Continuous Random Variables

Chia-Ping Chen

Professor
Department of Computer Science and Engineering
National Sun Yat-sen University

Probability

Outline

- Continuous Random Variables
- Probability Density Function
- Expectation and Variance
- Cumulative Distribution Function
- Uniform, Exponential, Normal
- Conditional Models
- Total Probability and Total Expectation
- Bayes rule

Random variable types

Let $(\Omega, \mathcal{F}, P)$ be a probability model. A random variable defined on Ω can be either discrete, continuous, or mixed.

- A discrete random variable has a discrete range, e.g.

$$\{0, 1, 2, \dots\}$$

- A continuous random variable has a continuous range, e.g.

$$(0,3),\ [0,3],\ [0,3),\ (0,3]$$

- A mixed random variable has a range which is the union of a discrete set and a continuous set, e.g.

$$\{0\} \cup [5, 10)$$

Example (별에서 온 그대 and 사랑의 불시착)

Continuous random variables (CRVs) arise naturally in scenarios where the quantities take continuous values, e.g. time and space.

- Arrival times of meteorites
- Landing spots of paragliding

Definition (probability density function)

Let $(\Omega, \mathcal{F}, P)$ be a probability model and X be a CRV defined on Ω with range $\mathcal{X}$.

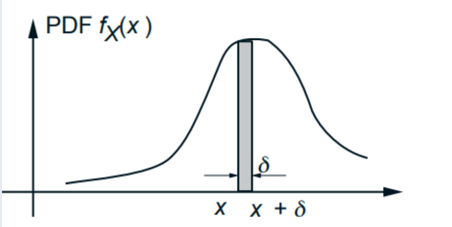
- Probability law over Ω is converted to distribution of X over $\mathcal{X}$, which is specified by a probability density function (PDF)
- An interval $(x, x+\delta)$ corresponds to an event $(X \in (x, x+\delta))$. Furthermore, the probability is given by

$$P(X \in (x, x+\delta)) = \underbrace{f_X(x)}_{\text{PDF of } X} \delta + o(\delta)$$

Small-interval probability

Let X be a CRV with PDF f_X.

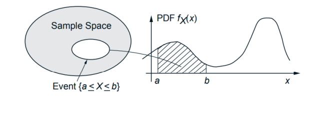
- $P(X \in (x, x+\delta)) \approx f_X(x)\delta$: the probability of a small interval is proportional to its length and the proportional constant is $f_X(x)$
- Thus $f_X(x)$ is the probability per unit length (density) at x
- Note $f_X(x)\delta$ is the area of the grey bar



Finite-interval probability

Let X be a CRV. For any interval

$$P(X \in (a,b)) = \int_a^b f_X(x)dx$$



Partition the interval

$$(X \in (a,b)) = \bigcup_{i=0}^{n-1} (X \in (x_i, x_{i+1})),\ x_i = a + i\delta,\ \delta = \left(\frac{b-a}{n}\right)$$

Probability of small intervals

$$P(X \in (a,b)) = \sum_{i=0}^{n-1} P(X \in (x_i, x_{i+1})) = \sum_{i=0}^{n-1} f_X(x_i)\delta + o(\delta)$$

As $\delta \to 0$, the infinite sum is integration

$$P(X \in (a,b)) = \int_a^b f_X(x)dx$$

PDF properties

Let X be a CRV with PDF f_X.

- f_X is non-negative

$$f_X(x) \geq 0$$

- f_X is normalized

$$\int_{-\infty}^{\infty} f_X(x)dx = 1$$

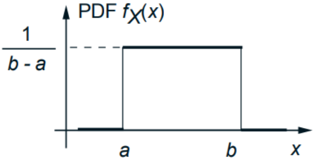
Uniform

A uniform CRV with range (a, b) is denoted by

$$X \sim \textbf{Uni}(a, b)$$

The PDF of X is

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$



Example (3.1 Uniform)

A **wheel of fortune** is uniform and continuously calibrated between 0 and 1. What is the PDF for the outcome of a spin?

Let X be the outcome of a spin. We have $X \sim \textbf{Uni}(0, 1)$. So the PDF of X is

$$f_X(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

Definition (step function)

The step function is defined by

$$u(x) = \begin{cases} 0, & x \le 0 \\ 1, & x > 0 \end{cases}$$

Using step function, the PDF of $X \sim$ **Uni**$(0, 1)$ is (except for one point, which is OK for finite PDF)

$$f_X(x) = u(x - 0) - u(x - 1)$$

More generally, the PDF of $X \sim$ **Uni**(a, b) is

$$f_X(x) = \frac{1}{b - a}(u(x - a) - u(x - b))$$

Example (3.2 Piece-wise uniform)

Alvin's driving time to work is uniform in 15–20 minutes (resp. 20–25) in a sunny day (resp. rainy day). Assume that a day is sunny with probability $\frac{2}{3}$, and rainy with probability $\frac{1}{3}$. What is the PDF of the driving time X?

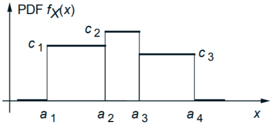
The PDF of X are constants in the intervals $(15, 20)$ and $(20, 25)$

$$f_X(x) = c_1(u(x-15) - u(x-20)) + c_2(u(x-20) - u(x-25))$$

To decide c_1 and c_2, we note $P(\text{sunny}) = P(X \in (15, 20)) = \frac{2}{3}$ and $P(\text{rainy}) = P(X \in (20, 25)) = \frac{1}{3}$. Thus

$$\frac{2}{3} = \int_{15}^{20} c_1\, dx \;\Rightarrow\; c_1 = \frac{2}{15}$$

$$\frac{1}{3} = \int_{20}^{25} c_2\, dx \;\Rightarrow\; c_2 = \frac{1}{15}$$

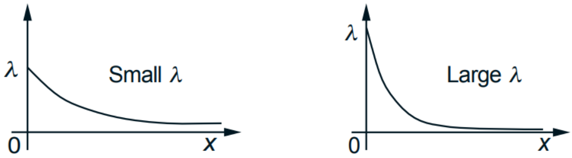


Exponential

An exponential random variable with parameter λ has PDF

$$f_X(x) = \lambda e^{-\lambda x} u(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

Examples



It is denoted by

$$X \sim \textbf{Exp}(\lambda)$$

Expectation and Variance

Definition (continuous expectation)

Let X be a CRV with PDF f_X. The expectation of X is

$$\mathbf{E}[X] = \int x\, f_X(x) dx$$

Expectation of a function of CRV

$$\mathbf{E}[g(X)] = \int g(x) f_X(x) dx$$

Definition (variance and moments)

Let X be a CRV with PDF f_X.

- The variance of X is

$$\textbf{var}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

- For $n \in \mathbb{N}$, the nth moment of X is

$$\mathbf{E}[X^n] = \int x^n f_X(x) dx$$

$$\begin{aligned}
\textbf{var}(X) &= \mathbf{E}[(X - \mathbf{E}[X])^2] \\
&= \int (x - \mathbf{E}[X])^2 f_X(x) dx \\
&= \int x^2 f_X(x) dx - 2\mathbf{E}[X] \int x f_X(x) dx + \mathbf{E}^2[X] \\
&= \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X]
\end{aligned}$$

Example (3.4 Uniform expectation and variance)

Consider $X \sim \textbf{Uni}(a, b)$.

$$\mathbf{E}[X] = \int x\, f_X(x) dx$$

$$= \frac{1}{b-a} \int_a^b x\, dx$$

$$= \frac{a+b}{2}$$

$$\textbf{var}(X) = \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X]$$

$$= \frac{1}{b-a} \int_a^b x^2\, dx - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{(b-a)^2}{12}$$

Example (Exponential expectation and variance)

Consider $T \sim \textbf{Exp}(\lambda)$.

$$\begin{aligned}
\mathbf{E}[T] &= \int t\, f_T(t) dt \\
&= \int_0^\infty t \left(\lambda e^{-\lambda t} \right) dt \\
&= \left(-t e^{-\lambda t} \right) \Big|_0^\infty - \int_0^\infty \left(-e^{-\lambda t} \right) dt \\
&= \left(-\frac{1}{\lambda} e^{-\lambda t} \right) \Big|_0^\infty \\
&= \frac{1}{\lambda}
\end{aligned}$$

$$\begin{aligned}
\mathbf{E}\left[T^2\right] &= \int_0^\infty t^2\, \lambda e^{-\lambda t} dt = -t^2 e^{-\lambda t} \Big|_0^\infty - \int_0^\infty \left(-2t e^{-\lambda t} \right) dt \\
&= \frac{2}{\lambda} \int_0^\infty t \left(\lambda e^{-\lambda t} \right) dt = \frac{2}{\lambda^2}
\end{aligned}$$

$$\textbf{var}(T) = \mathbf{E}\left[T^2\right] - \mathbf{E}^2[T] = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda} \right)^2 = \frac{1}{\lambda^2}$$

Example (3.5)

The time until a small meteorite first lands anywhere in the Sahara desert is modeled as $T \sim \textbf{Exp}(\lambda)$ with a mean of 10 days. The time is currently midnight. What is the probability that the first meteorite lands between 6 am and 6 pm of the first day?

We have

$$\mathbf{E}[T] = \frac{1}{\lambda} = 10 \;\Rightarrow\; \lambda = \frac{1}{10}$$

Thus

$$\begin{aligned} P\left(\frac{1}{4} < T < \frac{3}{4}\right) &= \int_{\frac{1}{4}}^{\frac{3}{4}} f_T(t)dt \\ &= \int_{\frac{1}{4}}^{\frac{3}{4}} \lambda e^{-\lambda t} dt \\ &= -e^{-\lambda t}\Big|_{\frac{1}{4}}^{\frac{3}{4}} \\ &= e^{-\frac{1}{40}} - e^{-\frac{3}{40}} \end{aligned}$$

Linear function

Let X be a CRV. We have

$$\mathbf{E}[aX + b] = a\mathbf{E}[X] + b$$

$$\textbf{var}(aX + b) = a^2\textbf{var}(X)$$

They are the same results as the discrete case.

Cumulative Distribution Function

Definition (cumulative distribution function)

Let $(\Omega, \mathcal{F}, P)$ be a probability model and X be a random variable defined on Ω with range $\mathcal{X}$.

- The cumulative distribution function of X is defined by

$$\underbrace{F_X(x)}_{\text{CDF of } X} = P(X \leq x)$$

- $F_X(x)$ fully specifies the distribution of X

CDF and PMF

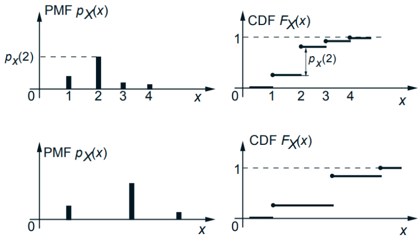
Let X be a DRV with CDF F_X and PMF p_X.

- F_X is the accumulation of p_X

$$F_X(x) = \sum_{x_i \le x} p_X(x_i)$$

- p_X is the difference of F_X

$$p_X(x_k) = F_X(x_k) - F_X(x_{k-1})$$



CDF and PDF

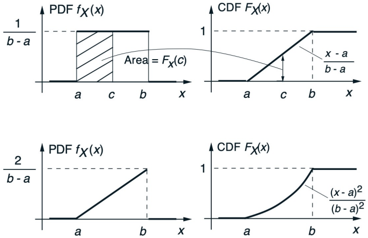
Let X be a CRV with CDF F_X and PDF f_X.

- F_X is the integration of f_X

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

- f_X is the differentiation of F_X

$$f_X(x) = \frac{dF_X(x)}{dx}$$



Example (CDF)

- $X \sim \textbf{Uni}(a,b)$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & x \ge b \end{cases}$$

- $X \sim \textbf{Exp}(\lambda)$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0, & x \le 0 \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}$$

- $X \sim \textbf{Geo}(p)$

$$F_X(x) = \sum_{k=1}^{\lfloor x \rfloor} p_X(k) = \begin{cases} 0, & x < 1 \\ 1 - (1-p)^{\lfloor x \rfloor}, & x \ge 1 \end{cases}$$

CDF properties

For any random variable X, its CDF F_X has the following properties.

- Non-decreasing

$$x_1 \leq x_2 \xrightarrow{(X \leq x_1) \subset (X \leq x_2)} F_X(x_1)\, e F_X(x_2)$$

- Bounded

$$0 \leq F_X(x) \leq 1$$

- Limit values

$$\lim_{x \to -\infty} F_X(x) = P(X \leq -\infty) = 0$$

$$\lim_{x \to \infty} F_X(x) = P(X \leq \infty) = 1$$

Inverse function method

- Let X be CRV with CDF $F_X(x)$. Consider $Y = F_X(X)$. We have $0 \leq Y \leq 1$ and the CDF of Y is

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= P(F_X(X) \leq y) \\ &= P(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)) \\ &= P(X \leq F_X^{-1}(y)) \\ &= F_X(F_X^{-1}(y)) = y \end{aligned}$$

  Therefore, $Y \sim \textbf{Uni}(0,1)$.

- Let $U \sim \textbf{Uni}(0,1)$. Since $F_X(X)$ and U have the same PDF, X and $F_X^{-1}(U)$ have the same PDF. For example, let $X \sim \textbf{Exp}(\lambda)$ with $F_X(x) = 1 - e^{-\lambda x}$. We have

$$F_X^{-1}(U) = -\frac{1}{\lambda}\log(1-U) \sim \textbf{Exp}(\lambda)$$

Example (3.6 Maximum CDF)

Let final score X be the maximum of independent scores X_1, X_2, X_3 with $X_i \sim \textbf{Uni}[1, 10]$. Find the PMF of X.

We derive PMF from CDF. Consider $(X \leq t)$. Since $(X \leq t) = (X_1 \leq t) \cap (X_2 \leq t) \cap (X_3 \leq t)$, we have

$$\begin{aligned} P(X \leq t) &= P((X_1 \leq t) \cap (X_2 \leq t) \cap (X_3 \leq t)) \\ &= P(X_1 \leq t)\, P(X_2 \leq t)\, P(X_3 \leq t) \end{aligned}$$

At $k = 1, \ldots, 10$ the values of the CDF of X are

$$F_X(k) = F_{X_1}(k) F_{X_2}(k) F_{X_3}(k) = \left(\frac{k}{10}\right)^3$$

At $k = 1, \ldots, 10$, the values of the PMF of X are

$$p_X(k) = F_X(k) - F_X(k-1) = \left(\frac{k}{10}\right)^3 - \left(\frac{k-1}{10}\right)^3$$

Geometric and exponential

Consider $N \sim \textbf{Geo}(p)$, $W = N\delta$ and $T \sim \textbf{Exp}(\lambda)$.

- The CDF of N is

$$F_N(t) = 1 - (1-p)^{\lfloor t \rfloor}$$

- The CDF of W is

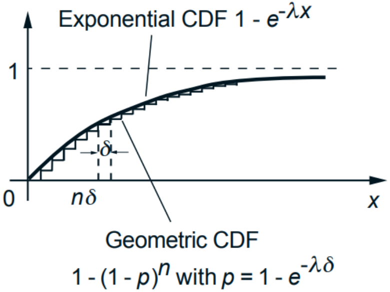
$$F_W(x) = F_N\left(\frac{x}{\delta}\right) = 1 - (1-p)^{\lfloor \frac{x}{\delta} \rfloor}$$

- The CDF of T is

$$F_T(x) = 1 - e^{-\lambda x} = 1 - \left(e^{-\lambda\delta}\right)^{\frac{x}{\delta}}$$

- If λ, p, δ are related by $1 - p = e^{-\lambda\delta}$, then

$$F_T(x) \approx F_W(x)$$



Exponential CDF $1 - e^{-\lambda x}$
1
0
$n\delta$
δ
x
Geometric CDF
$1 - (1 - p)^n$ with $p = 1 - e^{-\lambda\delta}$

Normal Random Variables

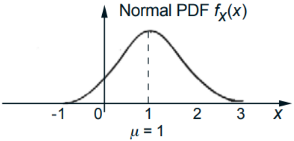
Definition (Normal random variable)

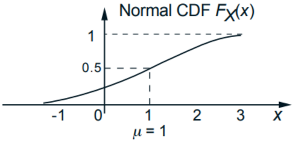
A Normal (a.k.a. Gaussian) random variable with parameters μ and σ^2 has PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

This is denoted by

$$X \sim \mathcal{N}(\mu, \sigma^2)$$





Normal properties

- Consider $X \sim \mathcal{N}(\mu, \sigma^2)$. The mean and variance of X are

$$\mathbf{E}[X] = \mu$$

$$\textbf{var}(X) = \sigma^2$$

- Let X be a CRV with PDF $f_X(x)$ such that

$$f_X(x) \propto e^{-a^2x^2+bx}$$

Then

$$X \sim \mathcal{N}\left(\mu = \frac{b}{2a^2}, \sigma^2 = \frac{1}{2a^2}\right)$$

Definition (standard Normal)

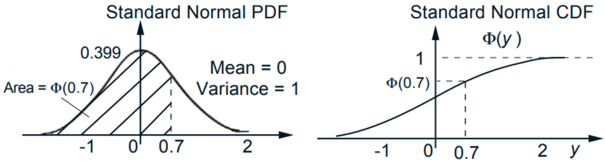
The random variable $Y \sim \mathcal{N}(0,1)$ is called a **standard Normal**.

- The PDF of Y (a standard Normal) is

$$f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$$

- The CDF of Y is denoted by Φ

$$F_Y(y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \Phi(y)$$



Standard Normal table

A standard Normal table stores CDF values of a standard Normal.

Frequently referenced values of Φ are

$$\Phi(0) = 0.5000$$
$$\Phi(1) = 0.8413 \text{ (1 standard deviation)}$$
$$\Phi(2) = 0.9772 \text{ (2 standard deviations)}$$
$$\Phi(3) = 0.9987 \text{ (3 standard deviations)}$$

One-sided table

The CDF value of a negative argument is related to the value of a positive argument by

$$\Phi(y) = 1 - \Phi(-y)$$

$$\begin{aligned}
\Phi(y) + \Phi(-y) &= \int_{-\infty}^{y} f_Z(z)dz + \int_{-\infty}^{-y} f_Z(z)dz \\
&= \int_{-\infty}^{y} f_Z(z)dz + \int_{y}^{\infty} f_Z(z)dz \\
&= \int_{-\infty}^{\infty} f_Z(z)dz \\
&= 1
\end{aligned}$$

Table lookup

Consider $X \sim \mathcal{N}(\mu, \sigma^2)$. The CDF values of X can be looked up in standard Normal table.

- $Y = \frac{X-\mu}{\sigma}$ is a standard normal
- We have equivalent events

$$(X \leq x) = \left(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right)$$
$$= \left(Y \leq \frac{x-\mu}{\sigma}\right)$$

- It follows that

$$P(X \leq x) = P\left(Y \leq \frac{x-\mu}{\sigma}\right)$$
$$= \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Example (3.7 Normal)

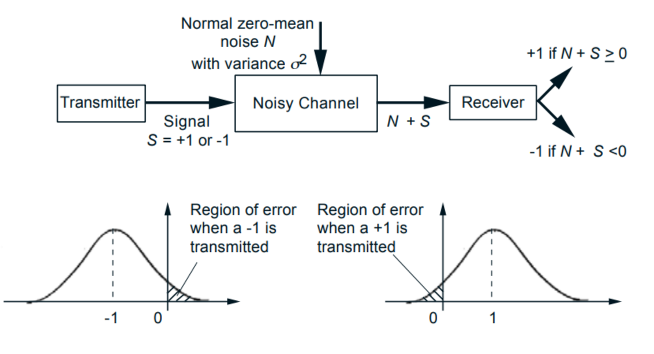
The yearly snowfall at **Mountain Rainier** is modeled as a Normal random variable with a mean of 60 inches and a standard deviation of 20 inches. What is the probability that this year's snowfall will be at least 80 inches?

Let X be the snowfall this year. We have

$$\begin{aligned} P(X \geq 80) &= 1 - P(X < 80) \\ &= 1 - P\left(\frac{X-60}{20} < \frac{80-60}{20}\right) \\ &= 1 - \Phi(1) \\ &= 1 - 0.8413 \\ &= 0.1587 \end{aligned}$$

Example (3.8 Normal noise)

A binary message is transmitted as a signal S, which is either -1 or $+1$. The channel corrupts the transmission by an additive noise $N \sim \mathcal{N}(0, \sigma^2)$. The receiver receives $Y = S + N$ and decides that $S = -1$ (resp. $S = +1$) if $Y < 0$ (resp. $Y \geq 0$). What is the probability of error in transmission?



The error event is $E = \{\text{transmitted signal} \neq \text{decided signal}\}$. The total probability of E with partition $((S = +1), (S = -1))$ is

$$
\begin{aligned}
P(E) &= P(E \cap (S = 1)) + P(E \cap (S = -1)) \\
&= P(E \mid S = 1)P(S = 1) + P(E \mid S = -1)P(S = -1) \\
&= P(Y < 0 \mid S = 1)P(S = 1) + P(Y \geq 0 \mid S = -1)P(S = -1) \\
&= P(N < -1)P(S = 1) + P(N \geq 1)P(S = -1) \\
&= P(N \geq 1)(P(S = 1) + P(S = -1)) \\
&= P(N \geq 1) \\
&= 1 - P(N < 1) \\
&= 1 - P\left(\frac{N - 0}{\sigma} < \frac{1 - 0}{\sigma}\right) \\
&= 1 - \Phi\left(\frac{1}{\sigma}\right)
\end{aligned}
$$

Multiple Random Variables

Definition (joint probability density function)

Let $(\Omega, \mathcal{F}, P)$ be probability model and X, Y be CRVs defined on Ω with ranges $\mathcal{X}$ and $\mathcal{Y}$.

- The distribution of X and Y over $\mathcal{X} \times \mathcal{Y}$ can be specified by a joint probability density function (joint PDF)
- It is such that

$$
\begin{aligned}
& P(X \in (x, x+\delta_x) \cap Y \in (y, y+\delta_y)) \\
= & \underbrace{f_{XY}(x,y)}_{\text{joint PDF of } X \text{ and } Y} \delta_x \delta_y + o(\delta_x \delta_y)
\end{aligned}
$$

Small-region probability

Let X and Y be CRVs with joint PDF f_{XY}.

- $P(X \in (x, x+\delta_x) \cap Y \in (y, y+\delta_y)) \approx f_{XY}(x,y)\,\delta_x\delta_y$
- The probability of a small rectangle is proportional to its area, and the proportional constant is $f_{XY}(x,y)$
- f_{XY} is probability density

Finite-region probability

Let X and Y be CRVs with joint PDF f_{XY}. For any region $S \subset \mathbb{R}^2$, the probability of $(X,Y) \in S$ is the integral of f_{XY} over S. That is

$$P((X,Y) \in S) = \iint_S f_{XY}(x,y)\,dx\,dy$$

Joint PDF to marginal PDF

Let X and Y be CRVs with joint PDF f_{XY}. We have

$$f_X(x) = \int f_{XY}(x,y)dy$$
$$f_Y(y) = \int f_{XY}(x,y)dx$$

From $(X \in (x, x+\delta)) = (X \in (x, x+\delta)) \cap (Y \in (-\infty, \infty))$

$$\begin{aligned} P(X \in (x, x+\delta)) &= \int_x^{x+\delta} \int f_{XY}(x,y)dxdy \\ &= \int_x^{x+\delta} \left(\int f_{XY}(x,y)dy \right) dx \end{aligned}$$

Thus

$$\begin{aligned} f_X(x)\delta + o(\delta) &= \left(\int f_{XY}(x,y)dy \right) \delta + o(\delta) \\ \Rightarrow \ f_X(x) &= \int f_{XY}(x,y)dy \end{aligned}$$

Example (3.9 Joint PDF)

Recall the example about Romeo and Juliet in Chapter 1. Let X and Y be their delays for the date. What is f_{XY}?

The probability density is constant over $\mathcal{X} \times \mathcal{Y}$. That is

$$f_{XY}(x, y) = \begin{cases} c, & \text{if } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}$$

The constant c is determined by normalization

$$\iint f_{XY}(x, y)\, dx\, dy = \int_0^1 \int_0^1 c\ dx\, dy = 1$$

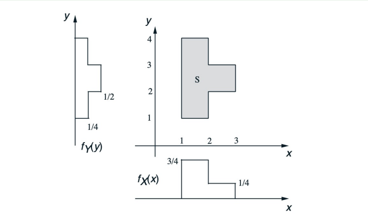
so $c = 1$.

Example (3.10 Joint PDF to marginal PDF)

Suppose the distribution of X and Y is uniform over S and is zero outside S. That is

$$f_{XY}(x, y) = \begin{cases} c, & \text{if } (x, y) \in S \\ 0, & \text{otherwise} \end{cases}$$

Determine c and the PDF f_X.

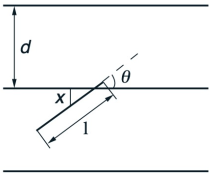


$$\iint f_{XY}(x, y)\, dx\, dy = 1 \;\Rightarrow\; 4c = 1 \;\Rightarrow\; c = \frac{1}{4}$$

$$f_X(x) = \int f_{XY}(x, y) dy = \begin{cases} \frac{3}{4}, & \text{if } 1 < x < 2 \\ \frac{1}{4}, & \text{if } 2 < x < 3 \\ 0, & \text{otherwise} \end{cases}$$

Example (3.11 Buffon's needle)

A surface is ruled with parallel lines separated by distance d. A needle of length $l < d$ is dropped on the surface. What is the probability that the needle crosses a line?



Specific position of the needle can be represented by x and θ, where x is the distance to the closest line from center and θ is the acute angle between needle and line. The needle crosses a line if $\frac{x}{\sin\theta} < \frac{l}{2}$.

Random position of the needle can be represented by random variables X and Θ. Assume uniform joint PDF

$$f_{X\Theta}(x,\theta) = \frac{4}{\pi d}, \quad 0 < x < \frac{d}{2},\ 0 < \theta < \frac{\pi}{2}$$

Consider $A = \{\text{needle crosses a line}\} = \left(X < \frac{l}{2}\sin\Theta\right)$. The probability is

$$\begin{aligned} P(A) &= P\left(X < \frac{l}{2}\sin\Theta\right) \\ &= \iint_{x<\frac{l}{2}\sin\theta} f_{X\Theta}(x,\theta)\,dx\,d\theta \\ &= \int_0^{\frac{\pi}{2}} \int_0^{\frac{l}{2}\sin\theta} \frac{4}{\pi d}\,dx\,d\theta \\ &= \frac{2l}{\pi d}\int_0^{\frac{\pi}{2}} \sin\theta\,d\theta \\ &= \frac{2l}{\pi d} \end{aligned}$$

Conditional Probability Models

Definition (conditional probability density function)

Let $(\Omega, \mathcal{F}, P)$ be a probability model, X be a CRV defined on Ω with range $\mathcal{X}$, and $A \in \mathcal{F}$ have non-zero probability. Conditional on A, the distribution of X can be specified by a conditional probability density function, such that

$$P(X \in (x, x+\delta) \,|\, A) = \overbrace{f_{X|A}(x)}^{\text{conditional PDF}} \delta + o(\delta)$$

For a finite interval (l, u), we have

$$P((X \in (l, u)) \,|\, A) = \int_l^u f_{X|A}(x) dx$$

PDF conditional on $X \in (a, b)$

Consider $A = (X \in (a, b))$. The conditional PDF of X is

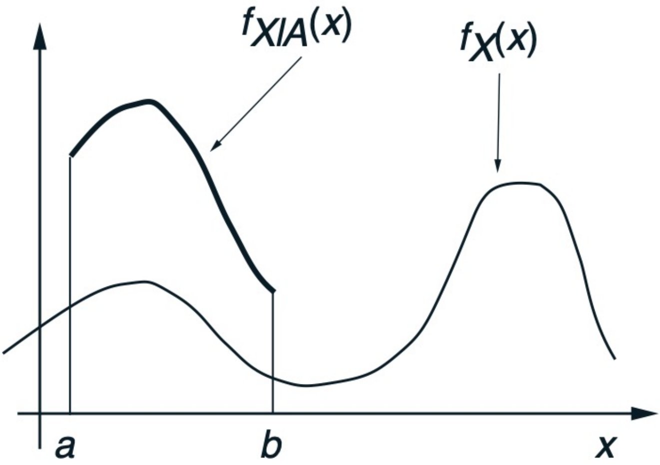
$$f_{X|X\in(a,b)}(x) = \begin{cases} \frac{f_X(x)}{P(X\in(a,b))}, & \text{if } a < x < b \\ 0, & \text{otherwise} \end{cases}$$

$$P(X \in (x, x+\delta) \,|\, X \in (a, b)) = \frac{P(X \in (x, x+\delta) \cap X \in (a, b))}{P(X \in (a, b))}$$

The numerator is 0 if $x \notin (a, b)$. For $x \in (a, b)$

$$f_{X|X\in(a,b)}(x)\,\delta = \frac{P(X \in (x, x+\delta))}{P(X \in (a, b))} = \frac{f_X(x)\,\delta}{P(X \in (a, b))}$$

$$\Rightarrow \; f_{X|X\in(a,b)}(x) = \frac{f_X(x)}{P(X \in (a, b))}$$


$f_{X|A}(x)$
$f_X(x)$
a
b
x

Example (3.13 Conditional PDF)

The time T until a new light bulb burns out is an exponential random variable with parameter λ. Alice turns the light on, leaves the room, and when she returns, t time units later, finds that the light bulb is still on. Let X be the additional time for the light bulb to burn out. What is the PDF of X?

Consider $X = T - t$ conditional on $(T > t)$. For $\tau > 0$

$$\begin{aligned} f_{X|T>t}(\tau) &= f_{T|T>t}(t+\tau) = \frac{f_T(t+\tau)}{P(T>t)} = \frac{\lambda e^{-\lambda(t+\tau)}}{e^{-\lambda t}} \\ &= \lambda e^{-\lambda \tau} \end{aligned}$$

Note a used bulb has the same PDF as a new bulb

$$f_{X|T>t}(\tau) = f_T(\tau)$$

This is the memoryless property of exponential random variables.

Definition (conditional PDF with 2 CRVs)

Let $(\Omega, \mathcal{F}, P)$ be a probability model and X, Y be CRVs defined on Ω with joint PDF f_{XY}. The conditional PDF of X given $Y = y$ is

$$\underbrace{f_{X|Y}(x|y)}_{\text{conditional PDF}} = \frac{f_{XY}(x,y)}{f_Y(y)}$$

Note that conditional PDF can be obtained from joint PDF

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_{XY}(x,y)}{\int f_{XY}(x,y)dx}$$

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(x,y)}{\int f_{XY}(x,y)dy}$$

Rules for joint/marginal/conditional PDFs

Let X and Y be CRVs.

- Factorization

$$f_{XY}(x,y) = f_X(x) f_{Y|X}(y|x)$$

$$f_{XY}(x,y) = f_Y(y) f_{X|Y}(x|y)$$

- Marginalization

$$f_X(x) = \int f_{XY}(x,y)dy, \quad f_Y(y) = \int f_{XY}(x,y)dx$$

- Bayes

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int f_X(x') f_{Y|X}(y|x') dx'}$$

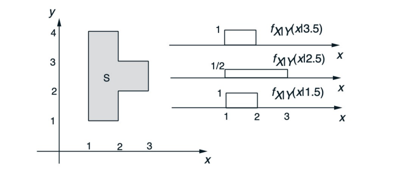
$$f_{Y|X}(y|x) = \frac{f_Y(y) f_{X|Y}(x|y)}{\int f_Y(y') f_{X|Y}(x|y') dy'}$$

Example (Conditional PDF)

Assume uniform distribution over S

$$f_{XY}(x,y) = \frac{1}{4}$$

Find $f_{X|Y}$.



Consider $1 < y < 2$ (similarly for other intervals).

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_{XY}(x,y)}{\int f_{XY}(x,y)dx} = u(x-1) - u(x-2)$$

Thus

$$f_{X|Y}(x|y) = \begin{cases} u(x-1) - u(x-2), & 1 < y < 2 \\ \frac{1}{2}(u(x-1) - u(x-3)), & 2 < y < 3 \\ u(x-1) - u(x-2), & 3 < y < 4 \end{cases}$$

Total probability theorem

Let X be a CRV and $(A_1, \ldots, A_n)$ be a partition of Ω.

$$f_X(x) = \sum_{i=1}^{n} P(A_i) f_{X|A_i}(x)$$

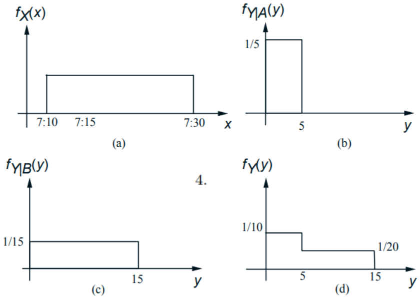
Apply total probability theorem to event $X \in (x, x+\delta)$.

$$\begin{aligned} P(X \in (x, x+\delta)) &= \sum_i P(X \in (x, x+\delta) \cap A_i) \\ &= \sum_i P(A_i) P(X \in (x, x+\delta) \mid A_i) \\ \Rightarrow \ f_X(x)\delta + o(\delta) &= \sum_i P(A_i)(f_{X|A_i}(x)\delta + o(\delta)) \\ \Rightarrow \ f_X(x) &= \sum_i P(A_i) f_{X|A_i}(x) \end{aligned}$$

Example (3.14 TPT)

Trains arrive at a station every quarter. Every morning, Yu walks in the station between 7:10 and 7:30, with all times equally likely. What is the PDF of the waiting time for a train to arrive?

Let X be the time elapsed from 7:10 to the walk-in time and Y be the waiting time. We have $f_X(x) = \frac{1}{20}(u(x-0) - u(x-20))$.



Define $A = \{\text{catch 7:15 train}\}$ and $B = \{\text{catch 7:30 train}\}$.

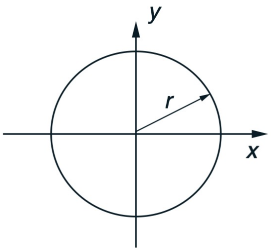
$$P(A) = P(0 < X < 5) = \frac{1}{4},\ f_{Y|A}(y) = \frac{1}{5}(u(y) - u(y-5))$$
$$P(B) = P(5 < X < 20) = \frac{3}{4},\ f_{Y|B}(y) = \frac{1}{15}(u(y) - u(y-15))$$

By total probability theorem

$$\begin{aligned} f_Y(y) &= P(A)f_{Y|A}(y) + P(B)f_{Y|B}(y) \\ &= \frac{1}{4} \cdot \frac{1}{5}(u(y) - u(y-5)) + \frac{3}{4} \cdot \frac{1}{15}(u(y) - u(y-15)) \\ &= \frac{1}{20}(u(y) - u(y-5)) + \frac{1}{20}(u(y) - u(y-15)) \end{aligned}$$

Example (3.15 Joint PDF to conditional PDF)

Ivan throws a dart at a circular target of radius r. We assume that he always hits the target, and that all points of impact are equally likely. Let (X, Y) be the random point of impact. What is the conditional PDF $f_{X|Y}$?



The joint PDF is uniform $f_{XY}(x,y) = (\pi r^2)^{-1}$ for $x^2 + y^2 < r^2$.

$$f_Y(y) = \int f_{XY}(x,y)dx = \int_{-\sqrt{r^2-y^2}}^{\sqrt{r^2-y^2}} \frac{1}{\pi r^2} dx = \frac{2\sqrt{r^2-y^2}}{\pi r^2}, \; |y| < r$$

$$\Rightarrow \; f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{1}{2\sqrt{r^2-y^2}}, \; x^2+y^2 < r^2$$

Example (3.16 Conditional PDF to joint PDF)

The speed of a vehicle that drives past a police radar is modeled as an exponential random variable X with mean 50 miles per hour ($\lambda = \frac{1}{50}$). The police radar's measurement Y of the vehicle's speed has an error which is modeled as a Normal random variable with zero mean and standard deviation equal to one tenth of the vehicle's speed. What is the joint PDF of X and Y?

By factorization (multiplication rule)

$$
\begin{aligned}
f_{XY}(x,y) &= f_X(x) f_{Y|X}(y|x) \\
&= \left(\frac{1}{50} e^{-\frac{x}{50}} u(x)\right) \left(\frac{1}{\sqrt{2\pi}\left(\frac{x}{10}\right)}\, e^{-\frac{(y-x)^2}{2\left(\frac{x}{10}\right)^2}}\right)
\end{aligned}
$$

Definition (conditional expectation)

A conditional expectation of X is the expectation of X with respect to a conditional model of X.

- The expectation of X conditional on A is

$$\mathbf{E}[X|A] = \int x\, f_{X|A}(x) dx$$

- Let Y be CRV. The expectation of X conditional on $Y = y$ is

$$\mathbf{E}[X|Y = y] = \int x\, f_{X|Y}(x|y) dx$$

- Define the conditional expectation of X given Y

$$\mathbf{E}[X|Y] = g(Y) \text{ where } g(y) = \mathbf{E}[X|Y = y]$$

Total expectation theorem

Let $(\Omega, \mathcal{F}, P)$ be probability model and X be CRV defined on Ω.

- Let $A_i \in \mathcal{F}$ and $(A_1, \ldots, A_n)$ be a partition of Ω. Then

$$\mathbf{E}[X] = \sum_{i=1}^{n} P(A_i)\mathbf{E}[X|A_i]$$

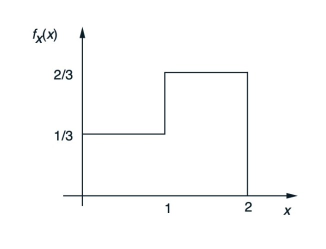
- Let Y be CRV. Then

$$\begin{aligned}
\mathbf{E}[\mathbf{E}[X|Y]] &= \int g(y) f_Y(y) dy \\
&= \int \mathbf{E}[X|Y = y] \, f_Y(y) dy \\
&= \iint x f_{X|Y}(x|y) f_Y(y) dx dy \\
&= \iint x f_{X,Y}(x, y) dx dy \\
&= \mathbf{E}[X]
\end{aligned}$$

Example (3.17 Total expectation)

Find $\mathbf{E}[X]$ and **var**(X) via the partition (A, A^c), where

$$A = (0 < X < 1)$$



$$\mathbf{E}[X] = P(A)\mathbf{E}[X|A] + P(A^c)\mathbf{E}[X|A^c] = \frac{1}{3}\frac{1}{2} + \frac{2}{3}\frac{3}{2} = \frac{7}{6}$$

$$\mathbf{E}\left[X^2\right] = P(A)\mathbf{E}\left[X^2|A\right] + P(A^c)\mathbf{E}\left[X^2|A^c\right] = \frac{1}{3}\frac{1}{3} + \frac{2}{3}\frac{7}{3} = \frac{15}{9}$$

$$\textbf{var}(X) = \mathbf{E}\left[X^2\right] - \mathbf{E}^2[X] = \frac{15}{9} - \frac{49}{36} = \frac{11}{36}$$

Definition (independent random variables)

Let $(\Omega, \mathcal{F}, P)$ be probability model and X and Y be CRVs defined on Ω. X and Y are said to be independent if

$$f_{XY}(x, y) = f_X(x)\, f_Y(y)$$

- The independence of X and Y is denoted by $X \perp\!\!\!\perp Y$
- For $X \perp\!\!\!\perp Y$, we have

$$f_{X|Y}(x|y) = f_X(x)$$

and

$$f_{Y|X}(y|x) = f_Y(y)$$

Bayes Rule

Given CRV conditional on CRV

Let $(\Omega, \mathcal{F}, P)$ be probability model and X and Y be CRVs defined on Ω. Let the PDF of X be f_X and the conditional PDF of Y given X be $f_{Y|X}$.

- Factorization. The joint PDF of X and Y is

$$f_{XY}(x, y) = f_X(x) f_{Y|X}(y|x)$$

- Marginalization. The PDF of Y is

$$f_Y(y) = \int f_{Y|X}(y|x) f_X(x) dx$$

- Bayes rule. The conditional PDF of X given Y is

$$f_{X|Y}(x|y) = \frac{f_X(x) f_{Y|X}(y|x)}{\int f_X(x') f_{Y|X}(y|x') dx'}$$

Example (3.19 Bayes rule)

Assume a light bulb has a lifetime $Y \sim \textbf{Exp}(\lambda)$. Since λ is unknown, we initially assume λ is drawn from **Uni** $\left(1, \frac{3}{2}\right)$. We test a light bulb and record its lifetime y. How can we update the uncertainty on λ?

We have 2 dependent random variables Λ and Y with joint PDF

$$f_{\Lambda Y}(\lambda, y) = f_{\Lambda}(\lambda) f_{Y|\Lambda}(y|\lambda)$$

By Bayes rule, the conditional PDF of Λ is

$$\begin{aligned} f_{\Lambda|Y}(\lambda|y) &= \frac{f_{\Lambda}(\lambda) f_{Y|\Lambda}(y|\lambda)}{\int f_{\Lambda}(\lambda') f_{Y|\Lambda}(y|\lambda') d\lambda'} \\ &= \frac{2\left(u(\lambda-1) - u\left(\lambda - \frac{3}{2}\right)\right) \lambda e^{-\lambda y}}{\int_1^{\frac{3}{2}} 2\lambda' e^{-\lambda' y} d\lambda'} \end{aligned}$$

Note that the updated PDF of Λ depends on y

$$f_{\Lambda|Y}(1^+ \,|\, 2) > f_{\Lambda|Y}\left(\frac{3}{2}^- \,|\, 2\right), \; f_{\Lambda|Y}\left(1^+ \,|\, \frac{1}{3}\right) < f_{\Lambda|Y}\left(\frac{3}{2}^- \,|\, \frac{1}{3}\right)$$

Given CRV conditional on DRV

Let Y be CRV and S be DRV. Let the PMF of S be p_S and the conditional PDF of Y given S be $f_{Y|S}$.

- Factorization. The joint probability of S and Y is

$$f_{SY}(s, y) = p_S(s) f_{Y|S}(y|s)$$

- Marginalization. The PDF of Y is

$$f_Y(y) = \sum_s p_S(s) f_{Y|S}(y|s)$$

- Bayes rule. The conditional PMF of S given Y is

$$p_{S|Y}(s|y) = \frac{f_{SY}(s, y)}{f_Y(y)} = \frac{p_S(s) f_{Y|S}(y|s)}{\sum_{s'} p_S(s') f_{Y|S}(y|s')}$$

Example (3.20 Bayes rule)

A signal S with $P(S = 1) = p$ and $P(S = -1) = 1 - p$ is transmitted, and received as $Y = S + N$, where $N \sim \mathcal{N}(0, 1)$. What is the probability of $(S = 1)$ given $(Y = y)$?

By Bayes rule

$$
\begin{aligned}
p_{S|Y}(1|y) &= \frac{p_S(1) f_{Y|S}(y|1)}{p_S(1) f_{Y|S}(y|1) + p_S(-1) f_{Y|S}(y|-1)} \\
&= \frac{p \, \frac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2}}{p \, \frac{1}{\sqrt{2\pi}} e^{-(y-1)^2/2} + (1-p) \, \frac{1}{\sqrt{2\pi}} e^{-(y+1)^2/2}} \\
&= \frac{pe^y}{pe^y + (1-p)e^{-y}}
\end{aligned}
$$

Given DRV conditional on CRV

Let Θ be CRV and N be DRV. Let the PDF of Θ be f_Θ and the conditional PMF of N given Θ be $p_{N|\Theta}$.

- Factorization. The joint probability of N and Θ is

$$f_{N\Theta}(n,\theta) = p_{N|\Theta}(n|\theta) f_\Theta(\theta)$$

- Marginalization. The PMF of N is

$$p_N(n) = \int p_{N|\Theta}(n|\theta) f_\Theta(\theta) d\theta$$

- Bayes rule. The conditional PDF of Θ given N is

$$f_{\Theta|N}(\theta|n) = \frac{f_{N\Theta}(n,\theta)}{p_N(n)} = \frac{p_{N|\Theta}(n|\theta) f_\Theta(\theta)}{\int p_{N|\Theta}(n|\theta') f_\Theta(\theta') d\theta'}$$

PDF

$$P(X \in (x, x+\delta)) \approx f_X(x)\delta$$

Common CRVs

$$\textbf{Uni}(a,b),\ \textbf{Exp}(\lambda),\ \mathcal{N}(\mu,\sigma^2)$$

CDF

$$F_X(x) = P(X \leq x)$$

Joint PDF

$$P(X \in (x, x+\delta_x) \cap Y \in (y, y+\delta_y)) \approx f_{XY}(x,y)\delta_x\delta_y$$

Factorization

$$f_{XY}(x,y) = f_Y(y) f_{X|Y}(x|y) = f_X(x) f_{Y|X}(y|x)$$

Marginalization

$$f_X(x) = \int f_{XY}(x,y)dy$$

Bayes rule

$$f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{\int f_{XY}(x,y)dx}$$

Total probability

$$f_X(x) = \sum_i P(A_i) f_{X|A_i}(x)$$

$$f_X(x) = \int f_{X|Y}(x|y) f_Y(y) dy$$

Total expectation

$$\mathbf{E}[X] = \sum_i P(A_i) \mathbf{E}[X|A_i]$$

$$\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X|Y]]$$