > 0$.

Cumulative Distribution Function (CDF)

- The cumulative distribution function (CDF) for r.v. $X \sim f_X(x)$ is defined by

$$F_X(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f_X(u)\, du \tag{4}$$

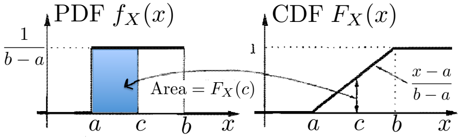
which is very similar to the discrete r.v. case.

  - The CDF "accumulates" all probability up to and including the value x.

- **Example: Uniform r.v.:** Consider r.v. X with PDF given in (3).

$$F_X(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f_X(u)\, du = \frac{1}{b-a} \int_{-\infty}^{x} \mathbb{1}_{[a,b]}(u) du$$

$$= \begin{cases} 0, & x < a \\ \dfrac{1}{b-a} \displaystyle\int_a^x du = \dfrac{x-a}{b-a}, & a \leq x \leq b \\ 1, & x > b. \end{cases}$$



Quiz 4.2

The cumulative distribution function of the random variable Y is

$$F_Y(y) = \begin{cases} 0 & y < 0, \\ y/4 & 0 \le y \le 4, \\ 1 & y > 4. \end{cases} \tag{1}$$

Sketch the CDF of Y and calculate the following probabilities:

(a) $P[Y \le -1]$

(b) $P[Y \le 1]$

(c) $P[2 < Y \le 3]$

(d) $P[Y > 1.5]$

Expected Values: Mean, Variance, Functions of R.V.

- The mean or expected value of a continuous r.v. $X \sim f_X(x)$ is defined as

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x) dx$$

- The expected value of $g(X)$, a function of r.v. $X \sim f_X(x)$ is defined as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx.$$

- The variance of continuous r.v. $X \sim f_X(x)$ is defined as

$$\text{Var}[X] = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 \cdot f_X(x) dx \triangleq \sigma_X^2.$$

  As in the discrete r.v. case, it follows that

$$0 \leq \text{Var}[X] = E[X^2] - E[X]^2.$$

- The standard deviation is the $\sigma_X = \sqrt{\text{Var}[X]}$.

Expected Values: Mean, Variance, Functions of R.V.

- The n-th moment and n-th central moment are respectively defined as

$$E[X^n] = \int_{-\infty}^{\infty} x^n \cdot f_X(x)dx, \text{ and } E[(X - E[X])^n] = \int_{-\infty}^{\infty} (x - E[X])^n \cdot f_X(x)dx.$$

- It can be shown, as in discrete r.v. case, that if $Y = aX + b$ where $a, b \in \mathbb{R}$ then

$$E[Y] = aE[X] + b, \;\; \text{Var}[Y] = a^2 \text{Var}[X].$$

- **Example:** Assume r.v. X is uniform on $[a, b]$. What is $E[X]$ and $\text{Var}[X]$?

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx = \frac{1}{b-a} \int_a^b x\, dx = \frac{1}{b-a} \left[\frac{1}{2} x^2 \Big|_a^b \right] = \frac{1}{(b-a)} \cdot \frac{(b^2 - a^2)}{2} = \frac{b+a}{2}$$

$$E[X^2] = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\, dx = \frac{1}{b-a} \int_a^b x^2\, dx = \frac{1}{b-a} \left[\frac{1}{3} x^3 \Big|_a^b \right] = \frac{1}{(b-a)} \cdot \frac{(b^3 - a^3)}{3}$$

$$= \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} = \frac{b^2 + ab + a^2}{3}$$

Expected Values: Mean, Variance, Functions of R.V.

- **Example Cont.:** The variance is

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{b+a}{2}\right)^2$$
$$= \frac{4(b^2 + ab + a^2) - 3(b^2 + 2ab + a^2)}{12} = \frac{b^2 + a^2 - 2ab}{12} = \frac{(b-a)^2}{12}.$$

- Note that the standard deviation is

$$\frac{1}{2\sqrt{3}}(b-a) = 0.29 \times (b-a) \approx \frac{1}{3} \times (\text{width of interval support}).$$

  - Thus, $\Pr(X \in [E[X] - \sigma_X, E[X] + \sigma_X]) = \frac{1}{\sqrt{3}} = 0.5774$; i.e. not quite, but close enough to $\sim \frac{2}{3} = 0.67$.

- Try calculating the mean and variance of $3X + \pi$...

Families of Continuous Random Variables

- The uniform distribution is discussed above as an example. So, we summarize briefly next.

Uniform Random Variable:

- X is said to be a uniform (a, b) r.v. if it has PDF given by (3). It has mean $(b+a)/2$ and variance $(b-a)^2/12$.

Exponential Random Variables:

- X is an exponential (λ) r.v. if for parameter $\lambda > 0$ it has PDF

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{Otherwise}, \end{cases} \tag{5}$$

  - Plot graph of PDF[1]...
- The CDF for an exponential (λ) r.v. is obtained via

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du = \begin{cases} 0, & x < 0 \\ \int_0^x \lambda e^{-\lambda u} du, & x \geq 0. \end{cases}$$

[1] Recall $e^{-1} = 0.3679$, $e^{-2} = 0.1353$, and $e^{-3} = 0.0498$, etc.

Families of Continuous R.V.: Exponential PDF

- Note that exponential CDF follows from

$$\int_0^x \lambda e^{-\lambda u} du = \lambda \int_0^x e^{-\lambda u} du = \lambda \int_0^{-\lambda x} e^t \frac{dt}{-\lambda} = \int_{-\lambda x}^0 e^t dt = e^t \Big|_{-\lambda x}^0 = 1 - e^{-\lambda x} \quad (6)$$

where change of variable $t = -\lambda u \implies dt = -\lambda du$ is adopted in second equality.

- It follows that the CDF of an exponential (λ) r.v. is

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0, & \text{Otherwise.} \end{cases}$$

(Plot graph of CDF...)

- The mean of exponential r.v. is $E[X] = 1/\lambda$ evaluated as:

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \lambda \int_0^{\infty} x \cdot e^{-\lambda x} dx = \lambda \cdot \frac{-x}{\lambda} e^{-\lambda x} \Big|_0^{\infty} - \lambda \int_0^{\infty} \frac{-1}{\lambda} e^{-\lambda x} dx \\ &= -xe^{-\lambda x} \Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x} dx = (0 - 0) + \frac{1}{\lambda} = \frac{1}{\lambda} \end{aligned} \quad (7)$$

where integration by parts uses variables $u = x, dv = e^{-\lambda x} dx \implies du = dx,\ v = -\frac{1}{\lambda} e^{-\lambda x}$.

Families of Continuous R.V.: Erlang PDF

- More on expontential PDF:
  - Use integration by parts to show that $E[X^2] = 2/\lambda^2$ for an exponential (λ) r.v. X.
  - The variance of an exponential (λ) r.v. is therefore

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

Erlang Random Variable: models the time of the k-th event in Poisson process

- X is an Erlang (r, λ) r.v. if for parameters $\lambda > 0$ and integer $r \geq 1$ it has PDF

$$f_X(x) = \begin{cases} \dfrac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}, & x \geq 0, \\ 0, & \text{Otherwise.} \end{cases} \tag{8}$$

  Parameter r is often called the order of the Erlang process.
- The Erlang PDF generalizes the exponential PDF (5) and is the same when $r = 1$.
- **A Useful Integral Identity:** Note that because (8) is a PDF, the following identity holds:

$$1 = \int_{-\infty}^{\infty} f_X(x) dx = \int_0^{\infty} \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} dx \Longrightarrow \int_0^{\infty} x^{r-1} e^{-\lambda x} dx = \frac{(r-1)!}{\lambda^r}. \tag{9}$$

Families of Continuous R.V.: Erlang PDF (optional)

- The n-th moment of an Erlang r.v. $X \sim \dfrac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!}$, $x \geq 0$ for $n \geq 1$:

$$E[X^n] = \int_{-\infty}^{\infty} x^n \cdot f_X(x) dx = \int_0^{\infty} x^n \cdot \frac{\lambda^r x^{r-1} e^{-\lambda x}}{(r-1)!} dx = \frac{\lambda^r}{(r-1)!} \int_0^{\infty} x^{(n+r)-1} e^{-\lambda x} dx$$

$$= \frac{\lambda^r}{(r-1)!} \cdot \frac{(n+r-1)!}{\lambda^{n+r}} = \frac{(n+r-1)(n+r-2)(n+r-3)\cdots r(r-1)\cdots 2 \cdot 1}{(r-1)(r-2)\cdots 2 \cdot 1} \cdot \frac{1}{\lambda^n}$$

$$= \frac{(n+r-1)(n+r-2)(n+r-3)\cdots(r+1)r}{\lambda^n}.$$

- Thus, the first and second moments are given respectively by

$$E[X] = \frac{r}{\lambda}, \qquad E[X^2] = \frac{(r+1)r}{\lambda^2}. \tag{10}$$

- The variance of an r-order Erlang r.v. is therefore

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \frac{(r+1)r}{\lambda^2} - \frac{r^2}{\lambda^2} = \frac{r^2 + r - r^2}{\lambda^2} = \frac{r}{\lambda^2}.$$

Quiz 4.3

Random variable X has probability density function

$$f_X(x) = \begin{cases} cxe^{-x/2} & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Sketch the PDF and find the following:

(a) the constant c

(b) the CDF $F_X(x)$

(c) $\text{P}[0 \leq X \leq 4]$

(d) $\text{P}[-2 \leq X \leq 2]$

Quiz 4.4

The probability density function of the random variable Y is

$$f_Y(y) = \begin{cases} 3y^2/2 & -1 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Sketch the PDF and find the following:

(a) the expected value $E[Y]$

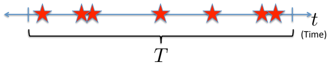
(b) the second moment $E[Y^2]$

(c) the variance $\text{Var}[Y]$

(d) the standard deviation σ_Y

Relationship Between Erlang PDF and Poisson PMF (optional)

- For a Poisson process with λ arrivals per unit time, the random $\#$ of arrivals in duration T is characterized by a Poisson (α) r.v. $K \sim P_K(k) = \frac{\alpha^k e^{-\alpha}}{k!}$, for $k = 0, 1, 2, \dots$ where $\alpha = \lambda T$:



- Let the arrival time of the r-th arrival be $\mathcal{T}_r$. Then,

$$\Pr(\mathcal{T}_r > t) = \Pr(r\text{-th arrival time exceeds } t) = \Pr\begin{pmatrix} \text{no more than } r-1 \text{ arrivals} \\ \text{occur in duration } t \end{pmatrix}$$

$$= \Pr(K \le r-1 \text{ for duration } T = t) = \sum_{k=0}^{r-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \Longrightarrow$$

$$F_{\mathcal{T}_r}(t) = \Pr(\mathcal{T}_r \le t) = 1 - \Pr(\mathcal{T}_r > t) = 1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!} = 1 - \frac{\Gamma(r, \lambda t)}{(r-1)!}$$

$$= 1 - \int_{\lambda t}^{\infty} \frac{u^{r-1} e^{-u}}{(r-1)!} du = 1 - \int_t^{\infty} \frac{(\lambda v)^{r-1} e^{-\lambda v}}{(r-1)!} \lambda dv = \int_0^t \frac{(\lambda v)^{r-1} e^{-\lambda v}}{(r-1)!} \lambda dv$$

$$= \int_{-\infty}^t U(v) \cdot \frac{(\lambda v)^{r-1} e^{-\lambda v}}{(r-1)!} \lambda dv = \int_{-\infty}^t f_{\mathcal{T}_r}(v) dv$$

Relationship Between Erlang PDF and Poisson PMF (optional)

- Regarding the previous calculation[2], the fourth to last equality follows from the change of variables $v = u/\lambda$; the third to last equality follows since $1 - \Pr(A) = \Pr(A^c)$ for any event A; the second to last equality introduces the unit step function[3]; the last equality simply recognizes that the integrand for an integral in this form must be the PDF by the fundamental theorem of calculus.
- Thus, the PDF of the r-th arrival time $\mathcal{T}_r$ is obtained via differentiation:

$$f_{\mathcal{T}_r}(t) = \frac{d}{dt} F_{\mathcal{T}_r}(t) = \begin{cases} \dfrac{(\lambda t)^{r-1} e^{-\lambda t}}{(r-1)!} \lambda, & t \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

which we recognize as the Erlang PDF (8).

[2]The complete Gamma function is defined as the integral $\Gamma(a,x) = \int_x^\infty u^{a-1} e^{-u} du$. When $a = m$, i.e. a is an integer, then $\Gamma(m,x) = (m-1)! e^{-x} \sum_{k=0}^{m-1} \frac{x^k}{k!}$.

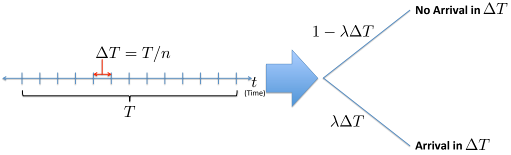
[3]The unit step function is $U(x) = 1$ for all $x \geq 0$, and $U(x) = 0$ for $x < 0$.

Relationship Between Erlang PDF and Poisson PMF (optional)

- Thus, we have the interesting result that the arrival time for the r-th arrival of a Poisson process has an Erlang PDF.
  - Note that the time elapsed before the first arrival, i.e. for the $r = 1$ case, is a first order Erlang or equivalently an exponential PDF: $\mathcal{T}_1 \sim \lambda e^{-\lambda t}$, $t \geq 0$.
  - Also, because adjacent time intervals are independent, the time between arrivals (called interarrival times) is also given by an exponential PDF.
  - It is noteworthy that the r-th arrival time $\mathcal{T}_r$ can be interpreted as the sum of r independent identically distributed interarrival times, i.e. $\mathcal{T}_r = X_1 + X_2 + \cdots + X_r$ where $X_i \sim$ exponential (λ), $i = 1, 2, \ldots, r$ and statistically independent.

Relationship Between Erlang PDF and Poisson PMF

- There's an alternative approach to establish this relationship between the Poisson PMF and the Erlang PDF.
  - Recall that a Poisson process can be decomposed into a series of independent Bernoulli trials by dividing the time interval duration T into many small time slots of duration ΔT:



  - The Poisson process is then determined by the number of successes in n independent Bernoulli trials, i.e. a Binomial process.
  - We showed that in the limit of $\Delta T \to 0$ that this Binomial distribution approaches the Poisson PMF.

Relationship Between Erlang PDF and Poisson PMF

A simple argument for the Erlang PDF:

- Consider the probability of the r-th arrival time, i.e. the duration up to and including the r-th arrival. By independence of nonoverlapping time intervals, it follows that for $t \geq 0$:

$$\begin{aligned}\Pr(t \leq \mathcal{T}_r \leq t + \Delta T) &= \Pr(K = r-1 \text{ in duration } t) \cdot \Pr(\text{one arrival in } \Delta T) \\ &= \Pr(Poi(\lambda t) = r-1) \cdot \lambda \Delta T = \frac{(\lambda t)^{r-1} e^{-\lambda t}}{(r-1)!} \cdot \lambda \Delta T\end{aligned}$$

$$\text{but as } \Delta T \to 0 \text{ recall from (2)} \Longrightarrow \Pr(t \leq \mathcal{T}_r \leq t + \Delta T) \approx f_{\mathcal{T}_r}(t) \cdot \Delta T$$

Thus, it follows that

$$f_{\mathcal{T}_r}(t) = \frac{(\lambda t)^{r-1} e^{-\lambda t}}{(r-1)!} \cdot \lambda, \quad \text{for } t \geq 0$$

which is again the Erlang PDF.

Moment Generating Functions (MGF)[4]

- Transforms such as the Laplace transform and Fourier transform play important roles in many areas of science, engineering, and mathematics. Probability theory is no exception.
- Transforms can be advantageous to establish important theorems, computing moments, and analyses of sums of random variables. We introduce the concept now.
- The moment generating function (MGF) of a r.v. X is defined as

$$\phi_X(s) = E[e^{sX}]$$

  for continuous and discrete random variables, differing only in formula for expectation.
- For continuous r.v. $X \sim f_X(x)$ the MGF is

$$\phi_X(s) = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} \cdot f_X(x)\, dx. \tag{12}$$

- The integral defining the MGF in (12) is simply the Laplace transform of the PDF. The values of s for which the integral converges ($\phi_X(s)$ exists) is the region of convergence.

[4]See Chapter 9.2 in Yates/Goodman [1]

Moment Generating Functions (MGF)

- For discrete r.v. $X \sim P_X(x)$ the MGF is

$$\phi_X(s) = E[e^{sX}] = \sum_{x \in S_X} e^{sx} \cdot P_X(x).$$

- Note from the definition of MGF that $\phi_X(0) = 1$:

$$\text{continuous r.v.:} \;\; \phi_X(0) = E[e^{0 \cdot X}] = \int_{-\infty}^{\infty} e^{0 \cdot x} \cdot f_X(x) dx = \int_{-\infty}^{\infty} 1 \cdot f_X(x) dx = 1.$$

$$\text{discrete r.v.:} \;\; \phi_X(0) = E[e^{0 \cdot X}] = \sum_{x \in S_X} e^{0 \cdot x} \cdot P_X(x) = \sum_{x \in S_X} 1 \cdot P_X(x) = 1.$$

  Thus, $s = 0$ is always in the region of convergence.

- Because of the uniqueness of Laplace transforms, knowledge of MGF $\phi_X(s)$ completely characterizes the PDF $f_X(x)$ (or PMF $P_X(x)$ if discrete). Thus, in theory, the MGF is a complete description of a r.v.

MGF and n-th Moments

- Random variable $X \sim f_X(x)$ with MGF $\phi_X(s)$ has n-th moment:

$$E[X^n] = \left. \frac{d^n \phi_X(s)}{ds^n} \right|_{s=0}.$$

$$\frac{d\phi_X(s)}{ds} = \frac{d}{ds} \int_{-\infty}^{\infty} e^{sx} \cdot f_X(x)\, dx = \int_{-\infty}^{\infty} \frac{d}{ds} e^{sx} \cdot f_X(x)\, dx = \int_{-\infty}^{\infty} x \cdot e^{sx} \cdot f_X(x)\, dx \Longrightarrow$$

$$\left. \frac{d\phi_X(s)}{ds} \right|_{s=0} = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx = E[X].$$

$$\frac{d^2\phi_X(s)}{ds^2} = \frac{d^2}{ds^2} \int_{-\infty}^{\infty} e^{sx} \cdot f_X(x)\, dx = \int_{-\infty}^{\infty} \frac{d^2}{ds^2} e^{sx} \cdot f_X(x)\, dx = \int_{-\infty}^{\infty} x^2 \cdot e^{sx} \cdot f_X(x)\, dx \Longrightarrow$$

$$\left. \frac{d^2\phi_X(s)}{ds^2} \right|_{s=0} = \int_{-\infty}^{\infty} x^2 \cdot f_X(x)\, dx = E[X^2],$$

$$\frac{d^n\phi_X(s)}{ds^n} = \frac{d^n}{ds^n} \int_{-\infty}^{\infty} e^{sx} \cdot f_X(x)\, dx = \int_{-\infty}^{\infty} \frac{d^n}{ds^n} e^{sx} \cdot f_X(x)\, dx = \int_{-\infty}^{\infty} x^n \cdot e^{sx} \cdot f_X(x)\, dx \Longrightarrow$$

$$\left. \frac{d^n\phi_X(s)}{ds^n} \right|_{s=0} = \int_{-\infty}^{\infty} x^n \cdot f_X(x)\, dx = E[X^n].$$

MGF and n-th Moments

- Regarding a discrete r.v. $X \sim P_X(x)$, one obtains

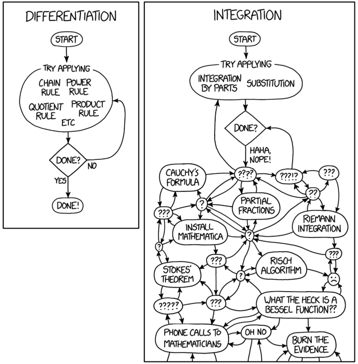
$$\frac{d^n \phi_X(s)}{ds^n} = \frac{d^n}{ds^n} \sum_{x \in S_X} e^{sx} \cdot P_X(x) = \sum_{x \in S_X} \frac{d^n}{ds^n} e^{sx} \cdot P_X(x) = \sum_{x \in S_X} x^n \cdot e^{sx} \cdot P_X(x) \Longrightarrow$$

$$\left. \frac{d^n \phi_X(s)}{ds^n} \right|_{s=0} = \sum_{x \in S_X} x^n \cdot P_X(x) = E[X^n].$$

Hence, named "moment" generating function.

- Evaluation of moments for a r.v. in several cases can be significantly simpler to do with the MGF than directly with the PDF because differentiation is often easier to do than integration.

A bit of humor from author Randall Munroe (`https://xkcd.com/2117/`) $\Longrightarrow$



MGFs: Example

- **Example:** Consider $X \sim f_X(x) = \lambda e^{-\lambda x}$, $x \geq 0$, i.e. an exponential (λ) r.v. The MGF is evaluated as

$$\phi_X(s) = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} \cdot f_X(x) dx = \int_0^{\infty} e^{sx} \cdot \lambda e^{-\lambda x}\, dx = \lambda \int_0^{\infty} e^{(s-\lambda)x}\, dx$$
$$= \lambda \int_0^{-\infty} e^u \frac{du}{s-\lambda} = \frac{\lambda}{s-\lambda} \cdot \left(e^u \big|_0^{-\infty} \right) = \frac{\lambda}{s-\lambda} \cdot (0-1) = \frac{\lambda}{\lambda - s}$$

  where variable change is $u = (s-\lambda)x$, $du = (s-\lambda)\, dx$ for $\mathrm{Re}(s-\lambda) < 0$.
  - As a test, note that $\phi_X(0) = \lambda/(\lambda - 0) = 1$.
  - The first moment can be obtained via
$$\frac{d\phi_X(s)}{ds} = \frac{\lambda(-1)}{(\lambda-s)^2} \cdot (-1) = \frac{\lambda}{(\lambda-s)^2} \implies \left.\frac{d\phi_X(s)}{ds}\right|_{s=0} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda} = E[X].$$
  - The second moment follows from
$$\frac{d^2\phi_X(s)}{ds^2} = \frac{\lambda(-2)}{(\lambda-s)^3} \cdot (-1) = \frac{2\lambda}{(\lambda-s)^3} \implies \left.\frac{d\phi_X(s)}{ds}\right|_{s=0} = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2} = E[X^2].$$

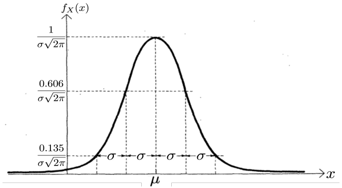
- A table of various MGFs is provided in Table 9.1 of Yates/Goodman [1] p. 312.

Gaussian Random Variables

- The Gaussian distribution appears quite frequently in many applications. Thus, it is a probability model with which it is worth becoming very familiar.
- X is a Gaussian (μ, σ) r.v. if it has PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad \text{for } -\infty \leq x \leq \infty \tag{13}$$

where parameters $\mu \in \mathbb{R}$ and $\sigma > 0$.
  - A Gaussian r.v. is also sometimes referred to as a normal r.v.
  - Shorthand notation $X \sim N(\mu, \sigma^2)$ is often used to indicate a Gaussian (μ, σ) r.v.
- Gaussian PDF illustrated below; the name "bell curve" has clear origin:



MGF of Gaussian R.V.s

- Since (13) is a PDF, it must hold that

$$1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx \Longrightarrow \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx = \sqrt{2\pi\sigma^2}. \qquad (14)$$

- The MGF for a Gaussian r.v. with PDF (13) is $\phi_X(s) = e^{s\mu + \frac{s^2\sigma^2}{2}}$:

$$\phi_X(s) = E[e^{sX}] = \int_{-\infty}^{\infty} e^{sx} \cdot f_X(x)dx = \int_{-\infty}^{\infty} e^{sx} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\, dx$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2x\mu+\mu^2+2\sigma^2 sx}{2\sigma^2}}\, dx = \frac{e^{\frac{\mu^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{x^2-2x(\mu+\sigma^2 s)}{2\sigma^2}}\, dx.$$

Completing the square in x we obtain

$$\phi_X(s) = \frac{e^{\frac{\mu^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{[x-(\mu+\sigma^2 s)]^2-(\mu+\sigma^2 s)^2}{2\sigma^2}}\, dx = \frac{e^{\frac{\mu^2-(\mu+\sigma^2 s)^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{[x-(\mu+\sigma^2 s)]^2}{2\sigma^2}}\, dx$$

$$= \frac{e^{-\frac{\mu^2-(\mu^2+2\mu\sigma^2 s+\sigma^4 s^2)}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \cdot \sqrt{2\pi\sigma^2} = e^{s\mu+\frac{s^2\sigma^2}{2}} \qquad \text{where integral by (14)}$$

Mean and Variance of Gaussian R.V.s

- The first moment $E[X]$ is easily deduced from the symmetry of the PDF (13), i.e. the center of mass is clearly $E[X] = \mu$. This can be established formally via:

$$\frac{d\phi_X(s)}{ds} = \left(\mu + s\sigma^2\right) e^{s\mu + \frac{s^2\sigma^2}{2}} \Longrightarrow \left.\frac{d\phi_X(s)}{ds}\right|_{s=0} = \mu = E[X].$$

- The second moment $E[X^2]$ follows from

$$\frac{d^2\phi_X(s)}{ds^2} = \left(\mu + s\sigma^2\right)^2 e^{s\mu + \frac{s^2\sigma^2}{2}} + \sigma^2 \cdot e^{s\mu + \frac{s^2\sigma^2}{2}} \Longrightarrow \left.\frac{d^2\phi_X(s)}{ds^2}\right|_{s=0} = \mu^2 + \sigma^2 = E[X^2].$$

  Thus, the variance of a Gaussian r.v. (13) is

$$\mathrm{Var}[X] = E[X^2] - E[X]^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2.$$

- The parameters of Gaussian PDF (μ, σ^2) are therefore respectively the mean and variance:

$$X \sim N(\mu, \sigma^2) \Longrightarrow E[X] = \mu, \text{ and } \mathrm{Var}[X] = \sigma^2.$$

Gaussians and Linear Transformations

- **Important property:** Gaussian r.v.'s regenerate under linear transformations.
  - If $X \sim N(\mu, \sigma^2)$ then $Y = aX + b \sim N(a\mu + b, a^2\sigma^2)$:

$$E[e^{sY}] = E[e^{s(aX+b)}] = e^{sb} \cdot E[e^{saX}] = e^{sb} \cdot \left. E[e^{\tilde{s}X}] \right|_{\tilde{s}=sa} = e^{sb} \cdot \left. e^{\tilde{s}\mu + \frac{\tilde{s}^2\sigma^2}{2}} \right|_{\tilde{s}=sa}$$

$$= e^{sb} \cdot e^{sa\mu + \frac{s^2a^2\sigma^2}{2}} = e^{s\mu_Y + \frac{s^2\sigma_Y^2}{2}}, \text{ where } \mu_Y = a\mu + b, \ \ \sigma_Y^2 = a^2\sigma^2.$$

  - Thus, by uniqueness of the Laplace transform, it follows that $Y \sim N(a\mu + b, a^2\sigma^2)$.

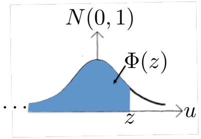
- Random variable $Z \sim N(0, 1)$ is said to be a standardized normal random variable. It has the property that $E[Z] = 0$ and $\mathrm{Var}[Z] = 1$.
  - Convince yourself that if $X \sim N(\mu, \sigma^2)$, then $Z = (X - \mu)/\sigma \sim N(0, 1)$.

- The CDF for a standardized normal random variable is a well-tabulated function (see Table 4.2, p. 143 in [1]. MATLAB also have special functions available for its calculation (see functions erf and erfc)).

CDF of Standardized Gaussian R.V.

- The r.v. $Z \sim N(0,1)$ has CDF given by

$$\Pr(Z \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-u^2}{2}} du \triangleq \Phi(z)$$



- The r.v. $Z \sim N(0,1)$ has complimentary CDF given by

$$\Pr(Z > z) = \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} e^{\frac{-u^2}{2}} du = 1 - \Phi(z) \triangleq Q(z).$$

where notation $Q(\cdot)$ is adopted to indicate this function.

- Consider r.v. $X \sim N(\mu, \sigma^2)$ and note that it has CDF

$$\begin{aligned} F_X(x) = \Pr(X \le x) &= \Pr(X - \mu \le x - \mu) = \Pr[(X-\mu)/\sigma \le (x-\mu)/\sigma] \\ &= \Pr[Z \le (x-\mu)/\sigma] = \Phi\left(\frac{x-\mu}{\sigma}\right). \end{aligned}$$

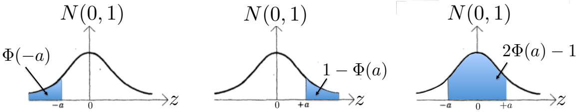
$$\Pr(a < X \le b) = \Pr[(a-\mu)/\sigma < Z \le (b-\mu)/\sigma] = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

Using CDF of Standardized Gaussian R.V.

- Consider that for $X \sim N(\mu, \sigma^2)$ we have

$$\Pr[|X-\mu| \le a\sigma] = \Pr[-a\sigma \le X-\mu \le a\sigma] = \Pr\left[-a \le \frac{X-\mu}{\sigma} \le a\right] = \Pr[-a \le Z \le a]$$
$$= \Phi(a) - \Phi(-a) = \Phi(a) - [1-\Phi(a)] = 2\Phi(a) - 1$$

where integrals and symmetry $\Phi(-a) = 1 - \Phi(a)$ are illustrated in Figure:



- As an example, note that for $X \sim N(\mu, \sigma^2)$:

$$\Pr[-a\sigma \le X-\mu \le a\sigma] = 2\Phi(a) - 1 = \begin{cases} 0.682, & a = 1 \\ 0.954, & a = 2 \\ 0.997, & a = 3. \end{cases}$$

  - Thus, there's 68% chance (i.e. $\sim \frac{2}{3} = 0.67$ probability) r.v. X belongs to $[\mu-\sigma, \mu+\sigma]$; Almost 70% of the time r.v. X will be within a standard deviation of its mean value.
  - There's a 95% chance r.v. X belongs to $[\mu-2\sigma, \mu+2\sigma]$, i.e. it will be within two standard deviations of its mean value with 95% certainty.

Exceeding Mean μ by Multiples of σ

- Note probability of X exceeding its mean by multiples of the standard deviation:

$$\Pr\left[X - \mu > a\sigma\right] = 1 - \Phi(a) = \frac{1}{2}\{1 - [2\Phi(a) - 1]\} = \begin{cases} 0.1587, & a = 1 \\ 0.0228, & a = 2 \\ 0.0013, & a = 3. \end{cases}$$

- Thus, exceedance of mean μ by σ is about $\sim 1/6$;
- exceedance of mean μ by 2σ is about $\sim 1/50$;
- exceedance of mean μ by 3σ is about $\sim 1/1000$;
- Last two events are quite rare.

Quiz 4.6

X is the Gaussian $(0, 1)$ random variable and Y is the Gaussian $(0, 2)$ random variable. Sketch the PDFs $f_X(x)$ and $f_Y(y)$ on the same axes and find:

(a) $\mathrm{P}[-1 < X \leq 1]$,

(b) $\mathrm{P}[-1 < Y \leq 1]$,

(c) $\mathrm{P}[X > 3.5]$,

(d) $\mathrm{P}[Y > 3.5]$.

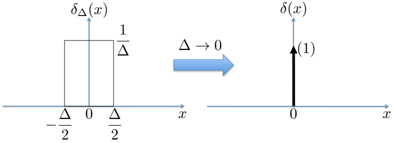
Delta Functions, and Mixed Random Variables

- So far we've considered only continuous random variables with ranges such that $S_X \subseteq \mathbb{R}$, and having CDF functions $F_X(x)$ that are continuous.
  - Recall that for such r.v.'s $\Pr(X = x_1) = \int_{x_1}^{x_1} f_X(u)du = 0$.
- Our discussion of discrete r.v.'s on the otherhand focused on r.v.'s with ranges $S_X = \{s_1, s_2, s_3, \ldots\} \subseteq \mathbb{R}$, i.e. consisting of a countable number of elements.
- It is desired to consider a mixed / hybrid r.v. that is like both a continuous r.v. and a discrete r.v., such that for a point $x_1 \in S_X$ it is possible to have nonzero $\Pr(X = x_1) = \int_{x_1}^{x_1} f_X(u)du > 0$ for X defined on an uncountable range $S_X \subseteq \mathbb{R}$.
- Such a mixed r.v. is made possible by consideration of the Dirac delta unit impulse function denoted as $\delta(x)$.
- The Dirac delta function $\delta(x)$ is defined by how it behaves under the integral sign, rather than the specific values it assume for $x \in \mathbb{R}$.
  - Such functions are called generalized function or distribution.

Delta Functions, and Mixed Random Variables

- Consider the function $\delta_\Delta(x)$ defined as

$$\delta_\Delta(x) = \begin{cases} \frac{1}{\Delta}, & -\frac{\Delta}{2} \leq x \leq \frac{\Delta}{2} \\ 0, & \text{Otherwise} \end{cases} \tag{15}$$



where clearly $\delta_\Delta(x)$ has unit area for all $\Delta > 0$, i.e.

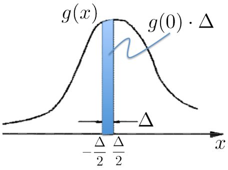
$$\int_{-\infty}^{\infty} \delta_\Delta(x)\, dx = \frac{1}{\Delta}\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} dx = \frac{1}{\Delta} x \Big|_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} = \frac{1}{\Delta}\left(\frac{\Delta}{2} - \frac{-\Delta}{2}\right) = \frac{\Delta}{\Delta} = 1.$$

- As $\Delta \to 0$, width becomes arbitrarily small while area under the pulse is fiexd to 1.
  - The length of interval $[\frac{-\Delta}{2}, \frac{\Delta}{2}]$, i.e. Δ, approaches zero.
- This assigns a nonzero probability (area) to a single point, i.e. a width zero interval.
- The limiting form of $\delta_\Delta(x)$ as $\Delta \to 0$ is denoted $\delta(x)$, and is graphically denoted as in above Figure where the number in parentheses $(\cdot)$ indicates the area, i.e.

$$\int_{-\infty}^{\infty} \delta(x)\, dx = 1.$$

Sifting Property of Dirac Delta Function

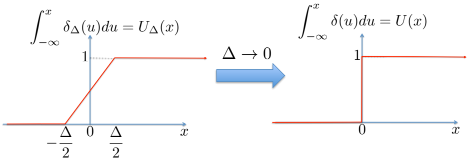
- Consider a function $g(x)$ and the area under $g(x) \cdot \delta_\Delta(x)$:
$$\int_{-\infty}^{\infty} g(x)\delta_\Delta(x)\,dx = \frac{1}{\Delta}\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} g(x)\,dx \approx \frac{1}{\Delta}\cdot g(0)\cdot\Delta = g(0)$$
where approximation is similar to that in (2) and exact as $\Delta \to 0$.



- Similarly, for function $g(x)$ the area under $g(x) \cdot \delta_\Delta(x - x_0)$:
$$\int_{-\infty}^{\infty} g(x)\delta_\Delta(x-x_0)\,dx = \frac{1}{\Delta}\int_{x_0-\frac{\Delta}{2}}^{x_0+\frac{\Delta}{2}} g(x)\,dx \approx \frac{1}{\Delta}\cdot g(x_0)\cdot\Delta = g(x_0)$$
where approximation improves as $\Delta \to 0$ for same reasons aforementioned.
- Thus, we have the sifting property of the Dirac delta function:
$$\int_{-\infty}^{\infty} g(x)\delta(x-x_0)\,dx = g(x_0), \tag{16}$$
  - i.e. integral of function times a delayed/shifted unit impulse equals function evaluated at delay/shift

Dirac Delta Relation to Unit Step Function

- From equation (15) and the figure below it, note that the cumulative integral of function $\delta_\Delta(x)$ is denoted $U_\Delta(x)$ and illustrated below:



- Thus, we have the following property that the cumulative integral of the unit impulse function is the unit step function:

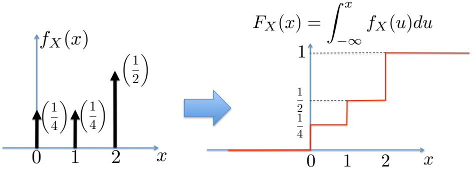
$$\int_{-\infty}^{x} \delta(u)du = U(x) = \begin{cases} 1, & x \geq 0 \\ 0, & \text{Otherwise} \end{cases} \tag{17}$$

Extending the fundamental theorem of calculus to "distributions", one writes

$$\delta(x) = \frac{d}{dx} U(x).$$

Example: Using PDF to Represent a Discrete R.V.

- Consider the PDF $f_X(x) = \frac{1}{4}\delta(x) + \frac{1}{4}\delta(x-1) + \frac{1}{2}\delta(x-2)$, illustrated below



- Using (17) the CDF is shown to be

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du = \int_{-\infty}^{x} \frac{1}{4}\delta(u) + \frac{1}{4}\delta(u-1) + \frac{1}{2}\delta(u-2)du$$
$$= \frac{1}{4}U(x) + \frac{1}{4}U(x-1) + \frac{1}{2}U(x-2) \ \text{(illustrated above)}$$

  - Clearly, unit impulse function provides a way of representing discrete events with nonzero probability within a continuous domain of infinite possibilities.
  - This example is essentially a discrete r.v. because the range of X is $S_X = \{0, 1, 2\}$ that is clearly finite in size and therefore countable.
  - This example shows how a discrete r.v. can be represented with a PDF (instead of a PMF).

Example: Using PDF to Represent a Discrete R.V.

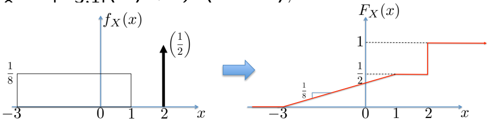
- The expected value of X is given by

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x \cdot f_X(x) dx = \int_{-\infty}^{\infty} x \cdot \left[\frac{1}{4}\delta(x) + \frac{1}{4}\delta(x-1) + \frac{1}{2}\delta(x-2)\right] dx \\
&= \frac{1}{4}\int_{-\infty}^{\infty} x \cdot \delta(x)\, dx + \frac{1}{4}\int_{-\infty}^{\infty} x \cdot \delta(x-1)\, dx + \frac{1}{2}\int_{-\infty}^{\infty} x \cdot \delta(x-2)\, dx \\
&= \frac{1}{4} \cdot 0 + \frac{1}{4} \cdot 1 + \frac{1}{2} \cdot 2 = \frac{5}{4}
\end{aligned}
$$

where that last equality follows from (16) the sifting property of the unit impulse.

Example: Mixed Random Variable

- Consider PDF $f_X(x) = \frac{1}{8} \cdot \mathbb{1}_{[-3,1]}(x) + \frac{1}{2}\delta(x-2)$, illustrated below



- This represents a mixed r.v. with continuous and discrete random components.
- Convince yourself that the CDF illustrated above and specified below is correct:

$$F_X(x) = \begin{cases} 0, & x \leq -3 \\ \frac{1}{8}(x+3), & -3 < x \leq 1 \\ \frac{1}{2}, & 1 < x < 2, \\ 1, & x \geq 2. \end{cases}$$

- CDF $F_X(x)$ is discontinuous in this case, i.e. at $x = 2$.
- At points of discontinuity x_D we can define the PDF value as $\Longrightarrow$

$$f_X(x_D) = \left.\frac{dF_X(x)}{dx}\right|_{x=x_D} = (\text{jump size}) \cdot \delta(x - x_D) + \text{any continuous part} \tag{18}$$

- Note that $\int_2^2 f_X(x)dx = \frac{1}{2} > 0$ in this case.

References

[1] Yates and Goodman, *Probability and Stochastic Processes*, 3^{rd} Ed., Wiley, 2014.

[2] D. P. Bertsekas and J. N. Tsitsiklis, *Introduction to Probability*, Athena Sci., 2002.

[3] A. W. Drake, *Fundamentals of Applied Probability Theory*, McGraw-Hill Inc., 1967.